Department of Genetics and Microbiology, Faculty of Science

Charles University in Prague

Gene expression in early development of *Xenopus laevis*.

Ph.D. thesis

Specialization: Molecular Biology, Genetics and Virology

Supervisor: Prof. MUDr. Jiří Jonák DrSc., Department of Gene Expression, Institute of Molecular Genetics, Academy of Science of the Czech Republic

Prague, 2008

Radek Šindelka

Prohlášení:

Prohlašuji, že jsem nepředložil tuto dizertační práci ani její podstatnou část k získání jiného nebo stejného akademického titulu.

V Praze dne 15. 2. 2008 Mgr. Radek Šindelka

Acknowledgements

I would like to thank to my supervisor Prof. MUDr. Jiří Jonák, DrSc. and Mikael Kubista, Ph.D., who allow me to do my Ph.D. work on exciting projects in their laboratories. I would also like to thank to all members of our laboratory who participated on these projects, especially Z. Ferjentsik for his assistance in many experiments, H. Šanderová, L. Krásný and others, for valuable discussion of results and comments. I would also like to thank A. Mášová, IOCHB AS CR for her kind assistance with some experiments. The work presented in this thesis was done in the Department of Gene Expression, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic. Special thanks go to Petra and my family for their patience, understanding and encouragement throughout my studies.

Table of contents

Abstract 3

1. Introduction 5

1.1 Gene expression and developmental processes 5

1.2 Early development of Xenopus 6

1.3 Methods to study gene expression- temporal and spatial studies 8

1.4 Quantitative real-time RT-PCR 9

1.5 qPCR tomography 10

2. Aims and results of my work 12

3. List of publications 14

Included papers: 14

Papers not included: 14

4. Conclusions 15

3.1 Normalization 15

3.2 Temporal expression profiles 15

3.3 Spatial expression profiles 15

3.4 Summary and future 16

5. References 17

Abbreviations:

A-V axis – animal-vegetal axis (first body axis in *Xenopus* development)

cDNA – complementary DNA

Ct – cycle of threshold

DABCYL – 4-(dimethylamino)azobenzene-4'-carboxylic acid

DNA – deoxyribonucleic acid

EtBr – ethidium bromide

FAM – 6-carboxy fluorescein

JOE – 2, 7-dimethoxy-4, 5-dichloro-6-carboxy-fluorescein

MBT – MidBlastula Transition

MC – melting curve

METRO – messenger transport organizer

qPCR – quantitative real-time RT-PCR

RNA – ribonucleic acid

ROX – 6-carboxy-X-rhodamine

SAGE – serial analysis of gene expression

TAMRA – 6-carboxy-tetra-methyl-rhodamine

Abstract

Every higher organism consists of several hundreds of different cell types. The cell differentiation is a process which requires highly precise regulation of mRNA production and subsequently protein synthesis. During development of an organism each cell gains a unique palette of mRNA and protein molecules, which reflects the cell´s fate. To understand the basic function and regulation of genes that are important during development, mRNA expression profiling is an irreplaceable tool. However, most techniques to determine the mRNA content, such as Northern blot, microarrays and *in situ* hybridization have some limitations in their specificity, dynamic range and/or sensitivity. Quantitative real-time PCR analysis (qPCR) for nucleic acids quantification was introduced in the last decade and it has overcome the above mentioned drawbacks. qPCR has rapidly become the golden standard in basic research as well as in many aspects of applied research such as molecular diagnostic, food pathogen detection and genetically modified organism (GMO) analysis.

We decided to apply qPCR in studies where it helps us to understand basic biological processes that take place in the developing organism. One of the focus areas of the Laboratory of gene expression at IMG AS CR, where I did my PhD thesis, was the role of Src tyrosine kinases in the early development of vertebrates studied on the African clawed frog *Xenopus laevis*. *Xenopus* oocyte and early embryos are huge compared to mammalian ones and contain enormous amount of biological material (RNA, proteins, ribosomes and mitochondria), which can be otherwise obtained only from thousands of somatic cells. Therefore, *Xenopus* has become one of the most popular model organisms for developmental studies.

I was dealing with five projects during my PhD studies:

1. To find reference genes suitable for normalization of temporal analysis of mRNA expression throughout early developmental period of *Xenopus*. Our qPCR examinations revealed that no one of so far routinely used *Xenopus* reference genes met criteria for a developmental reference gene. Instead, we found out that reliable data can be obtained by normalization against total RNA content in individual samples.

2. To determine time profiles of expression of a group of developmental genes at early development of *X. laevis*. To compare these profiles with predicted function(s) of these genes. The expression profiles of 21 important genes during early development were determined and tight connection between the roles of the genes in different developmental stages and their expression was found. New methods for data pretreatment and statistical analysis of multidimensional data were tested on our results.

3. To determine temporal and spatial expression profiles of *Xenopus* Src tyrosine kinases (STK) and Csk , the natural inhibitor of STK at early development. This was achieved by qPCR analysis and whole mount *in situ* hybridization analysis.

4. To determine profiles of spatial distribution of developmental mRNAs within *X. laevis* oocytes. For this purpose a new method, qPCR tomography was developed. mRNA molecules were found to form two distinct gradients along the animal –vegetal axis of the oocyte. The first group is predominantly localized in the animal hemisphere and consists of mRNA expressed from genes such as FoxH1, Oct60, Xmam, elongation factor 1-alpha, GAPDH, GSK3-beta, disheveled, beta-catenin, Tcf-3 and Xpar1. The second group of mRNAs forms a gradient with a maximum in the vegetal hemisphere and consists of mRNAs expressed from genes such as VegT, Vg1, Wnt11, Otx1, Deadsouth, Xcad2, Xpat and Xdazl.

5. To analyze gene expression during the immune response in flash-fly *Sarcophaga bullata*. The expression profiles of 8 genes (transferrin, sapecin, Ppo1, Ppo2, storage binding protein, cathepsin L, 18S rRNA, sarcocystatin) predicted to be involved in immune response were determined in collaboration with the Institute of Organic Chemistry and Biochemistry, AS CR in Prague.

Our results indicate that qPCR is a suitable method for studies of gene expression and mRNA localization during early development. Results obtained during my PhD studies were presented in 7 papers, in several oral presentations at international conferences and in numerous posters. One manuscript is in preparation.

1. Introduction

1.1 Gene expression and developmental processes

When do cells begin to differentiate? When does differentiation become irreversible? These are two basal questions of fundamental importance to understand biological development, and the answers to these questions will be of greatest value in the development of future medical therapies.

The development of organisms is a complicated and highly accurate process that eventually leads to mature animals containing hundreds of differentiated cell types. All cell types originate from a single cell, the fertilized egg. Fertilized egg is created after fusion of male sperm and female oocyte cells. Fusion process equip developing embryo with all necessary material, such as RNA, protein, mitochondria, ribosomes for substantial development.

It is generally accepted that major mechanisms of cellular differentiation and development are results of differential gene expression. The composition of the expression palette, or transcriptome, regulates the cell's biology and determines its fate. The functions of many developmental genes and their expression patterns in various tissues and developmental stages have been described. Typically, genes are part of signaling pathways with up and downstream regulations. Changes in the expression of one gene affect expression of many other genes, some of which may be members of other signaling pathways. Therefore, to understand the complex process of development it is essential to first determine the expression profiles and functional characteristics of key genes in different developmental stages (Koide et al. 2005; Heasman 2006). Most of our information about gene function and expression is based on experiments where expression was measured in a mixture of cells. Such data reflect properties of an average (and artificial) cell, and not the situation/potential/commitment of true cells. The expression among cells may vary greatly even within a cell population (Levsky et al. 2003). Such cellular heterogeneity, which may lead to differentiation and govern development, may be a stochastic consequence of a small number of molecules controlling the transcription machinery. It has been reported that a three-fold change in concentration of morphogenes can guide cells to completely different fates (Smith et al. 2004). These fundamental questions have not been possible to address experimentally, at least not with sufficient precision, until recently because suitable techniques have not been available.

1.2 Early development of Xenopus

Xenopus

Fig. 1 *Xenopus laevis*

Our prime model organism is the water living African clawed frog *Xenopus laevis*. *Xenopus laevis* and *Xenopus tropicalis* are ideal organisms for studies of the early development (Amaya 2005; Smith 2005; Showell et al. 2007). *Xenopus laevis* is better suited for surgical manipulations in early developmental stages (1.2 mm egg diameter) but it also has some disadvantages, such as a pseudotetraploid genome, and the relatively long time it takes to mature (2 years). *Xenopus tropicalis* has a diploid genome which is partially sequenced and it matures in half a year, but it is more difficult to handle and cultivate.

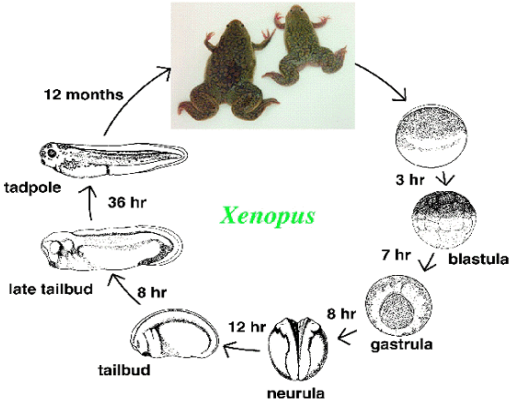


Fig. 2 Early development of *Xenopus laevis*. Time schedule at optimal growth temperature (~20°C).

From two to three *Xenopus* females thousands of embryos can be collected on a daily basis throughout the whole year. The eggs and early cells are huge in comparison to the mammalian ones and therefore highly suitable for microsurgical manipulations. Protein and mRNA content from one *Xenopus* egg is sufficient for thousands of somatic cells which will evolve from the fertilized egg by mere division. mRNA molecules produced during oogenesis are called maternal (White et al. 2007). The stored mRNAs code mainly for specific transcription factors that subsequently determine the fates of the different parts of the developing embryo and in some cases even the fates of individual cells. Furthermore, cell differentiation depends not only on the presence of particular mRNA molecules, but also on their expression levels.

***Xenopus* oocyte**

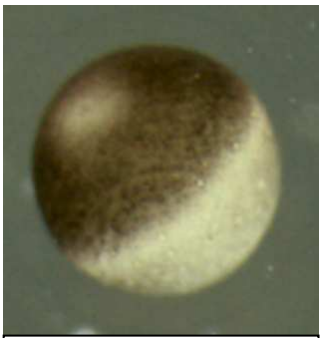


Fig. 3 *Xenopus laevis* oocyte

The *Xenopus* oocyte has two main regions – the animal hemisphere (pigmented- dark) and the vegetal hemisphere (yolky- light). It is clear that specific distribution of mRNAs along the animal-vegetal axis already exists before fertilization. This distribution is established during oogenesis. After fertilization these mRNAs are further distributed within the cell, and become unequally distributed among daughter cells giving rise to differentiation and later specialization. Two distinct pathways lead to vegetal polarization of RNAs in the oocyte with 20-50 fold variation in concentration. In the early or METRO pathway many coding and non-coding RNAs essential to germ cell development associate with a structure

called the mitochondrial cloud that locates towards the vegetal pole (Heasman et al. 1984). The late pathway locates germ layer determinants to the vegetal pole. These RNAs are excluded from the mitochondrial cloud in the early stages and accumulate within a wedge-like region within the vegetal ooplasm (Deshler et al. 1997). While the two pathways are mechanically distinct (Zearfoss et al. 2003; Zearfoss et al. 2004), a subset of vegetally localized RNAs has been identified with characteristics of both pathways. Another set of expression products of some twenty genes has been found to localize at the animal pole. The mechanism behind the animal pole localization is not known, and it leads to a concentration increase of merely 3-10-fold. Some of the vegetally localized mRNAs encode proteins that serve as mesoderm and endoderm inducers (e.g. Vg1, VegT, Xwnt11, Xcat-2, (Kofron et al. 1999; Mowry et al. 1999; MacArthur et al. 2000; Pannese et al. 2000; Kataoka et al. 2005; Machado et al. 2005; Tao et al. 2005). Other vegetal mRNAs, such as Xcat2, Xdaz1 and Deadsouth, that localize in the so called germ plasm, encode proteins that induce the formation of future germ cells (MacArthur et al. 1999; Houston et al. 2000). Unfortunately, functions of the most of these specifically localized mRNAs are unclear, particularly for those located in the animal hemisphere.

Another major redistribution of mRNA molecules occurs when a sperm enters the egg. This process is termed cortical rotation, during which the cortex rotates by about 30^{o} (Vincent et al. 1986; Gerhart et al. 1989) and redistributes mRNA molecules important for specification of the dorso-ventral axis (e.g. beta-catenin into the future dorsal site, (Wylie et al. 1996). The site of the sperm entry into the animal hemisphere determines the future ventral side.

Early development of *Xenopus* (from oocyte to tadpole)

After sperm entry, *Xenopus* oocyte undergoes rapid cell division. First 12 cell divisions occur in 30 minutes intervals. Resulting is an embryo of around 4000 cells. As mentioned above, transcription of zygotic genes is completely silenced till this stage. Therefore maternal RNA and protein content is only distributed into daughter cells and original localization of these factors in egg determines future cell fates.

While in mammals and birds zygotic mRNA production starts in one of the first developmental stages, most amphibians provide their progeny with amount of maternal RNA sufficient for these first developmental steps and *Xenopus* transcription is silenced until the early gastrula stage (MBT - MidBlastula Transition). Transcription of zygotic genes initiates only after this stage (Newport et al. 1982; Newport et al. 1982; Masui et al. 1998). Transcription activities of zygotic genes vary among cells, and depend particularly on the position of the cell in the embryo. Therefore, cell fate is determined by spatial and temporal distribution/expression of maternal and zygotic mRNAs.

Why we used *Xenopus*?

Xenopus laevis early development and transgenesis were one of the main subjects studied in the Laboratory of gene expression, Institute of Molecular Genetics AS CR (Takac et al. 1992; Habrova et al. 1996; Takac et al. 1998; Dvorakova et al. 2000; Jonak 2000; Krylov et al. 2003). Hundreds of identical eggs which each contain micrograms of RNA and proteins, can be obtained after stimulation of a single female. From this point of view *Xenopus* early development appears to be irreplaceable as a model system for large scale temporal and spatial expression profiling studies. Further detailed knowledge of gene expression patterns at the single cell level is necessary for understanding gene function, and for mapping signaling pathways (Koide et al. 2005). A reliable and sensitive analysis of gene expression patterns at the single cell level would correlate mRNA distribution with the fate of the cells and with developmental processes.

1.3 Methods to study gene expression- temporal and spatial studies

Several methods for quantification and detection of nucleic acids in biological samples (e.g. Northern blotting, SAGE, microarrays, *in situ* hybridization) were introduced into molecular biology in the last decades. But these techniques differ in specific properties and are used in different experiments.

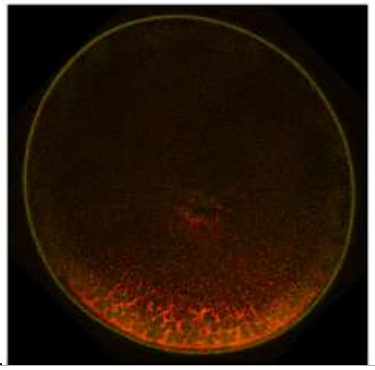


Fig. 4 Localization of Vg1 mRNA in vegetal pole of *Xenopus* egg

Usually target molecules are specifically detected and quantified in complex sample. Results are further analyzed and conclusions are derived. Most of these methods are frequently used to study biological processes such as silencing (Heasman 2002) and induction of transcription, RNA degradation (Fleige et al. 2006) and RNA transport (Zhou et al. 1996; Zhou et al. 2004). Two types of experiments can be distinguished: i) temporal approach, to determine expression changes in time, ii) spatial approach, to compare different localization of target molecules in space such as in organism body, tissue or even within single cell.

Methods used for temporal and spatial expression measurement in *Xenopus* studies:

a) temporal: Ribonuclease protection assay, Northern blotting (Watanabe et al. 2005), RT- PCR (Wardle et al. 2004), SAGE (serial analysis of gene expression) (Blomberg et al. 2004), microarrays and macroarrays (Altmann et al. 2001; Chalmers et al. 2005; Shin et al. 2005; Hufton et al. 2006).

b) spatial: *in situ* hybridization, whole mount *in situ* hybridization (Harland 1991; Kataoka et al. 2005).

All mentioned methods have some limitation. However, during the last decade, quantitative real-time reverse transcription PCR (RT-QPCR) technology has been developed, by which the expression of selected marker genes can be measured with unprecedented accuracy and sensitivity even in minute sample amounts.

1.4 Quantitative real-time RT-PCR

After its invention in 1983 by Kary Mullis (Saiki et al. 1985), polymerase chain reaction (PCR) became a golden standard for nucleic acid detection and amplification in biochemical and biological laboratories. Further steps involved development of procedures for nucleic acid quantification, particularly quantification in real time during cycling, first described in 1992 by Higuchi et. al (Higuchi et al. 1992). Rapid development of instrumentation that allows termocycling and parallel fluorescent measurement was used in a technique now called quantitative real-time RT-PCR, or shortly qPCR. To date qPCR is used for many different purposes, such as food pathogen detection, infection and virus detection, cancer diagnostic and of course for basic research (Bustin 2000; Bustin et al. 2004; Kubista et al. 2006; Nolan et al. 2006).

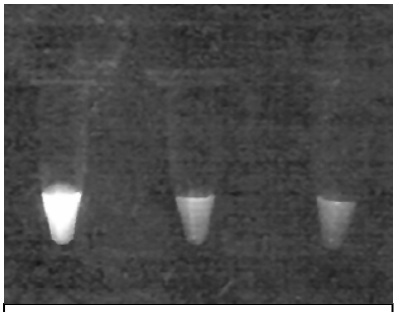


Fig. 5 Tubes with PCR product and ethidim bromide after different number of cycles. Number of cycles decreases from left to right. The product is detected by EtBr fluorescence.

The only difference between qPCR and classical PCR is the presence of a fluorescent marker in qPCR. Fluorescence of the marker is measured in every cycle of the PCR. The earlier the signal appears (lower cycle) the higher is the number of target molecules in the original reaction mixture. There are two types of fluorescent markers: i) nonspecific binding dyes and ii) sequence specific probes.

i) Nonspecific binding dyes

Ethidium bromide was the first dye used for quantification of nucleic acids. Plenty of different fluorescent intercalating dyes with similar sequences but different properties were introduced during the last decade (Bengtsson et al. 2003). Nowadays, the most widely used qPCR dye is SYBR Green I (Molecular Probes) (Wittwer et al. 1997; Morrison et al. 1998). Free in solution the dye absorbs proper light and convert its energy into heat. But when intercalated into a nucleic acid molecule, the dye becomes fixed and cannot turn light to heat. Instead, the dye emits fluorescence. Dyes intercalate into double stranded DNA molecules. This intercalation is sequence independent. Therefore, nonspecific DNA molecules produced by PCR also give a signal. Specific analysis of the PCR product is necessary. Melting curve analysis after the PCR, gel electrophoresis of the PCR products and sequencing are used most frequently.

Melting curve analysis

Melting curve analysis (Ririe et al. 1997) is based on the fact that different DNA molecules melt at different temperature. Melting temperature depends on length and base composition (longer molecules and molecules rich in Cs and Gs have higher melting temperature). Temperature is gradually increased in 0.5 – 1 ^{o}C steps and the fluorescent signal is measured. When reaction temperature reaches the melting point of a dsDNA molecule, its strands separate and the fluorescent dye is released. Then fluorescence immediately drops. Second derivate gives us peak report with melting temperatures.

High resolution melting curve

Modification of the melting curve analysis, called high resolution melting curve analysis, was introduced by (Wittwer et al. 2003). Fluorescent dyes with a lower binding strength, such as LC Green, Eva Green, BEBO (Bengtsson et al. 2003) are used. After completed PCR, temperature in the wells is gradually increased (~0.1°C) and the fluorescence is measured. Because of the high resolution of the melting curves, single nucleotide polymorphisms can be detected through subtle differences in stability of the duplex. The high resolution melting analysis can be carried out today not only in specialized instruments such as Idaho Technology- HR1, but also with two cyclers of the new generation for qPCR: Corbett Life Science- RotorGene 6000 and Roche- LightCycler 480.

ii) Probe based detection

Around 10 types of sequence specific probes for qPCR have been developed and introduced in last 20 years. Advantage of the specific probes compared to nonspecific dyes primarily lies with sequence specificity. The probe is usually a construct composed of a short oligonucleotide strand, a fluorescent dye (FAM, JOE, ROX) and a quencher (e.g DABCYL, TAMRA). Different design principles may lead to different mechanisms of detection. TaqMan probes have been broadly used for gene expression studies and are cleaved during the elongation step of PCR (Heid et al. 1996). Some probes such as Molecular Beacons (Bonnet et al. 1999), Scorpions (Saha et al. 2001) and hybridization probes are frequently used for mutation detection and fluorescent signal arises due to a change of the secondary structure of the probe. Other sets of probes with different fluorescent reporters also allow multiplexing.

Quantification and data analysis

Two approaches are available for quantification of qPCR data analysis. i) absolute quantification: Ct (cycle of threshold) values for unknown samples are interpreted against a standard curve, that is based on samples of known RNA, cDNA, DNA concentrations. ii) relative quantification: expression of a target gene is compared with a proper reference (such as the number of cells, sample volume, DNA content or the expression of one or more reference “housekeeping” genes) (Pfaffl 2001; Vandesompele et al. 2002). The expression of reference genes should be constant in all biological samples. This is why thorough validation of reference genes should be performed in every experiment (Bustin et al. 2004; Nolan et al. 2006).

Gene expression profiling of multiple genes in many biological samples generates large amounts of data. We have co-developed particularly powerful data mining methods for analysis of the very large data sets, that are implemented in the GenEx software from MultiD Analyses (www.multid.se) (Kubista et al. 2006; Kubista et al. 2007; Kubista et al. 2007). The data are statistically analyzed and visualized graphically to help determine biological conclusions.

1.5 qPCR tomography

qPCR is exclusively used in absolute and relative quantification and temporal expression studies, where expression of several genes is compared in several samples. *Xenopus* oocytes, as mentioned

above, have two differently pigmented hemispheres: animal and vegetal. This creates the first axis in *Xenopus* early development, the so called A-V axis. Along the A-V axis, gradients of biological factors (such as mRNA, proteins) are formed (Mowry et al. 1999; King et al. 2005). These factors are called morphogens (Neumann et al. 1997). Oocytes were cut by a cryostat along the A-V axis into slices. RNA was isolated from the slices and concentration of several mRNAs was determined by qPCR. From the qPCR results an expression pattern map for 16 maternal genes was constructed (Sindelka et al. 2008).

2. Aims and results of my work

At the beginning of my PhD studies I and Z. Ferjentsik were asked to introduce qPCR to the laboratory and to adopt it to efficiently monitor silencing effect of siRNA oligonucleotides, targeted against Src tyrosine kinase mRNAs, injected into *Xenopus laevis* one cell embryos. The role of Src kinases in *Xenopus* early development was one of the main subjects studied here for several years (see above). During that period, we discovered advantages of qPCR for studies of temporal expression profiling in early development. Then, the main aim of my work became application of qPCR for the description of correlations between mRNA expression, localization and gene function during early development of the frog *Xenopus laevis*.

Aims:

i) The first objective of the work was to determine, in the course of early development, expression stability of several genes (such as GAPDH, elongation factor 1 aplha, ODC, H4 and L8), widely used as references in mRNA studies.

ii) The second goal was to study correlation between regulation of gene expression and gene function by determining expression profiles of several genes in early developmental stages of *X. laevis*.

iii) The third objective originating from the longstanding interest of our laboratory in the role of Src tyrosine kinases in early development, was to look at temporal and spatial expression patterns of kinases (Src, Fyn, Yes, Lyn and Laloo) and their regulator (Csk kinase).

iv) The forth aim was to develop a technique for spatial expression analysis based on qPCR quantification. Our choice were *Xenopus* oocytes and embryos because of huge size and abundance of RNA. Gradients of maternal mRNA molecules were determined in slices prepared by cryosectioning.

v) Finally, in collaboration with the Department of Biochemistry and Molecular Biology of Institute of Organic Chemistry and Biochemistry, AS CR to measure gene activation at the level of mRNA, in flesh-fly *Sarcophaga bullata* during the immune response.

Results:

a) I determined temporal expression profiles of 6 reference genes (Sindelka et al. 2006). This experiment was designed to find suitable references for temporal expression studies.

b) I measured temporal expression profiles of 16 genes important for early development of *Xenopus*, in 16 developmental stages (Kubista et al. 2006; Kubista et al. 2007; Kubista et al. 2007). Genex statistical software, which was co-developed with MultiD, was applied for the statistical data analysis.

c) Together with my colleague Zoltan Ferjentsik, I determined temporal and spatial expression profiles of genes from the Src-kinases family (Ferjentsik et al. 2008). The mRNA profiles revealed where and when the kinases are expressed.

d) I developed qPCR tomography technique for spatial expression profiling in *Xenopus* oocytes and determined localization of maternal mRNAs in the oocytes (Sindelka et al. 2008).

e) In collaboration with IOCHB, AS CR expression of some genes during the immune response in the larvae of the flesh-fly *Sarcophaga bullata* was determined.

3. List of publications

Included papers:

I. R. Sindelka, Z. Ferjentsik, and J. Jonak, Developmental expression profiles of *Xenopus laevis* reference genes. Dev Dyn, 2006. 235(3): p. 754-8.

II. M. Kubista, J. M. Andrade, M. Bengtsson , A. Forootan, J. Jonák, K. Lind, R. Sindelka, R. Sjöback, B. Sjögreen, L. Strömbom, A.Ståhlberg & N. Zoric. The Real-time Polymerase Chain Reaction. Molecular Aspects of Medicine, 2006. 27, 95-125.

III. M. Kubista, B. Sjogreen, A. Forootan, R. Sindelka, J. Jonak, JM. Andrade. Real-time PCR gene expression profiling. European Pharmaceutical Reviews, 2007. Vol. 1.

IV. M. Kubista, R. Sindelka, A. Tichopad, A. Bergkvist, D. Lindh, A. Forootan, The Prime Technique. Real-time PCR data analysis. G.I.T. Laboratory Journal, 2007. 9-10: p. 33-35.

V. R. Sindelka, J. Jonák, R. Hands, SA. Bustin, M. Kubista, Intracellular expression profiles measured by real-time PCR tomography in the *Xenopus laevis* oocyte. Nucleic Acids Res, 2008. 36(2),387-92.

VI. Z. Ferjentsik, R. Sindelka, G. Lin, J. Jonák, Expression patterns of Src-family tyrosine kinases during *Xenopus laevis* development, Int. J. Dev. Biol., 2008.

Papers not included:

VII. A. Ciencialová, T. Neubauerová, M. Sanda, R. Sindelka, J. Cvačka, Z. Voburka, M. Buděšínský, V. Kašička, P. Sázelová, V. Solínová, M. Macková, B. Koutek, J. Jiráček, Mapping the peptide and protein immune response in the larvae of the fleshfly Sarcophaga bullata. J Pept Sci., 2007. 29

VIII. A. Ciencialová, R. Sindelka, M. Kubista, J. Jiráček, Gene expression after immune respond in the larvae of the fleshfly Sarcophaga bullata., in preparation

4. Conclusions

3.1 Normalization

We found that there is no suitable reference gene useful for temporal (Sindelka et al. 2006) and spatial (Sindelka et al. 2008) expression profiling in *Xenopus laevis* early development. Therefore, the normalization against total RNA is recommended as significantly more reliable than normalization against any of the so called reference genes. In agreement with this, the same amount of RNA from all biological samples should be used in the reverse transcription step (Stahlberg et al. 2004; Stahlberg et al. 2004).

3.2 Temporal expression profiles

Temporal expression profiles of all 21 developmental genes correlated with expected biological function of the genes. Based on the measured profiles, three different groups of genes were distinquished (Kubista et al. 2006; Kubista et al. 2007): i) maternally expressed genes such as Vg1, Wnt11, p53, VegT and dishevelled that prevail in stages before MBT, ii) MBT genes such as Xbra and cerberus, expression of which was highly increased around MBT stages, iii) zygotic genes, such as activin, HNF-3beta, goosecoid, siamois, follistation, N-CAM, chordin, derriere, expression of which began after MBT. Transcription of zygotic genes is usually controlled by transcription factors, that are produced during the first cell divisions after fertilization (such as beta-catenin, Vg1, Wnt11 and VegT). Indeed, mRNA molecules coding for these transcription factors are highly accumulated in oocytes in the maternal mRNA pool. Furthermore, the control of transcriptional induction and silencing is not dependent only on time, but also on spatial localization of these factors.

3.3 Spatial expression profiles

To detect spatial localization of mRNA molecules, I developed qPCR tomography. At first, I used it to measure intracellular gradients of mRNAs in the *Xenopus* oocyte. I found two distinct mRNA gradients after measurement of gene expression patterns of 20 maternal genes. The first group of genes called vegetal includes: VegT, Vg1, Wnt11, Otx1, Deadsouth, Xcad2, Xpat, Xdazl. All of these transcripts were localized in the oocyte in gradients and were the most abundant in the vegetal pole and the least abundant in the animal pole. The second group, which consists of FoxH1, Oct60, Xmam, elongation factor 1-alpha, GAPDH, GSK3-beta, disheveled, beta-catenin, Tcf-3 and Xpar1 transcripts, dominated in animal hemisphere. The qPCR tomography technology is being used in our laboratory to also study other systems such as the developing tooth and neurosphere formation and the results appear to be very promising.

3.4 Summary and future

In the last few years, qPCR technique has become a golden standard for quantification of nucleic acids in biological samples. Thanks to the extreme dynamic range, specificity and sensitivity of qPCR, it is now possible to study changes in RNA expression and localization even at the single cell level. We applied qPCR to describe basic biological processes occurring in early development of the frog *Xenopus laevis*. In the last several years we determined expression profiles of many of its genes in different early developmental stages not only during time, but also at the spatial level. Our results provide a small although vital piece to the understanding of biological functions of the genes expressed in early development and help us to decide how to continue further.

Functional tests, which could reveal real functions of the genes will follow, may be not in our laboratory, but in collaboration with groups that have complementary skills.

5. References

Altmann, C. R., E. Bell, A. Sczyrba, J. Pun, S. Bekiranov, T. Gaasterland and A. H. Brivanlou (2001). "Microarray-based analysis of early development in Xenopus laevis." Dev Biol **236**(1): 64-75.

Amaya, E. (2005). "Xenomics." Genome Res **15**(12): 1683-91.

Bengtsson, M., H. J. Karlsson, G. Westman and M. Kubista (2003). "A new minor groove binding asymmetric cyanine reporter dye for real-time PCR." Nucleic Acids Res **31**(8): e45.

Blomberg, L. A. and K. A. Zuelke (2004). "Serial analysis of gene expression (SAGE) during porcine embryo development." Reprod Fertil Dev **16**(2): 87-92.

Bonnet, G., S. Tyagi, A. Libchaber and F. R. Kramer (1999). "Thermodynamic basis of the enhanced specificity of structured DNA probes." Proc Natl Acad Sci U S A **96**(11): 6171-6.

Bustin, S. A. (2000). "Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays." J Mol Endocrinol **25**(2): 169-93.

Bustin, S. A. and T. Nolan (2004). "Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction." J Biomol Tech **15**(3): 155-66.

Chalmers, A. D., K. Goldstone, J. C. Smith, M. Gilchrist, E. Amaya and N. Papalopulu (2005). "A Xenopus tropicalis oligonucleotide microarray works across species using RNA from Xenopus laevis." Mech Dev **122**(3): 355-63.

Deshler, J. O., M. I. Highett and B. J. Schnapp (1997). "Localization of Xenopus Vg1 mRNA by Vera protein and the endoplasmic reticulum." Science **276**(5315): 1128-31.

Dvorakova, K., V. Habrova, M. Takac and J. Jonak (2000). "Depression in the level of cadherin and alpha-, beta-, gamma-catenins in transgenic Xenopus laevis highly expressing c-Src." Folia Biol (Praha) **46**(1): 3-9.

Ferjentsik, Z., R. Sindelka, G. Lin and J. Jonák (2008). "Expression patterns of Src-family tyrosine kinases during Xenopus laevis development." Int J Dev Biol **in press.**

Fleige, S. and M. W. Pfaffl (2006). "RNA integrity and the effect on the real-time qRT-PCR performance." Mol Aspects Med **27**(2-3): 126-39.

Gerhart, J., M. Danilchik, T. Doniach, S. Roberts, B. Rowning and R. Stewart (1989). "Cortical rotation of the Xenopus egg: consequences for the anteroposterior pattern of embryonic dorsal development." Development **107 Suppl**: 37-51.

Habrova, V., M. Takac, J. Navratil, J. Macha, N. Ceskova and J. Jonak (1996). "Association of rous sarcoma virus DNA with Xenopus laevis spermatozoa and its transfer to ova through fertilization." Mol Reprod Dev **44**(3): 332-42.

Harland, R. M. (1991). "In situ hybridization: an improved whole-mount method for Xenopus embryos." Methods Cell Biol **36**: 685-95.

Heasman, J. (2002). "Morpholino oligos: making sense of antisense?" Dev Biol **243**(2): 209-14.

Heasman, J. (2006). "Patterning the early Xenopus embryo." Development **133**(7): 1205-17.

Heasman, J., J. Quarmby and C. C. Wylie (1984). "The mitochondrial cloud of Xenopus oocytes: the source of germinal granule material." Dev Biol **105**(2): 458-69.

Heid, C. A., J. Stevens, K. J. Livak and P. M. Williams (1996). "Real time quantitative PCR." Genome Res **6**(10): 986-94.

Higuchi, R., G. Dollinger, P. S. Walsh and R. Griffith (1992). "Simultaneous amplification and detection of specific DNA sequences." Biotechnology (N Y) **10**(4): 413-7.

Houston, D. W. and M. L. King (2000). "A critical role for Xdazl, a germ plasm-localized RNA, in the differentiation of primordial germ cells in Xenopus." Development **127**(3): 447-56.

Hufton, A. L., A. Vinayagam, S. Suhai and J. C. Baker (2006). "Genomic analysis of Xenopus organizer function." BMC Dev Biol **6**: 27.

Jonak, J. (2000). "Sperm-mediated preparation of transgenic Xenopus laevis and transmission of transgenic DNA to the next generation." Mol Reprod Dev **56**(2 Suppl): 298-300.

Kataoka, K., A. Tazaki, A. Kitayama, N. Ueno, K. Watanabe and M. Mochii (2005). "Identification of asymmetrically localized transcripts along the animal-vegetal axis of the Xenopus egg." Dev Growth Differ **47**(8): 511-21.

King, M. L., T. J. Messitt and K. L. Mowry (2005). "Putting RNAs in the right place at the right time: RNA localization in the frog oocyte." Biol Cell **97**(1): 19-33.

Kofron, M., T. Demel, J. Xanthos, J. Lohr, B. Sun, H. Sive, S. Osada, C. Wright, C. Wylie and J. Heasman (1999). "Mesoderm induction in Xenopus is a zygotic event regulated by maternal VegT via TGFbeta growth factors." Development **126**(24): 5759-70.

Koide, T., T. Hayata and K. W. Cho (2005). "Xenopus as a model system to study transcriptional regulatory networks." Proc Natl Acad Sci U S A **102**(14): 4943-8.

Krylov, V., J. Macha, T. Tlapakova, M. Takac and J. Jonak (2003). "The c-SRC1 gene visualized by in situ hybridization on Xenopus laevis chromosomes." Cytogenet Genome Res **103**(1-2): 169-72.

Kubista, M., J. M. Andrade, M. Bengtsson, A. Forootan, J. Jonak, K. Lind, R. Sindelka, R. Sjoback, B. Sjogreen, L. Strombom, A. Stahlberg and N. Zoric (2006). "The real-time polymerase chain reaction." Mol Aspects Med **27**(2-3): 95-125.

Kubista, M., R. Sindelka, A. Tichopad, A. Bergkvist, D. Lindh and A. Forootan (2007). "The Prime Technique. Real-time PCR data analysis." G.I.T. Laboratory Journal **9-10**: 33-35.

Kubista, M., B. Sjogreen, A. Forootan, R. Sindelka, J. Jonak and J. M. Andrade (2007). "Real-time PCR gene expression profiling." European Pharmaceutical Reviews **1**.

Levsky, J. M. and R. H. Singer (2003). "Gene expression and the myth of the average cell." Trends Cell Biol **13**(1): 4-6.

MacArthur, H., M. Bubunenko, D. W. Houston and M. L. King (1999). "Xcat2 RNA is a translationally sequestered germ plasm component in Xenopus." Mech Dev **84**(1-2): 75-88.

MacArthur, H., D. W. Houston, M. Bubunenko, L. Mosquera and M. L. King (2000). "DEADSouth is a germ plasm specific DEAD-box RNA helicase in Xenopus related to eIF4A." Mech Dev **95**(1-2): 291-5.

Machado, R. J., W. Moore, R. Hames, E. Houliston, P. Chang, M. L. King and H. R. Woodland (2005). "Xenopus Xpat protein is a major component of germ plasm and may function in its organisation and positioning." Dev Biol **287**(2): 289-300.

Masui, Y. and P. Wang (1998). "Cell cycle transition in early embryonic development of Xenopus laevis." Biol Cell **90**(8): 537-48.

Morrison, T. B., J. J. Weis and C. T. Wittwer (1998). "Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification." Biotechniques **24**(6): 954-8, 960, 962.

Mowry, K. L. and C. A. Cote (1999). "RNA sorting in Xenopus oocytes and embryos." FASEB J **13**(3): 435-45.

Neumann, C. and S. Cohen (1997). "Morphogens and pattern formation." Bioessays **19**(8): 721-9.

Newport, J. and M. Kirschner (1982). "A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage." Cell **30**(3): 675-86.

Newport, J. and M. Kirschner (1982). "A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription." Cell **30**(3): 687-96.

Nolan, T., R. E. Hands and S. A. Bustin (2006). "Quantification of mRNA using real-time RT-PCR." Nat Protoc **1**(3): 1559-82.

Nolan, T., R. E. Hands, W. Ogunkolade and S. A. Bustin (2006). "SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations." Anal Biochem **351**(2): 308-10.

Pannese, M., R. Cagliani, C. L. Pardini and E. Boncinelli (2000). "Xotx1 maternal transcripts are vegetally localized in Xenopus laevis oocytes." Mech Dev **90**(1): 111-4.
Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time RT-PCR." Nucleic Acids Res **29**(9): e45.
Ririe, K. M., R. P. Rasmussen and C. T. Wittwer (1997). "Product differentiation by analysis of DNA melting curves during the polymerase chain reaction." Anal Biochem **245**(2): 154-60.
Saha, B. K., B. Tian and R. P. Bucy (2001). "Quantitation of HIV-1 by real-time PCR with a unique fluorogenic probe." J Virol Methods **93**(1-2): 33-42.
Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich and N. Arnheim (1985). "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia." Science **230**(4732): 1350-4.
Shin, Y., A. Kitayama, T. Koide, D. A. Peiffer, M. Mochii, A. Liao, N. Ueno and K. W. Cho (2005). "Identification of neural genes using Xenopus DNA microarrays." Dev Dyn **232**(2): 432-44.
Showell, C. and F. L. Conlon (2007). "Decoding development in Xenopus tropicalis." Genesis **45**(6): 418-26.
Sindelka, R., Z. Ferjentsik and J. Jonak (2006). "Developmental expression profiles of Xenopus laevis reference genes." Dev Dyn **235**(3): 754-8.
Sindelka, R., J. Jonak, R. Hands, S. A. Bustin and M. Kubista (2008). "Intracellular expression profiles measured by real-time PCR tomography in the Xenopus laevis oocyte." Nucleic Acids Res **36**(2): 387-92.
Smith, J. C. (2005). "Xenopus genetics and genomics." Mech Dev **122**(3): 259-62.
Smith, J. C. and J. B. Gurdon (2004). "Many ways to make a gradient." Bioessays **26**(7): 705-6.
Stahlberg, A., J. Hakansson, X. Xian, H. Semb and M. Kubista (2004). "Properties of the reverse transcription reaction in mRNA quantification." Clin Chem **50**(3): 509-15.
Stahlberg, A., M. Kubista and M. Pfaffl (2004). "Comparison of reverse transcriptases in gene expression analysis." Clin Chem **50**(9): 1678-80.
Takac, M., J. Cerna, V. Habrova, J. Stokrova, I. Rychlik and J. Jonak (1992). "Microinjection of cloned proviral Rous sarcoma virus DNAs into Xenopus laevis one cell embryos." Folia Biol (Praha) **38**(2): 65-77.
Takac, M., V. Habrova, J. Macha, N. Ceskova and J. Jonak (1998). "Development of transgenic Xenopus laevis with a high C-src gene expression." Mol Reprod Dev **50**(4): 410-9.
Tao, Q., C. Yokota, H. Puck, M. Kofron, B. Birsoy, D. Yan, M. Asashima, C. C. Wylie, X. Lin and J. Heasman (2005). "Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos." Cell **120**(6): 857-71.
Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe and F. Speleman (2002). "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes." Genome Biol **3**(7): RESEARCH0034.
Vincent, J. P. and J. C. Gerhart (1986). "A reinvestigation of the process of grey crescent formation in Xenopus eggs." Prog Clin Biol Res **217B**: 349-52.
Wardle, F. C. and J. C. Smith (2004). "Refinement of gene expression patterns in the early Xenopus embryo." Development **131**(19): 4687-96.
Watanabe, T., A. Takeda, K. Mise, T. Okuno, T. Suzuki, N. Minami and H. Imai (2005). "Stage-specific expression of microRNAs during Xenopus development." FEBS Lett **579**(2): 318-24.
White, J. A. and J. Heasman (2007). "Maternal control of pattern formation in Xenopus laevis." J Exp Zoolog B Mol Dev Evol.
Wittwer, C. T., M. G. Herrmann, A. A. Moss and R. P. Rasmussen (1997). "Continuous fluorescence monitoring of rapid cycle DNA amplification." Biotechniques **22**(1): 130-1, 134-8.

Wittwer, C. T., G. H. Reed, C. N. Gundry, J. G. Vandersteen and R. J. Pryor (2003). "High-resolution genotyping by amplicon melting analysis using LCGreen." Clin Chem **49**(6 Pt 1): 853-60.

Wylie, C., M. Kofron, C. Payne, R. Anderson, M. Hosobuchi, E. Joseph and J. Heasman (1996). "Maternal beta-catenin establishes a 'dorsal signal' in early Xenopus embryos." Development **122**(10): 2987-96.

Zearfoss, N. R., A. P. Chan, M. Kloc, L. H. Allen and L. D. Etkin (2003). "Identification of new Xlsirt family members in the Xenopus laevis oocyte." Mech Dev **120**(4): 503-9.

Zearfoss, N. R., A. P. Chan, C. F. Wu, M. Kloc and L. D. Etkin (2004). "Hermes is a localized factor regulating cleavage of vegetal blastomeres in Xenopus laevis." Dev Biol **267**(1): 60-71.

Zhou, Y. and M. L. King (1996). "RNA transport to the vegetal cortex of Xenopus oocytes." Dev Biol **179**(1): 173-83.

Zhou, Y. and M. L. King (2004). "Sending RNAs into the future: RNA localization and germ cell fate." IUBMB Life **56**(1): 19-27.

Paper I

R. Sindelka, Z. Ferjentsik, and J. Jonak, Developmental expression profiles of *Xenopus laevis* reference genes. Dev Dyn, 2006. 235(3): p. 754-8.

DEVELOPMENTAL DYNAMICS 235:754–758, 2006

TECHNIQUES

Developmental Expression Profiles of *Xenopus laevis* Reference Genes

Radek Šindelka, Zoltán Ferjentsik, and Jiří Jonák*

Cell differentiation depends mainly on specific mRNA expression. To quantify the expression of a particular gene, the normalisation with respect to the expression of a reference gene is carried out. This is based on the assumption that the expression of the reference gene is constant during development, in different cells or tissues or after treatment. *Xenopus laevis* studies have frequently used eEF-1 alpha, GAPDH, ODC, L8, and H4 as reference genes. The aim of this work was to examine, by real-time RT-PCR, the expression profiles of the above-mentioned five reference genes during early development of *X. laevis*. It is shown that their expression profiles vary greatly during *X. laevis* development. The developmental changes of mRNA expression can thus significantly compromise the relative mRNA quantification based on these reference genes, when different developmental stages are to be compared. The normalisation against total RNA is recommended instead. *Developmental Dynamics 235:754–758, 2006.* © 2006 Wiley-Liss, Inc.

Key words: *Xenopus*; real-time PCR; normalisation; housekeeping genes

Accepted 14 November 2005

INTRODUCTION

Cell properties depend mainly on the nucleic acid and protein content. The mRNA molecules represent a connection between DNA and protein. The spatial and temporal gene expression changes are a key mechanism in cell differentiation. Two different types of *Xenopus* mRNAs can be distinguished: (1) the maternal mRNAs, which occur in the oocyte before fertilization and originate from the female, and (2) the zygotic mRNAs, newly synthesized mainly after the midblastula transition (MBT). The correct spatial and temporal expression of both types of mRNAs is necessary for the first developmental processes such as the main body axis formation, gastrulation, germ layers induction, and others. The mRNA expression has been broadly studied by common methods such as Northern blotting, RNase protection assays, in situ hybridization, reverse transcription-polymerase chain reaction (RT-PCR), and microarray analysis. In the last decade, a new highly sensitive and specific method for RNA/DNA quantification, the real-time RT-PCR (qPCR), was introduced. The sensitivity, dynamic range, linearity of the measurements, and robustness make the qPCR a method of choice for the quantitative analysis of mRNA expression. For these reasons, it is also frequently used as an independent validation tool for the verification of microarray expression data (Giulietti et al., 2001).

The quantification can be carried out by two different approaches. The absolute method determines the number of mRNA copies in the sample from a calibration curve obtained from samples of cDNA complementary to mRNA of known concentrations. The relative approach compares copies of the target mRNA with those of a reference gene. The metabolic, structural, and ribosomal RNA genes, such as those coding for beta-actin, GAPDH, beta-tubulin, 18S rRNA, and so on, belong among the most popular reference genes (Suzuki et al., 2000). The normalisation against a reference gene requires a constant mRNA expression of the reference gene, which would not vary during the cell cycle, in different cell types, or during development. Moreover, the expression level of the

Department of Gene Expression, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic
Grant sponsor: Grant Agency of the Czech Republic; Grant number: 301/02/0408; Grant sponsor: Academy of Sciences of the Czech Republic; Grant number: AVOZ50521514.
***Correspondence to: Jiří Jonák, Department of Gene Expression, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo nam. 2, Prague 6, 16637 Czech Republic. E-mail: jjon@img.cas.cz**

DOI 10.1002/dvdy.20665
Published online 5 January 2006 in Wiley InterScience (www.interscience.wiley.com).

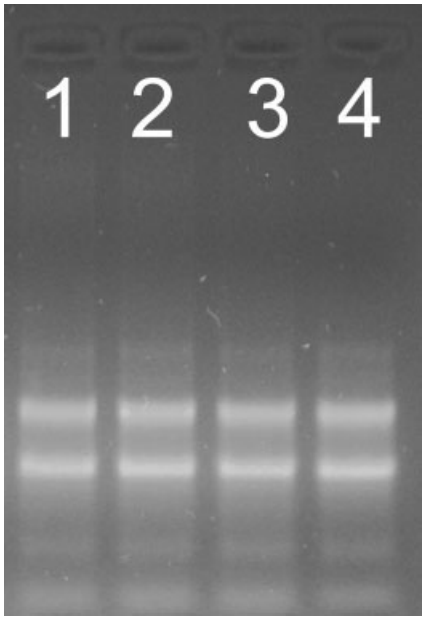


Fig. 1. One percent ethidium bromide agarose gel electrophoresis of 1 μg of total RNA from stage 10.5 (line 1 and line 2) and stage 13 (line 3 and line 4).

reference gene should be similar to that of the target gene. However, several recent studies have shown that, in vertebrate systems, the mRNA expression of reference genes differed among tissues, various cells, and after treatment (Zhang, 2003; Bas et al., 2004; de Kok et al., 2004; Bustin and Nolan, 2004; Vandesompele et al., 2002; Brunner et al., 2004; Radonic et al., 2004). To date, the most reliable method for normalisation appears to relate the mRNA data to the total RNA content of the sample preparation subjected to the reverse transcription reaction (Bustin, 2002). The most important step in the normalisation to total RNA is thus a precise determination of RNA concentration and the quality of RNA preparation (Fig. 1). *Xenopus laevis* mRNA expression studies have almost always used the genes coding for elongation factor eEF-1 alpha, ornithine decarboxylase (ODC), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), L8-ribosomal protein, and H4-histone protein as reference genes (Chang, 2004; Veldholen et al., 2002; Wardle and Smith, 2004). The aim of this study was to determine, by qPCR and normalisation to total RNA content, the mRNA expression profiles of the above-mentioned five reference genes, and, of two other genes, coding for N-tubulin and Xbra, in the course of *X. laevis* early development. The profiles are stage- and gene-specific and allow to determine stages at which the normalisation with the reference genes appears to be plausible and when it would give more or less biased results.

RESULTS

The concentration of total RNA per embryo in the course of early *Xenopus* embryogenesis stays essentially constant up to at least stage 32, and then a slow and continuous increase is observed so that the concentration of RNA at stage 44 is about two-and-half times higher than that at earlier stages (Fig. 2). On the contrary, as shown in Figure 3, the expression levels of all studied genes greatly varied during *X. laevis* development and among genes themselves. As expected, the levels of all tested mRNAs fluctuated less until the MBT stage, after which significant increases in expression and fluctuation were observed at various later stages of the development.

The level of elongation factor eEF-1 alpha mRNA changed only slightly by stage 9. Neither loss nor increase of the mRNA was apparent during this period. However, its concentration began to rise sharply after stage 10.5 and reached the maximum level at stage 26, which was about 80 times higher than at stage 1. This profile correlates nicely with the results that Krieg et al. (1989) obtained by a different method and is consistent with the idea that the MBT stage serves as an activation switch of the transcription of the *X. laevis* embryonic eEF-1 alpha gene.

The GAPDH mRNA kept an approximately constant level up to stage 24 with only slight, about twofold, fluctuations. Then, it rose steeply, about 10 times, and this elevated level was maintained until stage 44, the last stage we measured. The expression profile of L8 mRNA was similar to that of GAPDH except for the decrease at stages 33/34 and 37/38, followed by a steep increase at stage 44. The expression of the ODC mRNA, recently the most frequently used "gene" for normalisation, displayed only slight, about two- to three-fold fluctuations, both up and down, during the whole examined developmental period, except for stages 11.5 and 13 (tenfold decrease), and for stage 44 (tenfold increase).

The expression of the H4 gene shows a different profile when compared with other reference genes. The highest H4 mRNA level was detected in the egg. After stage 4 post-fertiliza-

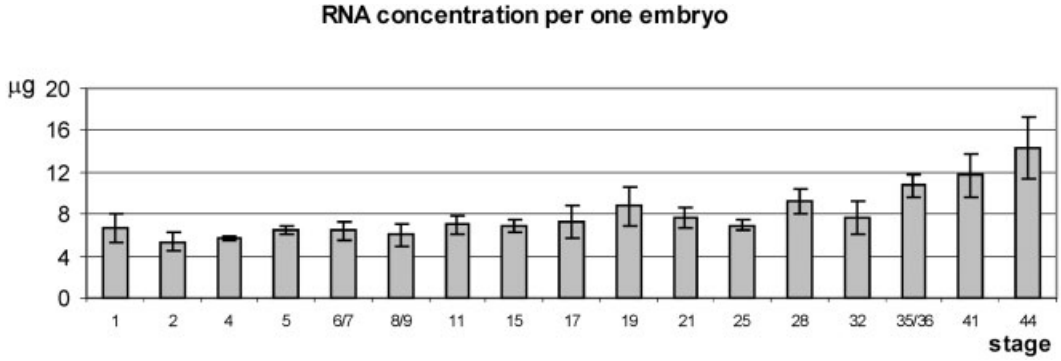


Fig. 2. Total RNA amount in *Xenopus laevis* early developmental stages. RNA was extracted from eggs/embryos and diluted to 50 μl as described in the Experimental Procedures section. Average RNA concentrations (columns) per embryo and standard deviations (bars) from 6 sets of the five eggs/embryos from stages indicated on the horizontal axes are shown.

tion, it gradually, about 5-fold, decreased with a minimum at stage 11. Subsequently, it recovered again, but the H4 mRNA level never exceeded that one detected in the egg.

Two non-reference genes, one coding for Xbra and the other for N-tubulin, were chosen to compare the reliability of our real-time PCR procedure with the already published data (Saka and Smith 2004). The *Xbra* gene belongs among the genes important for early *Xenopus* development. The *Xbra* mRNA expression profile presented here (Fig. 4a) had a similar shape to that described previously for stages 1 to 24 (Saka and Smith 2004). The N-tubulin gene transcription has been known to be closely related to the differentiation of the neural tissue. No expression of N-tubulin mRNA was detected until stage 13. Then, from stage 15, the level of the N-tubulin mRNA continuously rose, peaking around stage 28, the onset of its expression correlating with the onset of neurulation (Fig. 4b). Subsequently, the N-tubulin mRNA levels declined ~4 fold by stage 44. qPCR reactions with mRNA of stage 15 and later stages generated only one specific product with the right sequence and size as confirmed by sequencing and melting curve analyses. These two results prove the reliability of our real-time PCR procedure.

DISCUSSION

Various reference genes in various vertebrate organisms have been used to normalise mRNA expression in their tissues. In this work, the developmental profiles of mRNA expression of *X. laevis* reference genes most frequently used for relative mRNA quantification were determined. The results show fluctuations of at least an order of magnitude in the expression of the reference genes in the course of *X. laevis* early development, and different genes show different profiles. This restricts their use as reference genes to only particular developmental stages depending on the type of the reference gene. None of these genes is particularly well-suited as a reference gene throughout the developmental period most frequently studied in *Xenopus* experiments. The ODC mRNA level appears to be suitable as a refer-

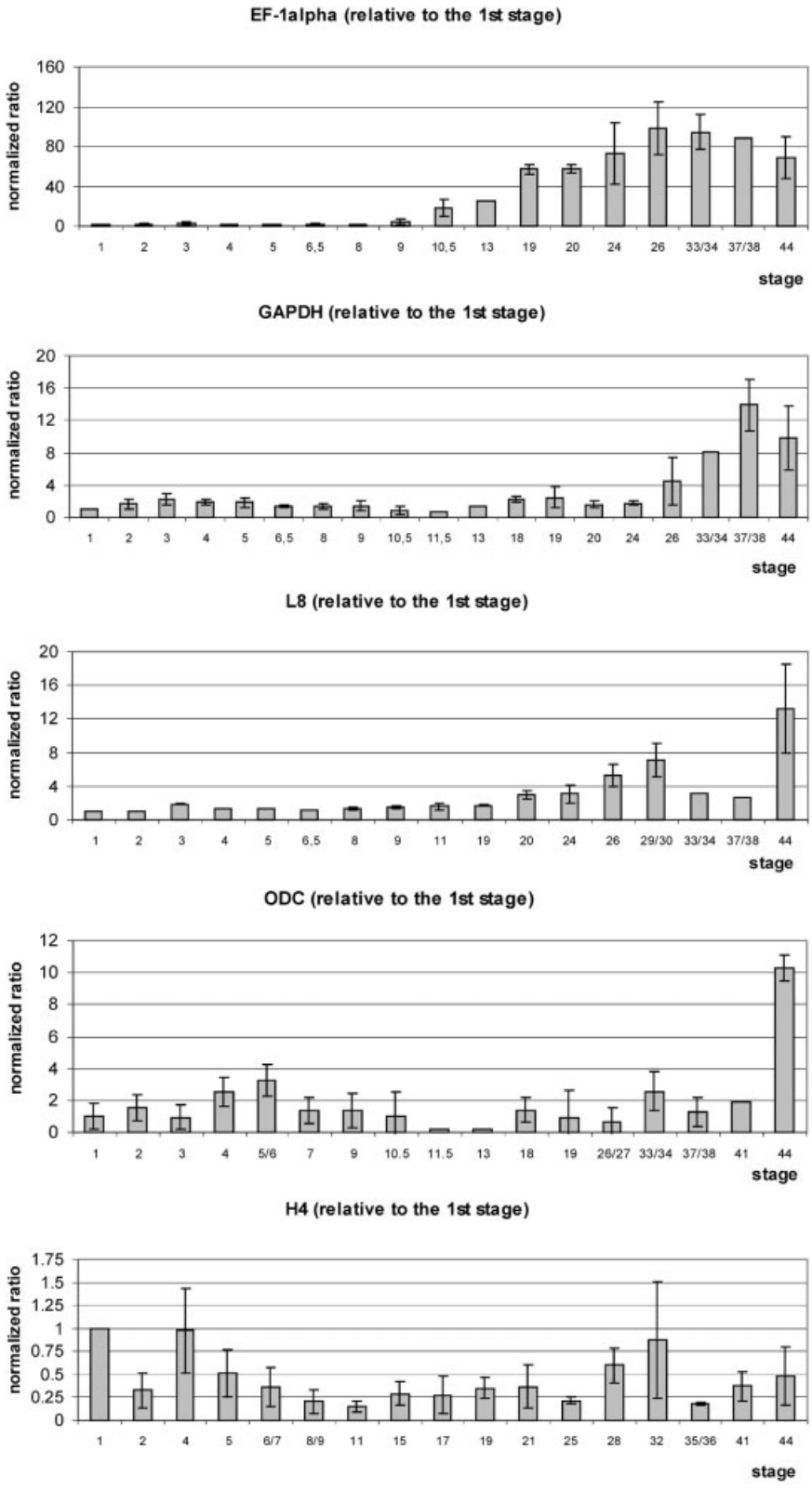


Fig. 3. The mRNA expression profiles of reference genes coding for (**a**) eEF-1 alpha, (**b**) GAPDH, (**c**) L8, (**d**) ODC, (**e**) H4 normalised to total RNA at stage 1 and expressed in arbitrary units. The numbers on the vertical axis represent the ratio between the average amount of copies of mRNA in a particular developmental stage and stage one normalised to the same amount of input RNA. For details see the Experimental Procedures section.

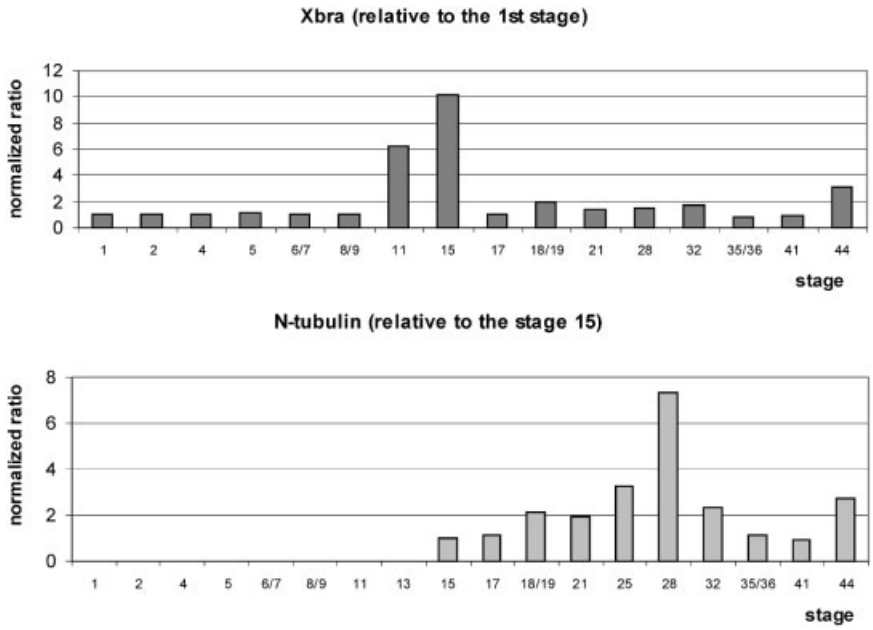


Fig. 4. The mRNA expression profiles of a non-reference gene *Xbra* (**a**) and a gene coding for N-tubulin (**b**) normalised to total RNA at stage 1 (*Xbra*) or stage 15 (N-tubulin) and expressed in arbitrary units. The expression of gene *Xbra* begins at stage 1 and the expression of N-tubulin is linked with neurogenesis and does not begin before the stage 15. The numbers on the vertical axis represent the ratio between the average amount of copies of mRNA in a particular developmental stage and stage one (*Xbra*) or stage 15 (N-tubulin) normalised to the same amount of input RNA. For details see the Experimental Procedures section.

ence throughout stages 1–10.5 and 18–41. The H4 mRNA level displays the smallest fluctuations over the entire developmental period; about fivefold change was detected throughout stages 1 to 44. The normalisation using the L8 and GAPDH mRNA appears to be suitable up to stage 19 whereas the normalisation using the eEF-1 alfa mRNA level should not exceed stage 9. Thus, the normalisation against total RNA according to Bustin (2002) seems to represent for *X. laevis* mRNA measurements a more practicable, reliable, convenient, and versatile method of choice than the normalisation against any one of the five reference genes described above. To find a true reference gene with constant and stable transcription in each tissue type and cell would be extremely useful. However, to date, such a gene has not been identified. As shown here, normalisation against any one of the examined five reference genes should be used only for the same *Xenopus* developmental stage or for a few carefully selected stages. The differential and stage-related expression of the reference genes should be taken into consideration in all relative mRNA quantification assays.

EXPERIMENTAL PROCEDURES

RNA Extraction and cDNA Preparation

X. laevis embryos were staged according to Nieuwkoop and Faber (1967). Two sets of 5 embryos from stages indicated on the horizontal axis in Figures 2–4 were collected from three independent in vitro fertilizations (three different females) and frozen immediately (Fig. 5). RNA from each sample was extracted using Trizol reagent (Invitrogen) according to instructions of the manufacturer and its concentration was determined by UV spectrophotometry at A_{260}. Each sample was measured three times and the average value was determined (Fig. 2). The quality of total RNA was analysed by 1% ethidium bromide agarose gel electrophoresis (Fig. 1).

cDNA was synthesized using 1 μg of total RNA and 10 pmol 25-dT oligo, the mixture was incubated at 72°C for 10 min, then 100 U MMLV reverse transcriptase (Promega), 12 U RNasin (Promega), 5 nmol dNTP were added to a total volume of 10 μl and incubation continued at 37°C for 70 min. The reactions were subsequently diluted to 50 μl and frozen.

Real-Time RT-PCR

Sequences of primers for *X. laevis* elongation factor eEF-1alpha, GAPDH, and N-tubulin cDNA amplification were designed according to XMMR (http://www.xenbase.org/WWW/Marker_pages/primers.html). Primers for ODC cDNA amplification were the same as designed by Heasman et al. (2000), primers for *Xbra* cDNA amplification were identical with those designed by Sun (1999), and primers for L8 and H4 cDNA amplification were designed by using the Beacon Designer 2.00 program (Premier Biosoft International). L8- Forward primer: TCCGTGGTGTGGCTATGAATCC; Reverse primer: GACGACCAGTACGACGAGCAG, H4- Forward primer: GACGCTGTCACCTACACCGAG; Reverse primer: CGCCGAAGCCGTAGAGAGTG. The experiments were performed according to the scheme given in Figure 5 (Stahlberg et al., 2004). The real-time RT-PCR mixture, in the final volume of 25 μl, contained 2 μl of cDNA, 50-fold diluted SYBRGreen solution (Molecular Probes), 0.5 mM forward and reverse primer and 1U Taq polymerase (Promega). The reactions were measured in iCycler (Bio-Rad) with cycling conditions: 95°C for 5 min, 40 cycles at 95°C for 15 sec, and 60°C for 60 sec. Serially diluted PCR fragments (standards), identical with those amplified in the real-time PCR experiment, were prepared to obtain calibration curves. Reaction efficiencies determined from calibration curves for each set of primers were between 85 and 100%. The Cts (threshold cycles) of the samples and standards were analyzed with Microsoft Excel program and copies of PCR products from particular stages of development were determined from calibration curves. The average deviation between Cts in parallel experiments did not exceed about 4.3% for all tested genes and stages. The expression profiles were derived from three independent *X. laevis* serial experiments. Specificity of every amplification reaction was verified by melting curve analysis and gel electrophoresis. The numbers on the vertical axis in Figures 3 and 4 represent the ratio between the average amount of copies of mRNA in a particular developmental stage and stage one normalised to the same amount of input RNA. The concen-

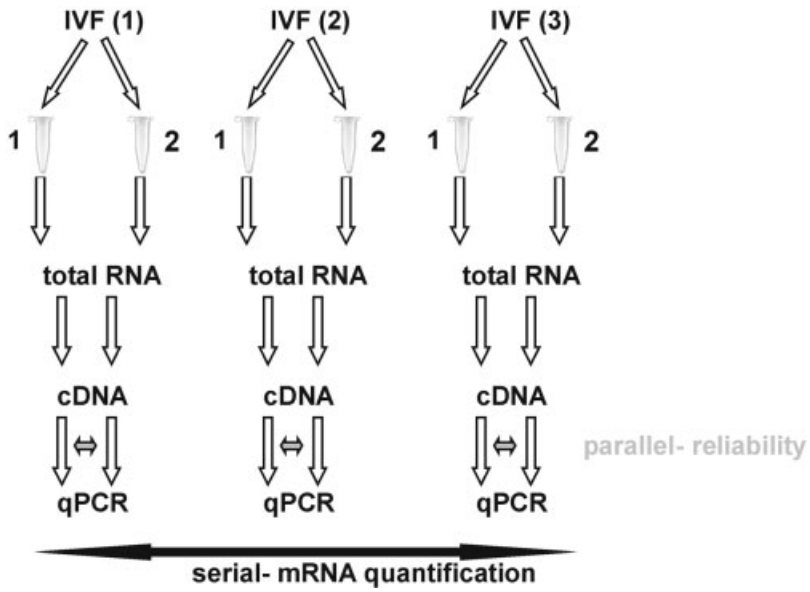


Fig. 5. A scheme of experiment: two samples (5 embryos) from three independent IVFs (in vitro fertilizations) were collected. The parallel samples served for the evaluation of the reproducibility of the procedure from the stage of RNA isolation to the qPCR measurement. The serial analysis shows variations between different IVF.

tration of every mRNA in stage one samples is arbitrarily taken as equal to 1. Standard deviation between results from serial experiments was determined and it is shown by error bars in Figure 3.

ACKNOWLEDGMENTS

This work was supported in part by grant No. 301/020408 from the Grant Agency of the Czech Republic (to J. Jonák) and by grant No. AVOZ50521514 from the Academy of Sciences of the Czech Republic.

REFERENCES

Bas A, Forsberg G, Hammarström S, Hammarström ML. 2004. Utility of the housekeeping genes 18S rRNA, beta-actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes. Scand J Immunol 59:566–573.

Brunner AM, Yakovlev IA, Strauss SH 2004. Validating internal controls for quantitative plant gene expression studies. BMC Plant Biol 4:1–7.

Bustin SA. 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J Mol Endocrinol 29:23–39.

Bustin SA, Nolan T. 2004. Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. J Biomol Tech 15:155–166.

Chang MM. 2004. Spatial and temporal expression patterns of Xenopus Nkx-2.3 gene in skin epidermis during metamorphosis. Gene Express Patterns 5:129–134.

de Kok JB, Roelofs RW, Giesendorf BA, Pennings JL, Waas ET, Feuth T, Swinkels DW, Span PN. 2004. Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. Lab Invest 85: 154–159.

Giulietti A, Overbergh L, Valckx D, Decallonne B, Bouillon R, Mathieu C. 2001. An overview of real-time quantitative PCR: applications to quantify cytokine gene expression. Methods 25:386–401.

Heasman J, Kofron M, Wylie C. 2000. Beta-catenin signaling activity dissected in the early Xenopus embryo: a novel antisense approach. Dev Biol 222:124–134.

Krieg PA, Varnum SM, Wormington WM, Melton DA. 1989. The mRNA encoding elongation factor 1-alpha (EF-1 alpha) is a major transcript at the midblastula transition in Xenopus. Dev Biol 133:93–100.

Nieuwkoop PD, Faber J. 1994. Normal table of *Xenopus laevis*. New York: Garland Publishing, Inc.

Radonic A, Thulke S, Mackay IM, Landt O, Siegert W, Nitsche A. 2004. Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun 313:856–862.

Saka Y, Smith JC. 2004. A Xenopus tribbles orthologue is required for the progression of mitosis and for development of the nervous system. Dev Biol 273:210–225.

Stahlberg A, Kubista M, Pfaffl M. 2004. Comparison of reverse transcriptases in gene expression analysis. Clin Chem 50: 1678–1680.

Sun BI, Bush SM, Collins-Racie LA, LaVallie ER, DiBlasio-Smith EA, Wolfman NM, McCoy JM, Sive HL. 1999. Derriere: a TGF-beta family member required for posterior development in Xenopus. Development 126:1467–1482.

Suzuki T, Higgins PJ, Crawford DR. 2000. Control selection for RNA quantitation. Biotechniques 29:332–337.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:1–12.

Veldholen N, Crump D, Werry K, Helbing CC. 2002. Distinctive gene profiles occur at key points during natural metamorphosis in the *Xenopus laevis* tadpole tail. Dev Dyn 225:457–468.

Wardle FC, Smith JC. 2004. Refinement of gene expression patterns in the early Xenopus embryo. Development 131:4687-4696.

Zhang QJ, Chadderton A, Clark RL, Augustine-Rauch KA. 2003. Selection of normalizer genes in conducting relative gene expression analysis of embryos. Birth Defects Res A Clin Mol Teratol 67:533–544.

Paper II

M. Kubista, J. M. Andrade, M. Bengtsson , A. Forootan, J. Jonák, K. Lind, R. Sindelka, R. Sjöback, B. Sjögreen, L. Strömbom, A.Ståhlberg & N. Zoric. The Real-time Polymerase Chain Reaction. Molecular Aspects of Medicine, 2006. 27, 95-125.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Aspects of Medicine 27 (2006) 95–125
MOLECULAR ASPECTS OF MEDICINE
www.elsevier.com/locate/mam

Review

The real-time polymerase chain reaction

Mikael Kubista [a,*], José Manuel Andrade [b], Martin Bengtsson [a,c], Amin Forootan [d], Jiri Jonák [e], Kristina Lind [a,f], Radek Sindelka [e], Robert Sjöback [a], Björn Sjögreen [d], Linda Strömbom [a], Anders Ståhlberg [a,g], Neven Zoric [a]

[a] *TATAA Biocenter, Medicinargatan 7B, 405 30 Göteborg, Sweden*
[b] *Department of Analytical Chemistry, University of A Coruna, A Zapateira s/n, E-15071 A Coruna, Spain*
[c] *Department of Experimental Medical Science, Lund University, 221 84 Lund, Sweden*
[d] *MultiD Analyses AB, Lotsgatan 5A, 414 58 Gothenburg, Sweden*
[e] *Laboratory of Gene Expression, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo n. 2, 16637 Prague 6, Czech Republic*
[f] *Department of Chemistry and Biosciences, Chalmers University of Technology, Sweden*
[g] *Stem Cell Center, Lund University, BMC, B10, Klinikgatan 26, 22184 Lund, Sweden*

Abstract

The scientific, medical, and diagnostic communities have been presented the most powerful tool for quantitative nucleic acids analysis: real-time PCR [Bustin, S.A., 2004. A–Z of Quantitative PCR. IUL Press, San Diego, CA]. This new technique is a refinement of the original Polymerase Chain Reaction (PCR) developed by Kary Mullis and coworkers in the mid 80:ies [Saiki, R.K., et al., 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, Science 230, 1350], for which Kary Mullis was awarded the 1993 year's Nobel prize in Chemistry. By

* Corresponding author. Tel.: +46 31 7733926; fax: +46 31 7733910.
E-mail address: mikael.kubista@tataa.com (M. Kubista).

0098-2997/$ - see front matter
doi:10.1016/j.mam.2005.12.007

PCR essentially any nucleic acid sequence present in a complex sample can be amplified in a cyclic process to generate a large number of identical copies that can readily be analyzed. This made it possible, for example, to manipulate DNA for cloning purposes, genetic engineering, and sequencing. But as an analytical technique the original PCR method had some serious limitations. By first amplifying the DNA sequence and then analyzing the product, quantification was exceedingly difficult since the PCR gave rise to essentially the same amount of product independently of the initial amount of DNA template molecules that were present. This limitation was resolved in 1992 by the development of real-time PCR by Higuchi et al. [Higuchi, R., Dollinger, G., Walsh, P.S., Griffith, R., 1992. Simultaneous amplification and detection of specific DNA-sequences. Bio-Technology 10(4), 413–417]. In real-time PCR the amount of product formed is monitored during the course of the reaction by monitoring the fluorescence of dyes or probes introduced into the reaction that is proportional to the amount of product formed, and the number of amplification cycles required to obtain a particular amount of DNA molecules is registered. Assuming a certain amplification efficiency, which typically is close to a doubling of the number of molecules per amplification cycle, it is possible to calculate the number of DNA molecules of the amplified sequence that were initially present in the sample. With the highly efficient detection chemistries, sensitive instrumentation, and optimized assays that are available today the number of DNA molecules of a particular sequence in a complex sample can be determined with unprecedented accuracy and sensitivity sufficient to detect a single molecule. Typical uses of real-time PCR include pathogen detection, gene expression analysis, single nucleotide polymorphism (SNP) analysis, analysis of chromosome aberrations, and most recently also protein detection by real-time immuno PCR.

Keywords: Real-time PCR; Gene expression profiling; GenEx; Principal component analysis; PCA; Multidimensional expression profiling

Contents

1. Real-time PCR basics . . . 97
   1.1. PCR amplification . . . 97
   1.2. Real-time monitoring of PCR . . . 98
   1.3. Fluorescence reporters . . . 102
   1.4. Real-time PCR instruments . . . 106
2. Gene expression measurements . . . 107
   2.1. The relative gene expression ratio . . . 110
   2.2. Real-time PCR expression profiling . . . 112
   2.3. Principal component analysis . . . 113
   2.4. Data pre-treatment . . . 114
   2.5. Embryonic development in *X. laevis* . . . 115
   2.6. Single cell gene expression profiling . . . 118
   Acknowledgement . . . 122
   References . . . 122

1. Real-time PCR basics

1.1. PCR amplification

PCR is performed on a DNA template, which can be single or double-stranded. Also needed are two oligonucleotide primers that flank the DNA sequence to be amplified, dNTPs, which are the four nucleotide triphosphates, a heat-stable polymerase, and magnesium ions in the buffer. The reaction is performed by temperature cycling. High temperature is applied to separate (melt) the strands of the double helical DNA, then temperature is lowered to let primers anneal to the template, and finally the temperature is set around 72 °C, which is optimum for the polymerase that extends the primers by incorporating the dNTPs (Fig. 1).

The melting shall be sufficient to fully separate the strands of the template, because partially strand separated structures will rapidly reanneal when temperature is dropped and will not be primed. The required melting temperature and duration of melting depends on the length and sometimes the sequence of the template and also on the instrument and the reaction containers used. With short amplified sequences, also called amplicons, and containers with rapid heat transfer, such as glass capillaries, it may be sufficient to touch 95 °C and then immediately start cooling. It may be necessary to heat more thoroughly during the first few cycles, because the original template is typically much longer than the amplicons that dominate later. Also, some hot-start polymerases require extensive heat activation.

The annealing temperature depends on the primers. Theoretically it shall be a few degrees below the melting temperature of the two primers (which shall be designed to have the same or similar melting temperatures), in order for them to form stable complexes with the targeted sequences but not with any other sequences. Several free and commercial softwares are available to estimate primer melting temperatures (http://frodo.wi.mit.edu/..cgi-bin/primer3/primer3_www.cgi, http://www.premierbiosoft.com/netprimer). These, however, do not account for the stabilizing effect of the

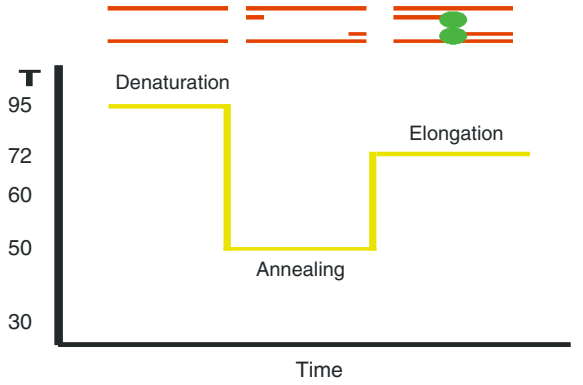


Fig. 1. The PCR temperature cycle: (1) the temperature is raised to about 95 °C to melt the double stranded DNA, (2) the temperature is lowered to let primers anneal, (3) the temperature is set to 72 °C to let the polymerase extend the primers.

polymerase, which binds the annealed primer and stabilizes the complex. In fact, the mechanism is probably that the polymerase binds the primer first and then the polymerase-primer complex binds target DNA.

Optimum temperature for Taq polymerase is about 72 °C, which is the elongation temperature used in most three-step PCR protocols. But it does not seem very important, and some protocols, particularly those based on Taqman probes (see below) elongate at 60 °C (Holland et al., 1991). Using elevated elongation temperature is probably more important to melt any secondary structures that may form in the template and may block extension. In real-time PCR amplicons are typically short with limited capability to fold. But sequential runs of guanines, even only 2–3 consecutive guanines, may fold the template into a tetraplex structure, which is exceedingly stable and cannot be transcribed by the polymerase (Simonson et al., 1998). Guanine tetraplexes form exclusively in the presence of K^+ ions (Simonsson, 2001), and the problem can be avoided by using K^+ free PCR buffer. Self complementary regions in the template can also cause problems by folding into hair-pins and other structures that may interfere with the extension. The same sequence features cause problems also if present in the primers. Complementarity between the primers, particularly in their 3′-ends, causes complications by forming aberrant PCR products called primer–dimers. Avoiding the formation of primer–dimer products is very important for quantitative PCR analysis of samples that contain only few target molecules because the PCR of the target and the PCR forming primer–dimers compete.

1.2. Real-time monitoring of PCR

Real-time PCR also needs a fluorescent reporter that binds to the product formed and reports its presence by fluorescence (Fig. 2). A number of probes and dyes are available and they are described in the next chapter. For now it is sufficient to know that the reporter generates a fluorescence signal that reflects the amount of product formed. During the initial cycles the signal is weak and cannot be distinguished from the background (Fig. 3). As the amount of product accumulates a signal develops that initially increases exponentially. Thereafter the signal levels off and saturates. The signal saturation is due to the reaction running out of some critical component. This can be the primers, the reporter, or the dNTPs (Kubista et al., 2001). Also the number of polymerase molecules may be limiting, in which case the exponential amplification goes over to linear amplification. It is worth noting that in a typical real-time PCR experiment all response curves saturate at the same level. Hence, end-point PCR measurements tell us nothing about the initial amounts of target molecules that were present in the samples; they only distinguish a positive from a negative sample. On the other hand the response curves are separated in the growth phase of the reaction. This reflects the difference in their initial amounts of template molecules. The difference is quantified by comparing the number of amplification cycles required for the samples' response curves to reach a particular threshold fluorescence signal level. The number of cycles required to reach threshold is called the CT value. The amplification response curves are expected to be parallel in the growth

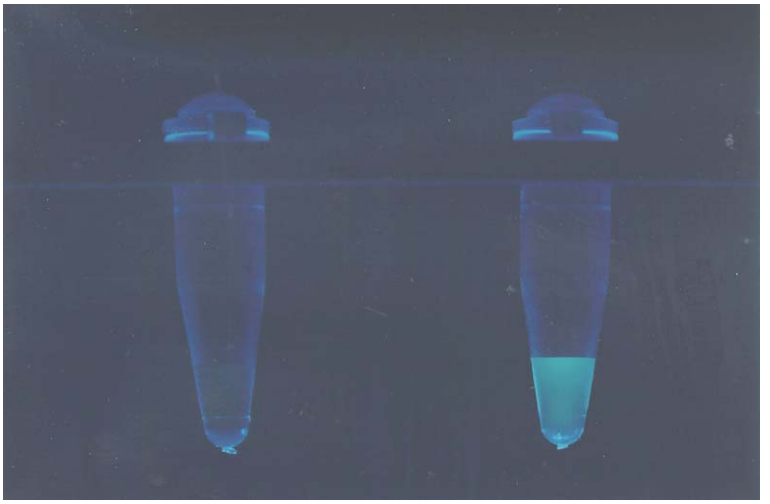


Fig. 2. Fluorescence from hybridized probe. Both PCR tubes contain DNA and probe. In the left tube the probe and the DNA are not complementary and the probe does not bind, while in the right tube the probe and the DNA are complementary.

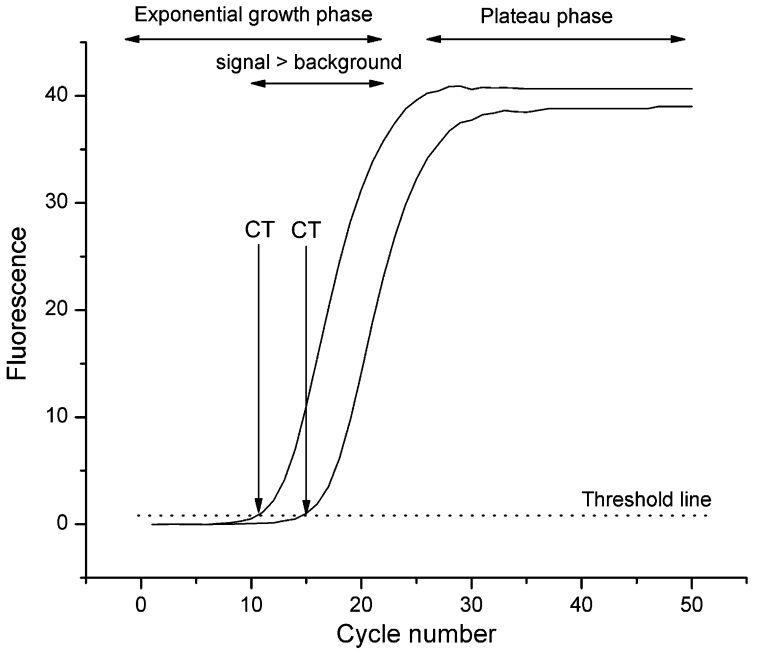


Fig. 3. Real-time PCR response curves. A threshold level is set sufficiently above background and the number of cycles required to reach threshold, CT, are registered.

phase of the reaction, and the setting of the threshold level should therefore not be critical. Different instrument softwares use different methods and algorithms to select the threshold, and most also let the user set it manually. The setting is therefore somewhat arbitrary and it does not affect significantly the differences between CT values, though it affects the values of the individual CTs. These are also affected by the setting of the instrument (filter, channel, gain, etc.). Hence, one should avoid comparing individual CT values between experiments, and include one reference per run to which all the other response curves can be related.

Assuming that the PCR is 100% efficient the ratio between the initial numbers of template copies in two samples is given by

$$\frac{[N_0]_A}{[N_0]_B} = 2^{(CT_B - CT_A)} \tag{1}$$

$[N_0]_A$ and $[N_0]_B$ are the initial numbers of template molecules in samples A and B, and CT_A and CT_B are the corresponding CT values. Suppose the response of sample A appears four cycles later than the response of sample B, i.e., four additional PCR cycles were needed to reach the same threshold level, sample A should initially have contained $2 \times 2 \times 2 \times 2 = 16$ times less template molecules than sample B. Note that the signal for the sample that initially contained less molecules requires higher number of amplification cycles and, hence, develops later.

If the PCR is not perfect, which it almost never is, the efficiency of the PCR enters the equation as

$$\frac{[N_0]_A}{[N_0]_B} = (1 + E)^{(CT_B - CT_A)} \tag{2}$$

Let say that the PCR is 90% efficient, which is quite typical when using biological samples. Four cycles difference between the two amplification curves then reflects a ratio of $(1 + 0.9)^4 = 13$ between the initial numbers of template copies in samples B and A. 16 and 13 are quite different estimates, emphasizing the importance of estimating the PCR efficiency and taking it into account.

The efficiency of a PCR assay can be estimated from a standard curve based on serial dilution of a standard, which can be a purified PCR product or a purified plasmid that contains the target sequence (Fig. 4) (Rutledge and Cote, 2003). The CT values of the diluted standards are read out, and plotted versus the logarithm of the samples' concentrations, number of template copies or dilution factor. The data are fitted to the equation:

$$CT = k \times \log(N_0) + CT(1) \tag{3}$$

The PCR efficiency is calculated from the slope as:

$$E = 10^{-\frac{1}{k}} - 1 \tag{4}$$

The intercept of the standard curve corresponds to the CT(1) of a diluted standard containing only a single target molecule.

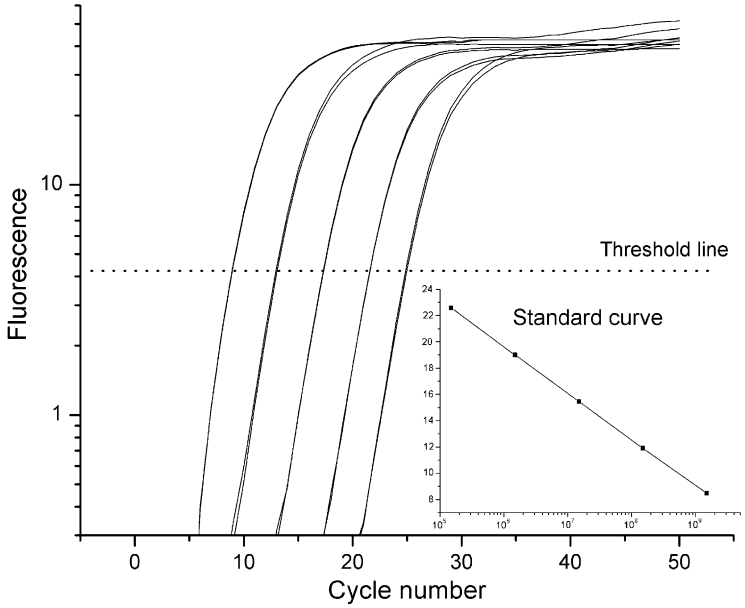


Fig. 4. Real-time PCR standard curve. Real-time PCR response curves shown in logarithmic scale for five standard samples. The crossing points with threshold line are the CT values. In the inset the CT values are plotted vs. the logarithm of the initial number of template copies in the standard samples.

The standard dilution series (briefly, standard curve) approach gives a good estimate of the efficiency of the PCR assay. But, most importantly, it does not tell us anything about the effect of the matrix of the real test sample. Biological samples are complex and may contain inhibitory substances that are not present in standards based on purified template, and this may reduce the PCR efficiency. Examples of common PCR inhibitors are heme, heparin, IgG, and lipids (Akane et al., 1994; Izraeli et al., 1991; Al-Soud et al., 2000). If there is enough sample it may be purified extensively and then diluted, which reduces inhibition. But some inhibitors are hard to remove by dilution. In such cases one may estimate the PCR efficiency of the test sample by either serial dilution of the sample or by the method of standard additions (Ståhlberg et al., 2003).

Some authors suggest estimating PCR efficiencies from the real-time PCR response curves. The idea is indeed attractive because it would allow more precise determination, and in an extension also absolute determination, of the number of target molecules in a sample (Rutledge, 2004; Van et al., 2005). Particularly when using dyes the rise of the amplification response curve is exceedingly difficult to model, because the amount of dye bound per amplicon changes when the DNA concentration increases, and the fluorescence of the bound dye depends on the binding ratio. But recently some important advances have been made and reliable corrections for PCR efficiency from the amplification curves may become reality sometime in the future (Rutledge, 2005).

1.3. Fluorescence reporters

Today fluorescence is exclusively used as the detection method in real-time PCR. Both sequence specific probes and non-specific labels are available as reporters. In his initial work Higuchi used the common nucleic acid stain ethidium bromide, which becomes fluorescent upon intercalating into DNA (Higuchi et al., 1992). Classical intercalators, however, interfere with the polymerase reaction, and asymmetric cyanine dyes such as SYBR Green I and BEBO have become more popular (Fig. 5) (Zipper et al., 2004; Bengtsson et al., 2003). Asymmetric cyanines have two aromatic systems containing nitrogen, one of which is positively charged, connected by a methine bridge. These dyes have virtually no fluorescence when they are free in solution due to vibrations engaging both aromatic systems, which convert electronic excitation energy into heat that dissipates to the surrounding solvent. On the other hand the dyes become brightly fluorescent when they bind to DNA, presumably to the minor groove, and rotation around the methine bond is restricted (Nygren et al., 1998). In PCR the fluorescence of these dyes increases with the amount of double stranded product formed, though not strictly in proportion because the dye fluorescence depends on the dye:base binding ratio, which decreases during the course of the reaction. The dye fluorescence depends also to some degree on the DNA sequence. But a certain amount of a particular double-stranded DNA target, in the absence of significant amounts of other double-stranded DNAs, gives rise to the same fluorescence every time. Hence, the dyes are excellent for quantitative real-time PCR when samples are compared at the same level of fluorescence in absence of interfering DNA. Although minor groove binding dyes show preference for runs of AT base-pairs (Jansen et al., 1993), asymmetric cyanines are considered sequence non-specific reporters in real-time PCR. They give rise to fluorescence signal in the presence of any double stranded DNA including undesired primer–dimer products. Primer–dimer formation interferes with the formation of specific products because of competition of the two reactions for reagents and may lead to erroneous readouts. It is therefore good practice to control for primer–dimer formation. This can be done by melting curve analysis after completing the PCR. The temperature is then gradually increased and the fluorescence is measured as function of temper-

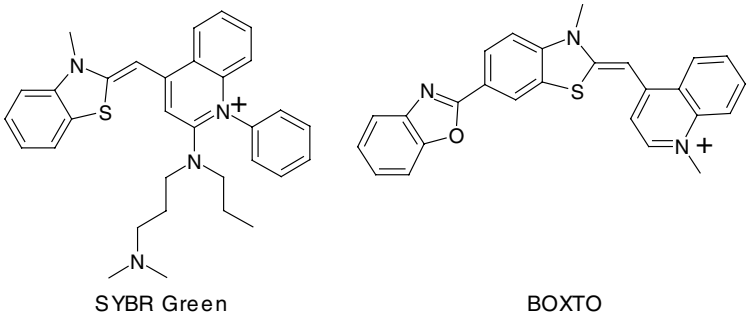


Fig. 5. The asymmetric cyanine dyes SYBR Green and BOXTO.

ature. The fluorescence decreases gradually with increasing temperature because of increased thermal motion which allows for more internal rotation in the bound dye (Nygren et al., 1998). However, when the temperature is reached at which the double stranded DNA strand separates the dye comes off and the fluorescence drops abruptly (Fig. 6) (Ririe et al., 1997). This temperature, referred to as the melting temperature, T_m, is easiest determined as the maximum of the negative first derivative of the melting curve. Since primer–dimer products typically are shorter than the targeted product, they melt at a lower temperature and their presence is easily recognized by melting curve analysis.

Labeled primers and probes are based on nucleic acids or some of their synthetic analogues such as the peptide nucleic acids (PNA) (Egholm et al., 1992) and the locked nucleic acids (LNA) (Costa et al., 2004). The dye labels are of two kinds: (i) fluorophores with intrinsically strong fluorescence, such as fluorescein and rhodamine derivatives (Sjöback et al., 1995), which through structural design are brought into contact with a quencher molecule, and (ii) fluorophores that change their fluorescence properties upon binding nucleic acids (Fig. 7). Examples of probes with two dyes are the hydrolysis probes, popularly called Taqman probes (Holland et al., 1991), which can be based either on regular oligonucleotides or on LNA (Braasch and Corey, 2001), Molecular Beacons (Tyagi and Kramer, 1996; Tyagi et al., 1998), Hybridization probes (Caplin et al., 1999), and the Lion probes (http://www.biotools.net). The dyes form a donor–acceptor pair, where the donor dye is excited and transfers its energy to the acceptor molecule if it is in proximity. Originally the acceptor molecule was also a dye, but today quencher molecules are more

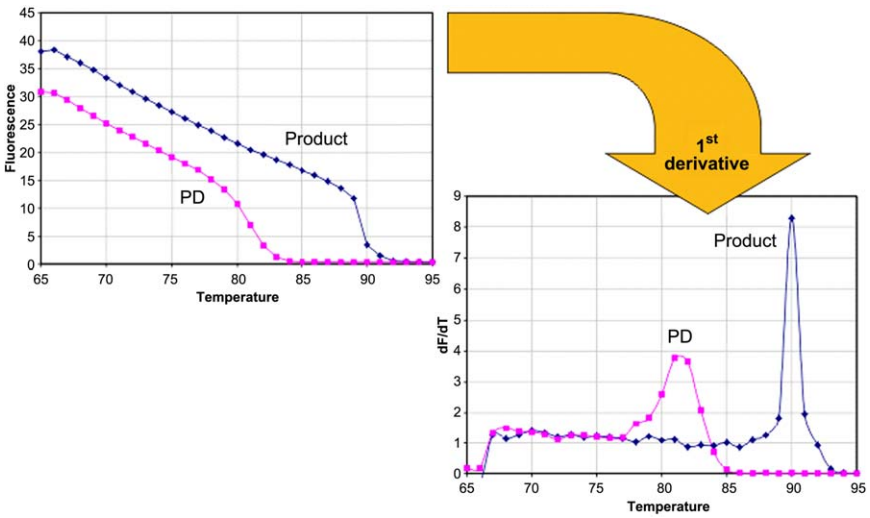


Fig. 6. Melting curve analysis. Dye fluorescence drops rapidly when the DNA melts. The melting point is defined as the inflection point of the melting curve, which is easiest determined as the maximum in the negative 1st derivative of the melting curve. The amplicon produced from the targeted product is typically longer and melts at higher temperature than the primer–dimers.

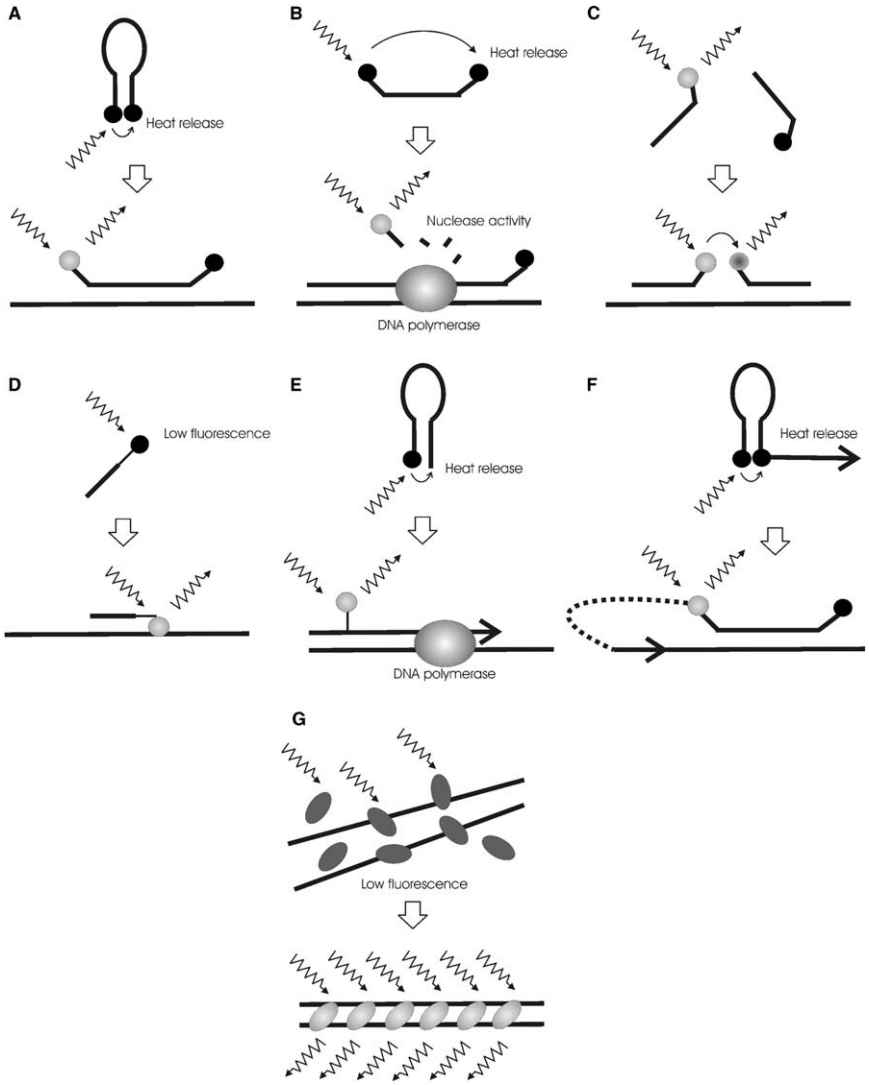


Fig. 7. Mechanisms of reporters used in real-time PCR: (A) the molecular beacon, (B) the Taqman probe, (C) the hybridization probes, (D) the LightUp probe, (E) the simple probe, (F) scorpion primer, (G) sequence non-specific dyes (SYBR Green/BOXTO).

popular (Wilson and Johansson, 2003). Energy transfer and quenching are distance dependent and structural rearrangement of the probe, or, in the case of hydrolysis probes, degradation, change the distance between the donor and acceptor and, hence, the fluorescence of the system.

Probes based on a single dye, whose fluorescence changes upon binding target DNA include the LightUp probes (Svanvik et al., 2000), AllGlo probes (http://

www.allelogic.com), Displacement probes (Li et al., 2002), and the Simple probes (http://www.idahotech.com/itbiochem/simpleprobes.html). Chemical modifications and alterations of the oligonucleotide backbone are employed in some probes to improve the binding properties to the target template. This makes it possible to use shorter probes, which is advantageous for the detection of targets with short conserved regions such as retroviruses. LightUp probes have a neutral peptide nucleic acid (PNA) backbone that binds to DNA with greater affinity than normal oligonucleotides. The LightUp probes are 10–12 bases, which is short compared to normal oligonucleotide probes that are usually at least 25 bases (http://www.lightup.se). LNA-probes make use of modified nucleotides to enhance binding affinity. MGB-probes are hydrolysis probes with a minor groove binding molecule attached to the end of the probe to increase affinity for DNA, which makes it possible to use shorter probes (Kutyavin et al., 2000). Examples of modified primers include: Scorpion primers (Whitcombe et al., 1999), LUX primers (Nazarenko et al., 2002), Ampliflour primers (Uehara et al., 1999), and the QZyme system (BD QZyme™ Assays for Quantitative PCR, 2003).

As long as a single target is detected per sample there is not much of a difference in using a dye or a probe. Assay specificity is in both cases determined by the primers. Probes do not detect primer–dimer products, but using non-optimized probe assays is hiding the problem under the rug. If primer–dimers form they cause problems whether they are seen or not. In probe based assays, particularly when high CT values are obtained, one should verify the absence of competing primer–dimer products. The traditional way is by gel electrophoresis. Recently, an alternative approach was proposed based on the BOXTO dye. BOXTO is a sequence non-specific double-stranded DNA binding dye that has distinct spectral characteristics to fluorescein and can be used in combination with FAM based probes. The BOXTO and the probe signals are detected in different channels of the real-time PCR instrument. While the probe reflects formation of the targeted product as usual, the BOXTO dye also reports the presence of any competing primer–dimer products, which can be identified by melting curve analysis (Lind et al., in press).

The great advantage of probes is for multiplexing, where several products are amplified in the same tube and detected in parallel (Wittwer et al., 2001). Today multiplexing is mainly used to relate expression of reporter genes to that of an exogenous control gene in diagnostic applications (Mackya, 2004), and for single nucleotide polymorphism (SNP) and mutation detection studies (Mhlanga and Malmberg, 2001; Mattarucchi et al., 2005). Multiplex assays are more difficult to design because when products accumulate the parallel PCR reactions compete for reagents. To minimize competition limiting amounts of primers must be used. Also, primer design is harder, because complementarity must be avoided between all the primers. Multiplex assays can be based either on probes or on labeled primers, where labeled primers usually give rise to signal from primer–dimer products, while probes do not.

The different probing technologies have their advantages and limitations. Dyes are cheaper than probes but they do not distinguish between products. Hairpin forming probes have the highest specificity, because the formation of the hairpin competes with the binding to mismatched targets. This makes them most suitable for

SNP and multi-site variation (MSV) analysis (Bonnet et al., 1999). Hydrolysis probes require two-step PCR to function properly, which is not optimal for the polymerase reaction, and short amplicons are necessary to obtain reasonable amplification efficiencies. But they are easier to design than hairpin forming probes and an 80% success rate was recently reported (Kubista, 2004).

In summary, a 'good' probe, independent of chemistry, should have low background fluorescence, high fluorescence upon target formation (high signal to noise ratio), and high target specificity. The dyes' excitation and emission spectra are important parameters to consider when designing multiplex reactions. Spectral overlap in excitation and emission should be minimized to keep cross-talk to a minimum.

1.4. Real-time PCR instruments

Today many instrument platforms are available for quantitative real-time PCR. The main differences between them are the excitation and emission wavelengths that are available, speed, and the number of reactions that can be run in parallel (Kubista and Zoric, 2004). Reaction containers also differ. Most popular are 96-well microtiter plates, which are becoming standardized and therefore available from multiple vendors. These are used in the Applied Biosystems 7300 and 7500 instruments (http://www.appliedbiosystems.com), the Exicycler from Bioneer (http://www.bioneer.co.kr), the Chromo4, the DNAEngine Opticon, the iCycler, the iQ, the MyiQ, and the iQ5 from BioRad (http://discover.bio-rad.com), the RealPlex from Eppendorf (http://www.eppendorf.com/mastercycler/index.html), the Mx3000p, the Mx3005p, and the Mx4000 from Stratagene (http://www.stratagene.com), and the Quantica from Techne (http://www.techne.com/CatMol/quantica.htm). Currently there are two 384-well plates instruments available on the market: the ABI PRISM 7900HT from Applied Biosystems (http://www.appliedbiosystems.com), and the LightCycler 480 system from Roche (http://www.roche-applied-science.com). For very high throughput we see an interesting development at Biotrove who have a through-hole array platform, called OpenArray™, enabling massively parallel real-time PCR. Passive microfluidics based on surface tension differentials load and retain in isolation 3072 33 nl reaction volumes in a footprint the size of microscope slide. Samples and primer pairs can be loaded into the OpenArray to give the user maximum flexibility in the number of transcripts measured per sample. Three such OpenArray plates can be run simultaneously in BioTrove's real-time PCR instrument enabling the study of up to 64 transcripts in 144 samples, or any combination thereof, equal to 9216 real-time PCR reactions (Brenan and Morrison, 2005) (http://www.biotrove.com).

The Rotor-Gene 3000 (4-Channel) and 6000 (6-Channel) from Corbett Research are based on a rotor platform to achieve the highest temperature uniformity between samples. They use plastic tubes, which are rather inexpensive (http://www.corbettresearch.com). The LightCycler from Roche uses glass capillaries (http://www.roche-applied-science.com). These have excellent optical properties, but are somewhat more expensive. But recently cheaper replica made of plastics has appeared. The Cepheid SmartCycler uses special containers for cycling (http://www.cepheid.com).

The instrument is designed for field testing, and can run several different assays that must not even be started simultaneously in parallel. The InSyte from Biobank has 96 independent wells, seven color multiplexing and very fast cycling by using special conductive plastic containers (http://www.biobank.co.kr/pcr/insyte.shtml). The fastest real-time PCR instrument of them all is the SuperConvector from AlphaHelix that runs at elevated G-forces, which cause turbulent flow and thereby rapid mixing resulting in extremely fast heat-transfer (http://www.alphahelix.com/pages/superconvector.html). For these more sophisticated instruments the higher cost for containers is probably negligible anyway. Then there is the very small DT-322 instrument from DNA Technology (http://www.dna-technology.com/catalog/ob_dt322_en.shtml), and a fiber optics based Line-Gene II from Bioer (http://www.bioer.com.cn/en/shengming_66pcr.htm). Fluidigm develops nanofluidic chips called dynamic arrays for QPCR. The instrument's footprint is W30'' × D30'' × H39''. It has five excitation filters, five emission filters, and a CCD that images the entire chip. Current chips are 48/48 dynamic arrays that yield 2304 real-time PCR reactions each being 10 nl (http://www.fluidigm.com) (A nanofluidic chip for absolute quantification of target nucleic acid sequences, Pharmaceutical Discovery October 1, 2005). Even a notebook real-time PCR instrument has been described (Belgrader et al., 2001). As for performance, sensitivity and accuracy the conventional instruments are virtually equivalent. Also the instrument softwares are pretty much the same at least concerning the basic features, such as setting up the experiment and specifying protocols, and simple pre-processing and processing of data, including base-line subtraction and calculating threshold cycle numbers. The prices for the instruments vary quite substantially, mainly depending on throughput and the number of colors they can handle. Unless you plan to do multiplexing and are not running a core facility, you can save a lot of money and still do excellent real-time PCR on anyone of the less expensive instruments.

2. Gene expression measurements

Before a gene expression measurement can be performed by real-time PCR, the mRNA in the sample must be copied to cDNA by reverse transcription (RT) (Fig. 8). The RT step is critical for sensitive and accurate quantification, since the amounts of cDNAs produced must correctly reflect the input amounts of the mRNAs. Comparing the technical reproducibilities of the RT reaction and the PCR it was found that the RT reaction contributes with most of the variation to the experimental determination of mRNA quantities (Ståhlberg et al., 2004a). Hence, it was concluded that substantially higher experimental accuracy is obtained when performing replicates starting with the RT reaction than when splitting the samples after the RT step to perform real-time PCR replicates only.

In an RT reaction RNA molecules are transcribed to DNA copies by a reverse transcriptase. The reaction can proceed without added primer, but higher efficiency is obtained when primers are added (Ståhlberg et al., 2004a). The three basic priming strategies are based on oligo(dT) primer, random sequence primers, and gene

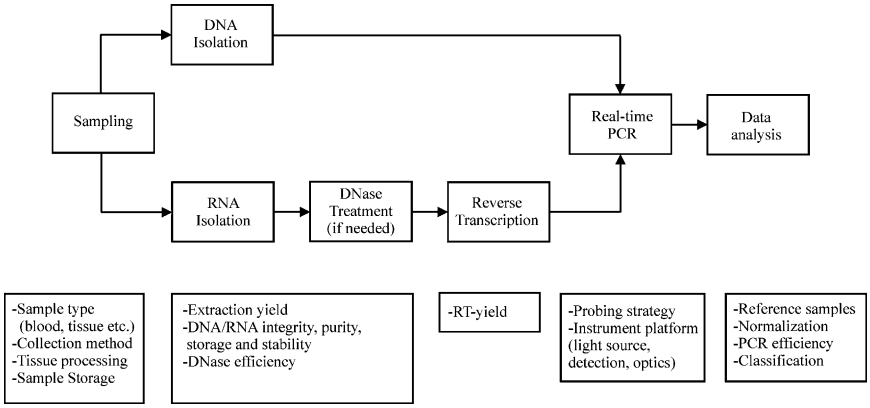


Fig. 8. Scheme of RNA and DNA analysis from biological samples. Sources of variation are indicated.

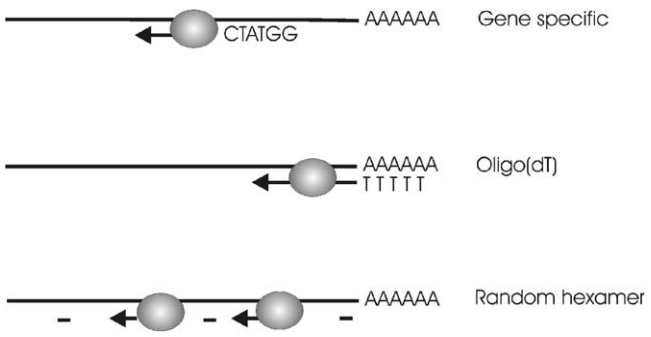


Fig. 9. RT priming strategies.

specific primers (Fig. 9). Oligo(dT) primers hybridize to the poly(A) tail present in most eukaryotic mRNAs and will initiate reverse transcription from the very beginning (3′-end) of the mRNA. This is important advantage if the cDNA shall be cloned. But for expression analysis this may be a disadvantage, because the transcription may not reach the PCR target sequence if the mRNA is not intact because of degradation, and a bias for amplicons located close to the mRNA 3′-end may be introduced. Particularly if the mRNA samples studied are of varying quality this may be a problem. Another limitation is that mRNAs without A-tails, such as histone genes in most eukaryotes, and most prokaryotic genes, are not transcribed. Ribosomal RNA and transfer RNA are not copied either, which precludes using rRNAs as internal standard. A variant are anchored oligo(dT) primers, which are oligothymines with one or two non-thymine bases in the 3′-end. These were initially designed for cloning purposes, but may be useful in diagnostic cDNA synthesis by anchoring the primer to the start of the A-tail. This results in homogeneous transcripts of each mRNA, which some vendors claim improves reproducibility.

One advantage of oligo(dT) primers is that their target sequence is a run of adenines, which does not fold into higher order structures. Oligo(dT) priming is therefore expected to be more efficient and less dependent on temperature than other priming strategies. Random sequence primers are short oligomers of all possible sequences. They are usually six (random hexamers) or nine (random nonamers) bases long. They are produced by random sequence synthesis, by essentially adding a blend of all four nucleotides in each step of the oligonucleotide synthesis. This gives rise to essentially all sequences, although some bias is expected due to variations in coupling efficiencies, where particularly some longer nucleotide runs are synthesized in poor yields. However, this bias is not likely to be important for their use as RT primers. Random sequence primers will copy all RNA, including tRNA, rRNA, and mRNA. It is the priming strategy of choice if rRNA shall be used as reference and for total reverse transcription of prokaryotic mRNA. Random nonamers bind stronger to target than random hexamers and may be preferred in high temperature RT protocols. On the other hand, they require access to larger regions in the target, which requires the mRNA to be more unfolded. Random oligomer priming may give rise to multiple transcripts, and may at least in theory give RT yields exceeding 100%. Transcription is expected to be initiated at more sites on longer RNAs than on shorter ones, and transcription yields will also depend on the RNAs secondary and tertiary structures. For total reverse transcription of RNA of lower quality, such as fixed and archival samples, random oligomers are preferred to oligo(dT) primers, because they are more likely to yield transcripts that extend past the PCR target sequences. For total reverse transcription, when it is important to copy as many of the different mRNAs as possible, one may use a blend of random oligomers and oligo(dT) primers. Third option is to use sequence specific primers. This is the preferred strategy when a limited number of mRNAs shall be analyzed. The hybridization of specific primers to mRNA is highly sequence dependent because of the folding of mRNA to secondary and tertiary structures. In non-purified samples access may also be limited by bound proteins. The primer must target an accessible sequence. What regions in the mRNA are accessible for priming cannot be predicted by inspection of the mRNA sequence only. It depends also on temperature. At elevated temperatures more sequence regions are accessible, which is the reason heat-stabile reverse transcriptases are gaining popularity. Software, such as *m*-fold, can aid in primer design by predicting the folded structures of RNAs (http://molbio.info.nih.gov/molbionih/mfold.html). However, structure prediction is only feasible for a few hundred base-pairs, while native mRNAs are several thousands or tens of thousands of base-pairs. A practical approach is to test a set of primers, and then go with the one that gave highest yields. Designing only one primer and hoping for the best, the yield may not be higher than when using random sequence oligomers or poly(dT) primers (Ståhlberg et al., 2004a). The RT primer may also be used as the reverse primer in the PCR, reducing the total number of primers needed. The RT and real-time PCR reactions can then be combined into a one-step RT-PCR reaction. This is convenient when analyzing only one or a few genes, and it reduces the risk of cross-contamination. However, optimal reaction conditions for the RT and real-time PCR are usually

somewhat different, and one-step reactions tend to be less sensitive than the regular two-step approach (Ståhlberg et al., 2004a; Bustin, 2002).

Several reverse transcriptases are today available commercially. Most are engineered forms of either the Moloney Murine Leukemia Virus (MMLV) or of the Avian Myeloblastosis Virus (AMV). They differ in size and in their temperature optima. Everything else being the same, one expects a smaller enzyme that is active at a higher temperature to give better RT yields because of more efficient priming. In an absolute quantification study based on an artificial gene an average RT yield of 30% was measured, but it was also found that the yield varied more than 100-fold with the enzyme used and the mRNA target (Ståhlberg et al., 2004b; Peters et al., 2004). Also the priming strategy affects the RT efficiency in a gene specific way. This may sound worrying, suggesting it is difficult to obtain comparable RT QPCR data. But this is not the case. The RT reaction is highly reproducible as long as the same experimental protocol and reaction conditions are used, and results are perfectly comparable (Ståhlberg et al., 2004a). Other sources of variation in quantitative expression analysis are sample preparation and RNA extraction. The variation in losses during RNA isolation between samples can be controlled by spiking the samples with a known concentration of exogenous mRNA. Still, this does not fully mimic the situation of native mRNA molecules that are localized in cell compartments, and we have no idea how the isolation yield varies among different mRNAs. Differences in length, folding, localization in the cell, and complexation to proteins are just some factors that may affect RNA extraction yield.

In mRNA quantification by RT QPCR false positive signals may arise from amplification of the gene or pseudo-gene in genomic DNA. This problem can be avoided by designing the two PCR primers to hybridize to different exons, hence having at least one intron in between, or designing one of the primers to span across an exon/exon boundary (Wang and Seed, 2003). If such designs are not possible the sample should be DNase treated and tested for genomic contamination. This is done by running a no-RT control, which is a normal sample but without reverse transcriptase added.

2.1. The relative gene expression ratio

Comparing samples requires normalization to compensate for differences in the amount of biological material in the tested samples. A number of normalization procedures have been suggested based on physical parameters, such as volume, mass, size and cell number. Due to the heterogeneity of biological samples, these methods are usually unpractical and unreliable. More convenient is to normalize to total RNA amount, to ribosomal RNA, to externally added RNA standard, or to internal reference genes. The latter is currently the most popular strategy. However, finding appropriate reference genes for data normalization is one of the most challenging problems today. Extensive evidence suggests that all genes are regulated under some conditions, and the field will probably have to face that there is no universal reference gene with a constant expression in all tissues (Bustin, 2000; Gibbs et al., 2003). Due to this

uncertainty, any system relying on reference genes should be carefully validated. Panels of potential reference genes are now available for testing (http://www.tataa.com/referencepanels.htm), and softwares have been developed to find the optimal reference genes for defined systems (Pfaffl et al., 2004; Vandesompele et al., 2002). These methods assume that the genes with highest correlated expressions are the most appropriate reference genes. This assumption has the problem that regulated genes may have highly correlated expressions due to co-regulation, and those are not suitable references. The risk of selecting improper reference genes is minimized by choosing genes with different metabolic functions (http://www.tataa.com/referencepanels.htm).

The relative expression of two genes in a same sample is given by Ståhlberg et al. (2005):

$$\frac{N_A}{N_B} = K_{RS}\frac{\eta_B(1+E_B)^{CT_B-1}}{\eta_A(1+E_A)^{CT_A-1}} \tag{5}$$

N_A and N_B are the numbers of mRNA molecules of gene A and gene B, respectively, that were present in the test sample. K_{RS} is the relative sensitivity of the detection chemistries of the two assays (Ståhlberg et al., 2003), and η_A and η_B are the cDNA synthesis yields of gene A and gene B, respectively, defined as the fractions of mRNA molecules that are transcribed to cDNA in the RT reaction (Ståhlberg et al., 2004a). The exponent CT-1 accounts for the production of double stranded DNA in the first PCR cycle from the single stranded cDNA template generated by the reverse transcription reaction. η is assumed to be independent of both the total RNA and target mRNA concentrations.

The large number of parameters makes it quite complicated to determine the expression ratio of two genes in a single sample. In most applications the expression ratio of two genes is therefore compared in two or more samples; so called comparative quantification (Pfaffl et al., 2002). Typically, one gene is the reporter whose expression is expected to be affected by the condition studied and the other is a reference gene whose expression should be constant. Assuming the same RT yields in the samples K_{RS} and η cancel and the comparative expression ratio of the two samples is given by

$$\frac{\text{Sample 1}}{\text{Sample 2}} = \frac{\left[\frac{N_A}{N_B}\right]_{\text{Sample 1}}}{\left[\frac{N_A}{N_B}\right]_{\text{Sample 2}}} = \frac{\left[\frac{(1+E_B)^{CT_B-1}}{(1+E_A)^{CT_A-1}}\right]_{\text{Sample 1}}}{\left[\frac{(1+E_B)^{CT_B-1}}{(1+E_A)^{CT_A-1}}\right]_{\text{Sample 2}}} \tag{6}$$

Further assuming that the PCR efficiencies in the two samples are the same the expression simplifies to:

$$\frac{\text{Sample 1}}{\text{Sample 2}} = \frac{(1+E_B)^{CT_{B1}-CT_{B2}}}{(1+E_A)^{CT_{A1}-CT_{A2}}} \tag{7}$$

Finally, one may also assume 100% PCR efficiency. This gives:

$$\frac{\text{Sample 1}}{\text{Sample 2}} = \frac{2^{CT_{B1}-CT_{B2}}}{2^{CT_{A1}-CT_{A2}}} = 2^{(CT_{B1}-CT_{B2})-(CT_{A1}-CT_{A2})} = 2^{\Delta\Delta CT} \tag{8}$$

which is the well-known ΔΔCT-method.

2.2. Real-time PCR expression profiling

Characterizing samples through the expression of a single reporter gene normalized with the expression of a reference gene is an excellent approach for many simple studies, but it is not sufficient to classify complex samples. These are traditionally studied using microarray techniques, by which the expression of all the genes is assessed. However, in most cases it is not important to measure the expression of all the genes. In most tissues under reasonably well defined conditions only a fraction of the genome is active, and a limited number of genes have their transcriptional levels significantly altered by external stimuli or moderate environmental changes. A powerful experimental strategy for expression profiling is therefore to first study a small number of representative samples using microarray technology to identify the genes that are most sensitive to the studied condition, and then study these genes in greater detail and in many more samples by the more sensitive and cost efficient real-time PCR technique. Our yet limited experience of real-time PCR expression profiling suggests that the expression of some 20–50 genes catches most of the variation in the expression of the transcriptome under defined study conditions. The selected genes may constitute an expression pathway or a signaling chain, or they can be members of an operon, or respond to a certain environmental change, or reflect a perturbed or disease state of the organism, whose expression can be used to diagnose and classify the disease, and also for making prognosis.

A real-time PCR expression profiling experiment generates a CT value for each gene in each sample that reflects the transcriptional activity of that gene in the particular sample. As we shall see, from such data very valuable and accurate information about the transcriptional response of the studied system can be extracted. Even more powerful experimental design is to study the expression profile of a set of samples also as function of a third parameter such as time after treatment, drug load, etc. Such studies generate so called 3-way data, which are exceedingly informative (Smilde et al., 2004) (http://www.multid.se).

When analyzing the expression of many reporter genes the best approach is to look for characteristic expression profiles. This is a common approach in microarray studies. Having properly selected the genes for the real-time PCR study, expression profiling by real-time PCR has many important advantages to expression profiling by microarrays. Data quality is much better, sensitivity is higher, dynamic range is wider, and all those genes that are not pertinent to the studied conditions and only contribute with noise to the measurement are excluded. Further, the cost for real-time PCR measurements is way lower than for microarray studies, and one can study much larger number of samples and perform more biological repeats with real-time PCR, which is most important for the statistical analysis of data. Usually, the more

significantly expressed genes that are considered the better, but the approach is applicable even for two genes. As shown by Ståhlberg et al., reliable classification of a disease can be obtained by measuring the expression of two marker genes that are reciprocally expressed (Ståhlberg et al., 2003). In their work, summarized in the review by Leijon et al., in this issue, they classified non-Hodgkin lymphoma by measuring the relative expression of the immunoglobulin kappa (IgLκ) and lambda (IgLλ) light chains. The same strategy was recently used to diagnose mantle cell lymphoma by measuring the CCND1:CCND3 expression ratio (Jones et al., 2004).

When classifying data by expression patterns instead of relative expressions, data pre-treatment and normalization become less important. For example, its quite complex to accurately estimate IgLκ:IgLλ expression ratios, which are given by (see Eq. (5)):

$$\frac{N_{\mathrm{IgL}\kappa}}{N_{\mathrm{IgL}\lambda}} = K_{\mathrm{RS}} * \frac{\eta_{\mathrm{IgL}\lambda}(1+E_{\mathrm{IgL}\lambda})^{\mathrm{CT}_{\mathrm{IgL}\lambda}-1}}{\eta_{\mathrm{IgL}\kappa}(1+E_{\mathrm{IgL}\kappa})^{\mathrm{CT}_{\mathrm{IgL}\kappa}-1}} \tag{9}$$

Although the unknown parameters can be determined, it is time and resource consuming. If one instead plots the data in a regular scatter plot negative and positive samples are readily distinguished (Fig. 9 in the review by Mikael Leijon et al., in this issue).

2.3. Principal component analysis

A scatter plot is the most intuitive way to visualize the expression of two genes and to classify samples based on their co-expression pattern. Expression of three genes can be visualized in a 3D scatter plot. Studies based on any larger number of genes cannot be directly presented in a scatter plot, because we have no convenient way to plot data in more than three dimensions. To deal with higher order data multivariate chemometric tools are required to reduce the number of dimensions without loss of essential information. A classical, widely available, tool is Principal Components Analysis (PCA), which allows scientists to study many variables simultaneously. PCA not only informs about how the original variables are correlated, but it also shows how the samples are grouped. Principal Components (PCs) are mathematical constructs that can be interpreted as linear combinations of the studied variables with the following important properties:

(i) The PCs are orthogonal. Once a PC is linked to the behavior of one or several genes, one can be reasonably sure that this information will be unique and these genes will not correlate substantially with other PCs. The numerical coefficients, ranging from -1 to $+1$, given to each gene in each PC are called loadings and reflect how important the gene is to define this PC.
(ii) The PCs are sequential. This means that the first and most significant PC can be interpreted as the line in the original multidimensional space of all the genes that best fits the expression data and, hence, explains most of the observed variability and account for most of the information. The second most significant

PC is a vector perpendicular to the first PC that fits the expression data best, and accounts for most of the variability that is not accounted for by the first PC. Additional PCs are defined analogously. This means that one can extract PCs until a certain percentage of all the information, let say 80%, is accounted for, and then discard the remaining PCs.

Once the PCs have been calculated, the samples can be located in this new space. This is done using the scores, which specify the location of each sample on each PC. The original data can now be presented in a simple scatter plot of the scores to reveal groups among the samples or of the loadings to reveal groups among the genes. Many times PC1–PC2 and PC1–PC3 scatter plots are sufficient, but one can also construct PC1–PC2–PC3 scatter plots.

2.4. Data pre-treatment

We assume that all samples are based on, or normalized to, the same amount of material in some way. It can be same amount of total RNA, total mRNA, total cDNA, the same number of cells, body fluid or tissue. We also assume that the data are arranged in a matrix with the genes as columns and the samples as rows.

Real-time PCR raw data are typically expressed in CT values. For classification CTs should be converted to copy numbers using equation:

$$N_0 = (1 + E)^{(\mathrm{CT(sc)} - \mathrm{CT})} \tag{10}$$

This requires knowing CT(sc), which is the CT expected for a single template copy sample, and the PCR efficiency. These can be estimated from the slope and intercept of a standard curve (see Eq. (3)), or by calibration using standard additions. Their precise values are not very important for expression profiling by PCA, since the genes and samples are classified by way of their expression patterns, and reasonable estimates are good enough. It is usually sufficient to set all CT values above the lowest CT of primer–dimer signals to this value. Then, when converting CT values to copy numbers, set CT(sc) to the CT of the primer–dimers. PCR efficiencies are usually assay specific and they may also vary from sample to sample. But again classification based on expression profiling is quite insensitive to the efficiency values, and it is usually good enough to assume the same efficiency for all genes in all samples, and set it to a value typical of the particular sample matrix. For blood and many tissue samples this is 0.85–0.90.

Samples are often analyzed in duplicates or even more replicates, and many users average the repeats. For PCA we recommend not averaging repeats, and instead treat them as independent samples. Replicates should lie close to each other in the score plots and their spread gives a very good idea about the reproducibility of the analytical methodology and of each biological replicate.

We must decide if the data shall be analyzed in linear or logarithmic scale. We usually think of the order of enhancement or the order of suppression of expression, which is logarithmic scale. To analyze data in logarithmic scale, we shall calculate the logarithm of the copy numbers. Traditionally in expression profiling this is done

using log with base 2. In $\log_2$ a difference of 2 corresponds to 2-fold increase in expression, 3 corresponds to 4-fold increase, 4 to 8-fold increase, etc., while −2 corresponds to 2-fold reduction. But any other base for the logarithm can also be used.

Next we must decide about normalization. In RT QPCR one often normalizes with the expression of reference genes. However, for PCA this is not necessary. In fact, better classification is usually obtained when the data are not normalized because the normalization adds variance to the data due to the biological variation in the expression of the reference gene(-s). Instead the data can be autoscaled. Typically autoscaling is done for each gene, although the data can also be autoscaled for samples. In autoscaling the mean expression of each gene is first subtracted and then the results are divided by the standard deviation of the expression of each gene. Hence, the autoscaled expression data have zero mean and unit variance for every gene. The consequence of autoscaling is that all genes in the analysis are treated as equally important. Our experience is that autoscaling works very well when the genes are all significantly affected by the studied conditions. Autoscaling is also recommended when the genes are expressed at very different levels. Hence, in summary, we recommend the following data pre-treatment:

(1) Set all CT values that are higher than the lowest CT of primer–dimers to the same value.
(2) Convert CT values to copy numbers assuming an E typical of the matrix and CT(sc) to that of the primer–dimers.
(3) Normalize to the same total amount of RNA/cells/blood etc. Optionally normalize with the expression of reference gene(-s).
(4) Convert data to $\log_2$ base.
(5) Autoscale data.

Below we exemplify the procedure by analyzing real-time PCR gene expression profiles of the embryonic development of the African claw frog *Xenopus laevis*.

2.5. Embryonic development in X. laevis

Developmental biology is the effort to describe the relation between mRNA expression, translation and gene function throughout development. RNA expression can be studied on two levels: spatial and temporal distribution. Spatial expression profiles are measured within different organs or different parts of embryos, while temporal expression profiles are measured as function of developmental stage or time. Temporal expression profiling during early development of mammals is limited by the small cellular amounts of protein and RNA (about 20 pg per cell). In contrast, amphibian eggs and embryos, such as those of *X. laevis*, contain several μg of total RNA. This feature further adds to the picture of *X. laevis* as one of the most popular model organisms for developmental studies. Two different groups of mRNA molecules are present during *Xenopus* early development. One group is the maternal mRNA molecules that were expressed during oogenesis and that are present in oocyte already prior to fertilization. Some of the maternal mRNAs (e.g. β-catenin,

VegT, Vg1, Wnt11) are heterogeneously distributed throughout the egg and are responsible for the main body axis formation and germ layer induction (Mowry and Cote, 1999). The second group of mRNA molecules is called zygotic. Zygotic mRNAs are transcribed after a specific developmental phase, which is called mid-blastula transition (MBT) and occurs during the gastrulation process. Among zygotic mRNAs are genes that antagonize bone morphogenetic protein (BMP) and Wnt signaling (e.g. follistatin, cerberus, chordin, noggin), as well as genes that are important for the forthcoming development and organogenesis (e.g. N-CAM and N-tubulin important for neurulation are highly expressed in neural tissue, whereas cardiac actin is expressed in heart tissue).

We have measured the expression of activin, Xbra, cerberus, chordin, derriere, dishevelled, follistatin, goosecoid, GSK3, HNF-3beta, N-CAM, p53, siamois, VegT, Vg1, and Xnot in the *Xenopus* developmental stages 1, 2, 4, 5, 6–7, 8–9, 11, 15, 17, 18–19, 21, 28, 32, 35–36, 41 and 44 assigned according to Nieuwkoop and Faber (1994). This is a total of 16 genes studied in 16 developmental stages. All genes weremeasured in at least duplicates, starting from different samples (biological repeats). In total 39 expression measurements were carried out in 16 developmental stages.

Two sets of three embryos from the different stages were collected from one in vitro fertilization and frozen immediately at −70 °C. RNA from each sample was extracted using Trizol reagent (Invitrogen) according to the instructions of the manufacturer and its concentration was determined by absorption. The RNA quality was analyzed by 1.5% ethidium bromide agarose gel electrophoresis. cDNA was produced starting with 1 μg of total RNA and 10 pmol 25-dT oligo. After incubation at 72 °C for 10 min, 100 U MMLV reverse transcriptase (Promega), 12 U RNasin (Promega) and 5 nmol dNTP were added to a total volume of 10 μl, and incubated at 37 °C for another 70 min. The reactions were subsequently diluted to 200 μl and frozen. The PCR reaction mixture had a final volume of 20 μl and contained 2 μl of cDNA, 10.000-fold diluted SYBR Green solution (Molecular Probes), 0.4 mM forward and reverse primer, 0.3 mM dNTPs, 3 mM $MgCl_2$ and 1 U Taq polymerase (Promega). Real-time PCR data were collected on the BioRad iCycler iQ and the Corbett Research Rotor-Gene 3000 with cycling conditions: 95 °C for 3 min, 40 cycles at 95 °C for 20 s, 60 °C for 20 s, and 72 °C for 20 s. Finally the samples were cooled to 60 °C and a melting curve was recorded between 60 °C and 95 °C with steps of 0.5 °C.

Real-time PCR expression measurements are frequently normalized with the expression of reference genes. But this approach is highly unsuitable for development studies, because no gene seems to be expressed at a constant level during development. In a previous study we measured the expression of ODC, EF-1α, L8, GAPDH and H4, which are five popular *Xenopus* reference genes, during development from the egg (stage 1) to the tadpole (stage 44), and found very large variations when normalizing their expressions to the total amount of RNA (Sindelka et al., in press). Autoscaling is therefore much better option.

The RT QPCR data were analyzed using the GenEx software (http://www.multid.se). In the pre-treatment all CT values above 30 were set to 30, which was

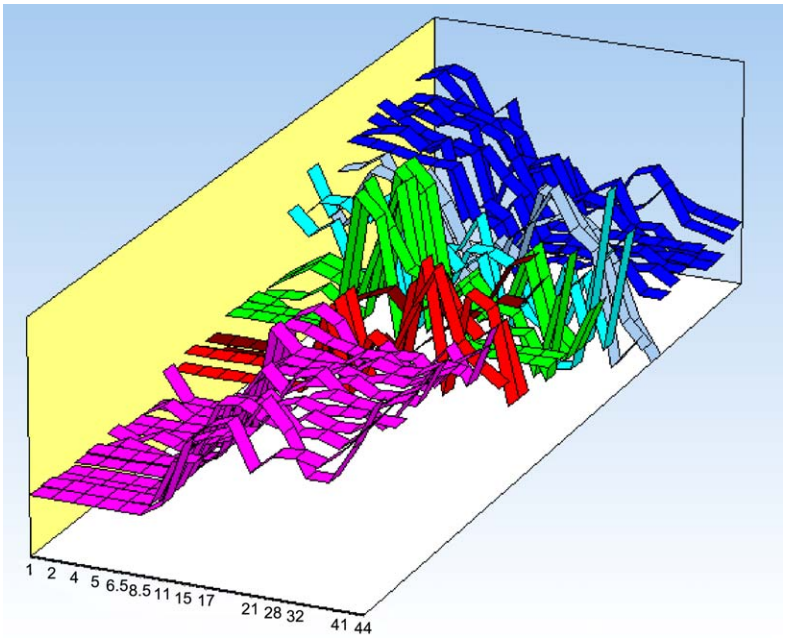


Fig. 10. Expression of *Xenopus laevis* developmental genes during development. Maternal genes are shown in blue (dishevelled, p53, VegT, and Xnot), cyan (Vg1) and sky blue (GSK-3beta), MBT genes are shown in green (Xbra and cerberus), and the late genes are shown in pink (activin, chordin, derriere, follistatin, goosecoid, HNF-3beta, and siamois) and in red (N-CAM). (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

the lowest CT value observed for primer–dimers. The CT values were then converted to copy numbers assuming a PCR efficiency of 0.90 and CT(sc) = 30. The data were then converted to $\log_2$ scale and autoscaled (Fig. 10).

Principal components were calculated with GenEx, and it was found that the first two PCs account for 76% of the total variation in the data and the first three PCs account for 85% of the variation. Fig. 11 shows the genes in a PC1–PC2 scatter plot. Three groups are seen. In different shades of blue[1] we have the genes dishevelled, GSK-3beta, p53, VegT, Vg1, and Xnot that are expressed in the early stages of development. Genes shown in different shades of red (activin, chordin, derriere, follistatin, HNF-3beta, N-CAM, and siamois) are expressed in late stages. The genes shown in green (Xbra and cerberus) have maximum expression at the MBT stage. By comparing the spread of biological replicates we can distinguish subgroups. In the blue group biological replicates of Vg1 (cyan) and GSK-3beta (sky blue) are distinctly separated from the other early genes. Closer inspection of the expression profiles reveals that both Vg1 and GSK-3beta are expressed also at a later stage of development, while the other early genes are not. Among the late genes the N-CAM repeats

[1] For interpretation of color in Figs. 10–13, the reader is referred to the web version of this article.

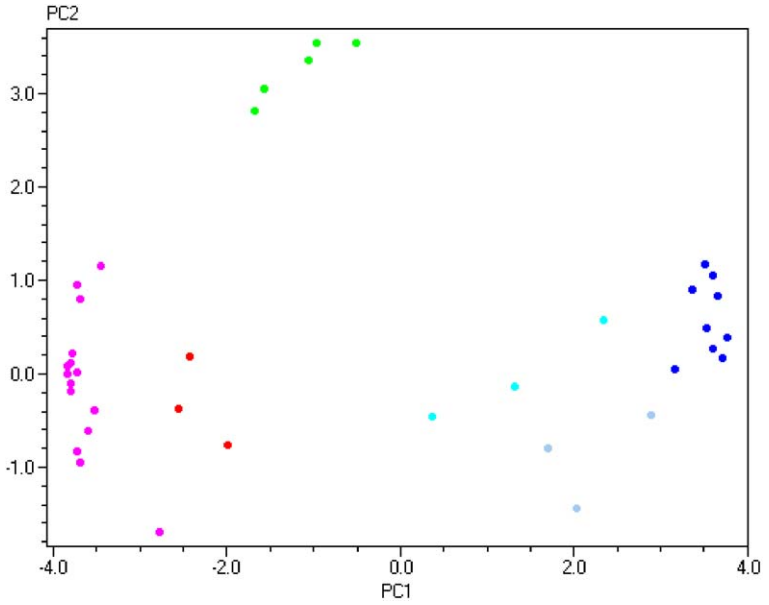


Fig. 11. PC1 vs. PC2 scatter plot of the expression of *Xenopus laevis* developmental genes. Same color coding as in Fig. 10. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

(red) are distinct. Closer inspection of the expression data reveals that N-CAM is also highly expressed at MBT, while the other late genes are not. There is one outlier among the pink samples. It is one of the goosecoid biological replicates. Since the other replicate is within the cluster we cannot make any conclusions about biological significance. The groups are even more distinct when viewed in a PC1–PC2–PC3 scatter plot, which accounts for 85% of the variation in the data (Fig. 12).

The stages can also be classified by PCA. Fig. 13 shows the developmental stages in a PC1–PC2 plot classified on the basis of the expressions of the genes. Also here three clusters are evident. First cluster is stages 1–8.5, second cluster is stages 11 and 15, and the third cluster is stages 17–44.

2.6. Single cell gene expression profiling

Cell measurements on tissue or culture in medical research are typically performed on a large number of cells. Gene expression measurements of cell cultures that have been treated differently and comparisons of healthy and abnormal tissues reveal the molecular mechanisms of function and disease. The methods most commonly used require thousands or millions of cells for analysis, and many fundamental breakthroughs in cell biology, physiology and pathology have sprung from these results. However, measuring on a large number of cells limits the findings to population-wide effects and does not address any important information about single cell

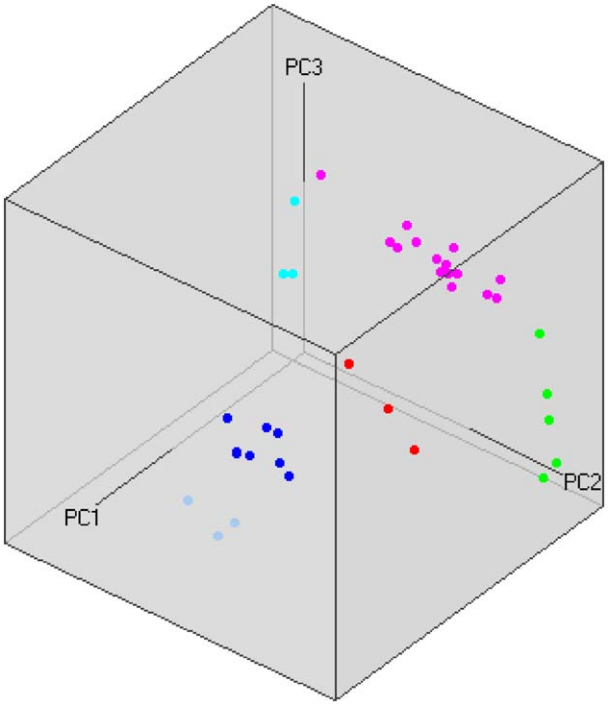


Fig. 12. PC1 vs. PC2 vs. PC3 scatter plot of the expression of *Xenopus laevis* developmental genes. Same color coding as in Fig. 10. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

biology. For example, a correlation between the expressions of two genes in a population only tells us that the two genes respond to the same external stimuli at approximately the same way. If we can verify that the correlation is also present on the one cell level, we can conclude that their transcriptional regulations are coupled and perhaps even controlled by the same molecular mechanism.

Another aspect of gene expression measurements is population heterogeneity. Tissues are made of many different cell types that are expected to respond differently to external stimuli and also to be differently affected by disease. In the pancreas, for example, the islets of Langerhans release hormones from at least five cell types, each with distinctly different characteristics (Bishop and Polak, 1997). To fully understand the physiology behind the complex regulatory mechanisms behind biological functions each cell type needs to be studied separately.

Very high sensitivity is required for analysis of transcriptional activity in individual cells. At any one time-point a typical eukaryotic cell contains abut 0.5 pg mRNA. This is equivalent to a few hundred thousand molecules transcribed from about ten thousand genes. To analyze the expression of these genes two methods are available today: In situ hybridization and nucleic acid amplification methods. In situ hybridization studies preserve the morphology of the tissues and expression

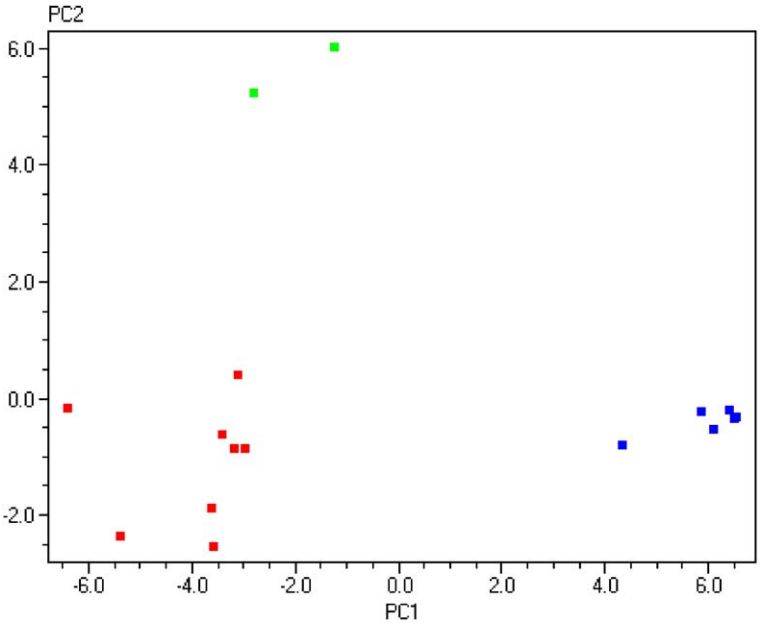


Fig. 13. PC1 vs. PC2 scatter plot of the developmental stages of *Xenopus laevis* classified by the expression of its developmental genes. Stages 1–8.5 are shown in blue, 11–15 in green, and 18.5–44 in red. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

is assessed by observing fluorescent probes inside the cell by microscopy (Levsky et al., 2002). Nucleic acid amplification methods are usually variants of either RT-PCR or T7 antisense RNA (aRNA) amplification. The aRNA method is an amplification process where a cDNA library is created from the mRNA, while extending the 5′-end with the T7 promoter sequence. With T7 RNA polymerase large quantities of antisense RNA copies are then produced from the cDNA library, which in turn are reverse transcribed back to cDNA (Wang et al., 2003). In combination with PCR, this method allows for global transcription profiling at the single cell level. However, there is a concern that not all transcripts are amplified with the same efficiency and that the results are biased in unknown way (Nygaard et al., 2003).

RT QPCR is the most versatile method for single cell mRNA analysis and it also offers quantitative information about transcript levels (Bengtsson et al., 2005). High standards on isolation methods are required for proper quantification. For accurate transcriptional profiling all mRNAs from the single cell must make its way to the reaction tube intact and accessible for the reverse transcriptase without introducing inhibitors of the downstream reactions. A number of methods are available. Laser-capture microdissection (LCM) allows handling of μm -sized pieces of tissue without compromising the cellular integrity. The laser beam cuts sections in thin slices of tissue fixed to transfer film, allowing selection based on morphology or staining (see

also the review by Panzini et al., in this issue). Fluorescence activated cell sorting (FACS) is commonly used to sort out a specific cell type from a heterogeneous mixture of cell populations and it can also be used to collect individual cells. Microscopes fitted with micromanipulators are used by electrophysiologists to record currents across membranes in single cells in so called patch-clamp recordings. With minor modifications this setup can be used together with RT QPCR to achieve a powerful combination of functional and transcriptional recordings (Sucher et al., 2000; Liss et al., 2001). For a single cell QPCR in pre-implantation diagnostics, see the review by Traeger-Synodinos in this issue.

As already mentioned the cell-to-cell variation is large even in seemingly homogenous synchronized cell cultures (Levsky and Singer, 2003). Events in the nucleus that determine the fate of the cell are highly probabilistic, or random, in nature. Chemical reactions that involve only a small number of molecules, such as the binding of transcription factors to DNA or the modifications of proteins, are all intrinsically stochastic events. Hence, unlike for the population of cells, the behavior of individual cells cannot be predicted because of a large degree of randomness, or noise that garbles the outcome. This view on gene expression is now established in the field of single cell biology.

We recently studied the transcript levels in a population of insulin producing β-cells and found that a small percentage of the cells express most of the mRNA in the population (Bengtsson et al., 2005). The observed variation was consistent with the lognormal distribution (Fig. 14). This implies that the typical cell in a population

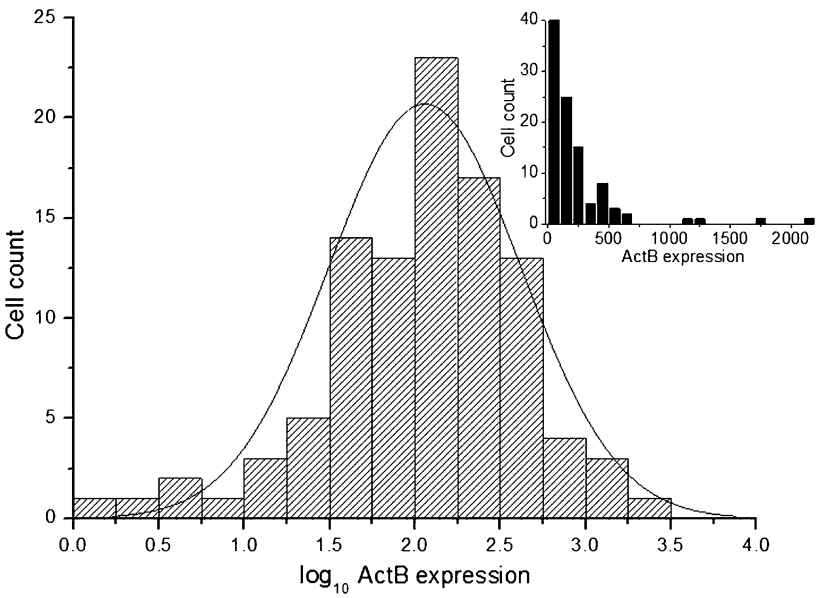


Fig. 14. Lognormal distribution of the expression of ActB in 121 individual β-cells from the pancreas of mouse. Inset shows the distribution in linear scale.

is not well described by the common arithmetic mean of the expression, which is the value measured on a population, but instead by the geometric mean. We further observed that among the five genes studied ActB, Ins1, Ins2, Abcc8, and Kcnj11, only the expressions of Ins1 and Ins2 correlated on the cell level. When glucose was added transcription of Ins1, Ins2 as well as of ActB increased. But ActB was expressed in different cells from those expressing Ins1 and Ins2 (Bengtsson et al., 2005). The origin of the lognormal distributions is unclear, but it has been suggested to arise from multiplicative propagation of fluctuations in correlated equilibria (Furusawa et al., 2005).

Acknowledgement

This work was supported in part by project no. AVOZ50520514 awarded by the Academy of Sciences of the Czech Republic.

References

A Nanofluidic Chip for Absolute Quantification of Target Nucleic Acid Sequences. Pharmaceutical Discovery, October 1, 2005.

Akane, A., Matsubara, H., Nakamura, S., Takahashi, S., Kimura, K., 1994. Identification of the heme compound copurified with deoxyribonucleic acid (DNA) from bloodstrains, a major inhibitor of polymerase chain reaction (PCR) amplification. J Forensic Sci. 39, 362–372.

Al-Soud, W.A., Jonsson, L.J., Radström, P., 2000. Identification and characterization of immunoglobulin G in blood as a major inhibitor of diagnostic PCR. J. Clin. Microbiol. 38, 345–350.

BD QZyme™ Assays for Quantitative PCR, 2003. Clontechniques 18 (4), 2–3.

Belgrader, P., Young, S., Yuan, B., Primeau, M., Christel, L.A., Pourahmadi, F., Northrup, M.A., 2001. A battery-powered notebook thermal cycler for rapid multiplex real-time PCR analysis. Anal Chem. 73 (2), 286–289.

Bengtsson, M., Karlsson, J.H., Westman, G., Kubista, M., 2003. A new minor groove binding asymmetric cyanine reporter dye for real-time PCR. Nucleic Acids Res. 31 (8), e45.

Bengtsson, M., Ståhlberg, A., Rorsman, P., Kubista, M., 2005. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Res. 15 (10), 1388–1392.

Bishop, A.E., Polak, J.M., 1997. The anatomy, organisation and ultrastructure of the islets of Langerhans. In: Pickup, J., Williams, G. (Eds.), Textbook of Diabetes. Blackwell Science, Oxford, UK, pp. 6.1–6.16.

Bonnet, G., Tyagi, S., Libchaber, A., Kramer, F.R., 1999. Thermodynamic basis of the enhanced specificity of structured DNA probes. Proc. Natl. Acad. Sci. USA 96 (5), 6171–6176.

Braasch, D.A., Corey, D.R., 2001. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. Chem. Biol. 8 (1), 1–7.

Brenan, C., Morrison, T., 2005. High throughput, nanoliter quantitative PCR. Drug Discovery Today: Technologies 2, 247–253.

Bustin, S.A., 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J. Mol. Endocrinol. 25, 169–193.

Bustin, S.A., 2002. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J. Mol. Endocrinol. 29 (1), 23–29.

Caplin, B.E., Rasmussen, R.P., Bernard, P.S., Wittwer, C.T., 1999. LightCycler™ hybridization probes–the most direct way to monitor PCR amplification and mutation detection. Biochemica 1, 5–8.

Costa, J.-M., Ernault, P., Olivi, M., Gaillon, T., Arar, K., 2004. Chimeric LNA/DNA probes as a detection system for real-time PCR. Clin. Biochem. 37, 930–932.

Egholm, M., Buchardt, O., Nielsen, P.E., Berg, R.H., 1992. Peptide nucleic acids (PNA). Oligonucleotide analogs with an achiral peptide backbone. J. Am. Chem. Soc. 114, 1895–1897.

Furusawa, C., Suzuki, T., Kashiwagi, A., Yomo, T., Kaneko, K., 2005. Ubiquity of log-normal distributions in intra-cellular reaction dynamicsBiophysics, Vol. 1. The Biophysical Society of Japan, pp. 25–31.

Gibbs, P.J., Cameron, C., Tan, L.C., Sadek, S.A., Howell, W.M., 2003. House keeping genes and gene expression analysis in transplant recipients: a note of caution. Transpl Immunol. 12 (1), 89–97.

Higuchi, R., Dollinger, G., Walsh, P.S., Griffith, R., 1992. Simultaneous amplification and detection of specific DNA-sequences. Bio-Technology 10 (4), 413–417.

Holland, P.M., Abramson, R.D., Watson, R., Gelfand, D.H., 1991. Detection of specific polymerase chain reaction product by utilizing the $5' \rightarrow 3'$ exonuclease activity of *Thermus aquaticus* DNA polymerase. Proc. Natl. Acad. Sci. USA 88 (16), 7276–7280.

Izraeli, S., Pfleiderer, C., Lion, T., 1991. Detection of gene expression by PCR amplification of RNA derived from frozen heparinized whole blood. Nucleic Acids Res. 19, 6051.

Jansen, K., Nordén, B., Kubista, M., 1993. Sequence dependence of 4′,6-diamidino-2-phenylindole (DAPI)-DNA interactions. J. Am. Chem. Soc. 115, 10527–10530.

Jones, C.D., Darnell, K.H., Warnke, R.A., Zehnder, J.L., 2004. Cyclin D1/Cyclin D3 ratio by real-time PCR improves specificity for the diagnosis of mantle cell lymphoma. J. Mol. Diagn. 6 (2), 84–89.

Kubista, M., 2004. Nucleic acid-based technologies: application amplified. Pharmacogenomics 5 (6), 767–773.

Kubista, M., Zoric, N., 2004. PCR platforms. Encyclopedia of Medical Genomics and Proteomics. Available from: <http://www.dekker.com/servlet/product/DOI/101081EEDGP120020685>.

Kubista, M., Ståhlberg, A., Bar, T., 2001. Light-up probe based real-time Q-PCR. In: Raghavachari, R., Tan, W. (Eds.), Genomics and Proteomics Technologies, Proceedings of SPIE, 4264, pp. 53–58.

Kutyavin, I.V., Afonina, I.A., Mills, A., et al., 2000. 3′-Minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. Nucleic Acids Res. 28 (2), 655–661.

Levsky, J.M., Singer, R.H., 2003. Gene expression and the myth of the average cell. Trends Cell Biol. 13, 4–6.

Levsky, J.M., Shenoy, S.M., Pezo, R.C., Singer, R.H., 2002. Single-cell gene expression profiling. Science 297, 836–840.

Li, Qingge, Luan, Guoyan, Guo, Qiuping, Liang, Jixuan, 2002. A new class of homogeneous nucleic acid probes based on specific displacement hybridization. Nucleic Acids Res. 30, e5.

Lind, K., Stahlberg, A., Zoric, N., Kubista, M., in press. Combining sequence specific probes and DNA binding dyes in real-time PCR for specific nucleic acid quantification and melting curve analysis. Biotechniques.

Liss, B., Franz, O., Sewing, S., Bruns, R., Neuhoff, H., Roeper, J., 2001. Tuning pacemaker frequency of individual dopaminergic neurons by Kv4.3L and KChip3.1 transcription. EMBO J. 20, 5715–5724.

Mackya, I.M., 2004. Real-time PCR in the microbiology laboratory. Clin. Microbiol. Infect. 10, 190–212.

Mattarucchi, E., Marsoni, M., Binelli, G., Passi, A., Lo Curto, F., Pasquali, F., Porta, G., 2005. Different real time PCR approaches for the fine quantification of SNP's alleles in DNA pools: assays development, characterization and pre-validation. J. Biochem. Mol. Biol. 38 (5), 555–562, Sep 30.

Mhlanga, M.M., Malmberg, L., 2001. Using molecular beacons to detect single-nucleotide polymorphisms with real-time PCR. Methods 25 (4), 463–471.

Mowry, K.L., Cote, C.A., 1999. RNA sorting in *Xenopus* oocytes and embryos. FASEB J. 13 (3), 435–445.

Nazarenko, I., Lowe, B., Darfler, M., Ikonomi, P., Schuster, D., Rashtchian, A., 2002. Multiplex quantitative PCR using self-quenched primers labelled with a single fluorophore. Nucleic Acids Res. 30 (9), e37.

Nieuwkoop, P.D., Faber, J., 1994. Normal Table of *Xenopus laevis*. Garland Publishing, Inc., New York, London.

Nygaard, V., Loland, A., Holden, M., Langaas, M., Rue, H., Liu, F., Myklebost, O., Fodstad, O., Hovig, E., Smith-Sorensen, B., 2003. Effects of mRNA amplification on gene expression ratios in cDNA experiments estimated by analysis of variance. BMC Genomics 4, 11.

Nygren, J., Svanvik, N., Kubista, M., 1998. The interaction between the fluorescent dye thiazole orange and DNA. Biopolymers 46, 39–51.

Peters, I.R., Helps, C.R., Day, M.J., 2004. Real-time RT-PCR: considerations for efficient and sensitive assay design. J. Immunol. Methods 286 (1-2), 203–217.

Pfaffl, M.W., Horgan, G.W., Dempfle, L., 2002. Relative expression software tool (REST) for groupwise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. 30 (9), E36.

Pfaffl, Michael W., Tichopád, Aleš, Prgomet, Christian, Neuvians, Tanja P., 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—Excel-based tool using pair-wise correlations. Biotechnol. Lett. 26, 509–515.

Primer3. Available from: <http://frodo.wi.mit.edu/..cgi-bin/primer3/primer3_www.cgi>. Netprimer. Available from: <http://www.premierbiosoft.com/netprimer>.

Ririe, K.M., Rasmussen, R.P., Wittwer, C.T., 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Anal. Biochem. 245 (2), 154–160.

Rutledge, R.G., 2004. Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. Nucleic Acids Res. 32, e178.

Rutledge, R.G., 2005. Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. Nucleic development of a kinetic-based sigmoid model for the polymerase chain reaction and its application to quantitative PCR. In: 2nd International qPCR Symposium, Freising.

Rutledge, R.G., Cote, C., 2003. Mathematics of quantitative PCR and the applications of standard curves. Nucleic Acids Res. 31, e93.

Simonson, T., Pecinka, P., Kubista, M., 1998. DNA tetraplex formation in the control region of c-myc. Nucleic Acids Res. 26, 1167–1172.

Simonsson, T., 2001. G-Quadruplex DNA structures—variations on a theme. Biol. Chem. 382, 621.

Sindelka, R., Ferjentsik, Z., Jonák, J., in press. Developmental expression profiles of *Xenopus laevis* reference genes. Dev. Dyn.

Sjöback, R., Nygren, J., Kubista, M., 1995. Absorption and fluorescence properties of fluorescein. Spetrochim. Acta Part A 51, L7–L21.

Smilde, A., Bro, R., Geladi, P., 2004. MultiWay Analysis. John Wiley & Sons Ltd., ISBN 0-471-98691-7.

Ståhlberg, A., Åman, P., Ridell, B., Mostad, P., Kubista, M., 2003. Quantitative real-time pcr method for detection of B-lymphocyte monoclonality by comparison of *k* and *l* immunoglobulin light chain expression. Clin. Chem. 49, 51–59.

Ståhlberg, A., Håkansson, J., Xian, X., Semb, H., Kubista, M., 2004a. Properties of the reverse transcription reaction in mRNA quantification. Clin. Chem. 50 (3), 509–515.

Ståhlberg, A., Kubista, M., Pfaffl, M., 2004b. Comparison of reverse transcriptases in gene expression analysis. Clin. Chem. 50 (9), 1678–1681.

Ståhlberg, A., Zoric, N., Åman, P., Kubista, M., 2005. Quantitative real-time PCR for cancer detection: the lymphoma case. Expert Rev. Mol. Diagn. 5, 221–230.

Sucher, N.J., Deitcher, D.L., Baro, D.J., Warrick, R.M., Guenther, E., 2000. Genes and channels: patch/voltage-clamp analysis and single-cell RT-PCR. Cell Tissue Res. 302, 295–307.

Svanvik, N., Westman, G., Dongyuan, W., Kubista, M., 2000. Light-up probes thiazole orange conjugated PNA for detection of nucleic acid in homogeneous solution. Anal. Biochem. 281, 26–35 http://www.lightup.se.

Tyagi, S., Kramer, F.R., 1996. Molecular Beacons: probes that fluorescence upon hybridization. Nat. Biotechnol. 14 (3), 303–308.

Tyagi, S., Bratu, D.P., Kramer, F.R., 1998. Multicolor molecular beacons for allele discrimination. Nat. Biotechnol. 16 (1), 49–53.

Uehara, H., Nardone, G., Nazarenko, I., Hohman, R.J., 1999. Detection of telomerase activity utilizing energy transfer primers: comparison with gel- and ELISA-based detection. Biotechniques 26 (3), 552–558.

Van, T.-L., Paquet, N., Calvo, E., Cumps, J., 2005. Improved real-time RT PCR method for high throughput measurements using second derivative calculations and double correction. Biotechniques 38, 287–293.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3 (7), 0034.I–0034.II.

Wang, X., Seed, B., 2003. A PCR primer bank for quantitative gene expression analysis. Nucleic Acids Res. 31 (24), e154.

Wang, J., Hu, L., Hamilton, S.R., Coombes, K.R., Zhang, W., 2003. RNA amplification strategies for cDNA microarray experiments. Biotechniques 34, 394–400.

Whitcombe, D., Theaker, J., Guy, Sp., Brown, T., Little, S., 1999. Detection of PCR products using self-probing amplicons and fluorescence. Nat. Biotechnol. 17 (8), 804–807.

Wilson, R., Johansson, M.K., 2003. Photoluminescence and electrochemiluminescence of a Ru(II) (bpy)3-quencher dual-labeled oligonucleotide probe. Chem. Commun. 21, 2710–2711.

Wittwer, C.T., Herrmann, M.G., Gundry, C.N., Elenitoba-Johnson, K.S., 2001. Real-time multiplex PCR assays. Methods 25 (4), 430–442.

Zipper, H., Brunner, H., Bernhagen, J., Vitzthum, F., 2004. Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. Nucleic Acids Res. 32 (12), e103.

Paper III

M. Kubista, B. Sjogreen, A. Forootan, R. Sindelka, J. Jonak, JM. Andrade. Real-time PCR gene expression profiling. European Pharmaceutical Reviews, 2007. Vol. 1.

Real-time PCR gene expression profiling

Mikael Kubista, TATAA Biocenter and MultiD Analyses AB, Sweden, **Björn Sjögreen**, Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, United States and MultiD Analyses AB, **Amin Forootan**, MultiD Analyses AB, **Radek Sindelka** and **Jiri Jonák**, Laboratory of Gene Expression, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic and **José Manuel Andrade**, Dept of Analytical Chemistry, University of A Coruna, Spain

Real-time PCR has rapidly become the preferred technique for quantitative analysis of nucleic acids. Its superior sensitivity, reproducibility and dynamic range make it the preferred choice for expression profiling in scientific, as well as routine, applications.[1]

Initially, most real-time PCR studies targeted a single gene of interest, whose expression reflects the state of disease, response to a drug or a change in the environment and the like. However, most biological phenomena are complex and cannot be described by the expression of individual genes. Instead expression profiles must be measured and interpreted. Traditionally this has been done using microarray techniques,[2] but the development of high throughput platforms opens up the possibility of using the more sensitive and cost efficient real-time PCR technology.

In gene expression profiling, the expression of many genes is measured in many samples. The genes are selected based on prior knowledge about their function (and often exploratory microarray studies) to be informative in respect to the studied condition. The data are then analysed to identify genes or samples with similar expressions. A few powerful biostatistical methods are available to find these similarities. Principal Components Analysis (PCA), Hierarchical clustering and Kononen Self Organizing Maps (SOM) are some of the most powerful. This article will outline an expression profiling experiment, pre-processing and scaling of the data in order to identify genes that have common regulations and samples

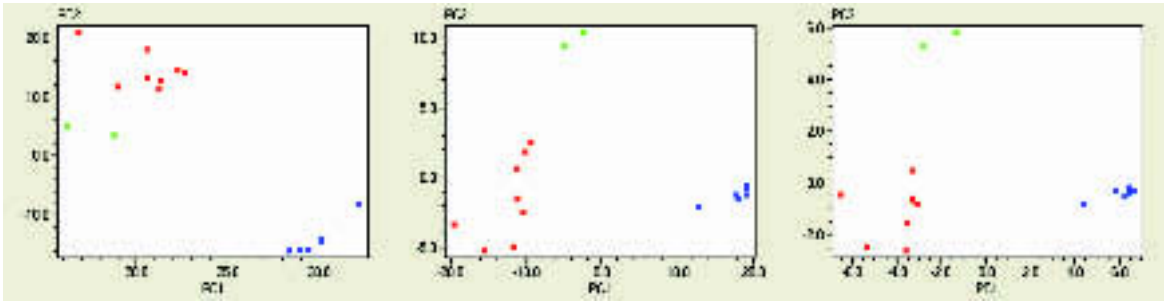


Figure 1: PC1 vs. PC2 scatter plots of the samples (scores). Stages 1-8.5 are shown in blue, stages 11 and 15 in green and stages 17 – 44 in red. From left to right: raw, mean centred and autoscaled data

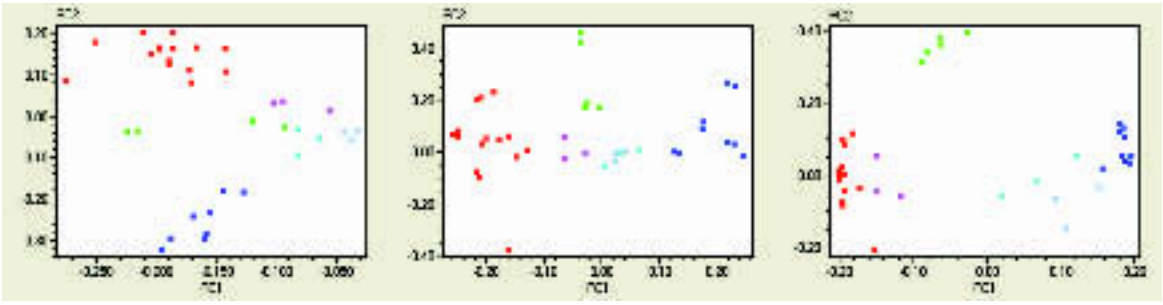


Figure 2: PC1 vs. PC2 scatter plots of the genes (loadings). Dishevelled, p53, VegT and Xnot are shown in blue, GSK-3beta in sky blue, Vg1 in aqua, Xbra and Cerberus in green, activin, chordin, derriere, follistatin, HNF-3beta and siamois in red, and N-CAM, in fuchsia. From left to right: raw, mean centred and autoscaled data

that show the same expression patterns. Our example is a study of the early development of the African claw frog, *Xenopus laevis*.

The expression of 16 genes was measured in 16 stages of development, ranging from the oocyte to the tadpole. All samples were measured in biological replicates and the RNA was extracted, reverse transcribed and analysed by real-time PCR as described previously.[1,3] For each reaction a CT value was registered.

There are two complications in this study, which are not uncommon in expression profiling. Firstly, many genes have virtually no expression in some of the stages, resulting in off-scale measurements. Secondly, there are no suitable reference genes for normalisation of the data because all Xenopus genes studied in any detail so far show variations in expression levels during development.[3] These are addressed in the pre-processing of the data.

Pre-processing of real-time PCR data for expression profiling has been described in detail before[1] and excellent on-line tutorials are available (www.multid.se). For the present data the level of detection (LOD), which is the highest CT value observed for a positive signal, was 30 and all CT values above 30 were set to 30. Setting off-scale values to 31 instead did not make any difference. A PCR efficiency of 90 per cent was assumed for all assays and the data were normalised to the total amount of RNA in the samples. No normalisation with reference genes was performed. The CT values were converted to relative quantities and then converted to $\log_2$ scale. Finally the data were mean centred or autoscaled. Mean center data is subtracting the mean expression of each gene. It removes the influence of overall expression levels in the classification, while maintaining the magnitudes of the changes. Autoscaling is mean centred followed by division with the standard deviation of the expression of each gene. This removes the influence of both the expression level and the magnitudes of the changes and gives rise to classification based on the relative changes in expression. All pre-processing and subsequent scaling of the real-time PCR data was performed using GenEx from MultiD (www.multid.se).

The expression of activin, Xbra, cerberus, chordin, derriere, dishevelled, follistatin, goosecoid, GSK3, HNF-3beta, N-CAM, p53, siamois, VegT, Vg1 and Xnot was measured in the *Xenopus* developmental stages 1, 2, 4, 5, 6-7, 8-9, 11, 15, 17, 18-19, 21, 28, 32, 35-36, 41 and 44 assigned according to Nieuwkoop and Faber.[4] 2-4 biological replicates were performed on each sample giving a total of 39 expression measurements in 16 developmental stages. The data were classified by PCA, Hierarchical clustering and the SOM.

Principal Component Analysis

The first multivariate method to be applied should be PCA.[5] Briefly, the principal components (PCs) are linear combinations of the original genes and samples defining a space of lower dimensionality in which the data can be visualised in scatter plots. PC1 vs. PC2 scatter plots of the stages (the 'scores') are shown in Figure 1. From left to right the data are unscaled, mean centered and autoscaled data. Three groups separate well in all scatter plots, although they are more differentiated for mean centred and autoscaled data. *Xenopus laevis* has very little transcription in the early stages of development and proteins are produced from translation of maternal mRNAs present in the oocyte. Transcription is initiated during a process called the mid-blastula transition (MBT). The three clusters we see in the PC1 vs. PC2 plot (Figure 1) should therefore represent the early or pre-MBT stage, the MBT and the late post-MBT stage. Among the blue stages we see that five are tightly clustered, while 8.5 is off from the group's centre. This suggests the latter has begun to differentiate into MBT.

Scores	1	2	4	6	8.5	8.5	11	15	17	18.5	21	28	32	35.5	41	44
PC1	6.50	6.39	6.53	5.86	6.12	4.33	-1.28	-2.80	-2.98	-3.10	-3.16	-3.03	-3.42	-3.56	-5.37	-6.40
PC2	-0.36	-0.20	-0.31	-0.24	-0.52	-0.79	6.04	5.23	-0.84	0.40	-0.94	-1.98	-0.62	-2.55	-2.36	-0.17

Figure 3: Scores of the first two principal components of the autoscaled data. Samples contributing with positive scores to PC1 are indicated in blue, samples contributing to PC1 with negative scores are indicated in red and samples contributing with positive scores to PC2 are indicated in green

Figure 2 shows PC1 vs. PC2 scatter plots for the genes (the 'loadings') based on unscaled, mean centred and autoscaled data. The genes indicated in red and blue separate in all three scatter plots, showing that they are expressed at different stages during development. The genes indicated by green colour cluster more clearly in the autoscaled plot, suggesting that they have the same expression profiles.

The biological replicates of Vg1 (cyan) and GSK-3beta (sky blue) separate from other genes in all plots. In the scatter plot of autoscaled data Vg1 and GSK-3beta are close to the genes in blue colour suggesting they have a similar expression profile. Likewise, N-CAM (fuchsia) has a similar profile to the genes in red.

We can identify genes critical for the different developmental stages by inspecting the scores and the loadings. The loadings are the contributions from the genes to the principal components and the scores are the contributions from the samples. The larger the loading, the more important the gene for a particular PC is; and the larger the score, the more important the sample is. Since the best PCA separation of the genes is obtained for the autoscaled data, let us inspect the corresponding loadings. In the loadings plot we see that all genes labeled bluish have positive PC1 values (Figure 2) and stages 1-8.5 have positive PC1 scores (Figure 1). Hence, Dishevelled, p53, VegT, Xnot, GSK-3beta and Vg1 are predominantly expressed during the early stages of development. The genes labeled reddish have negative PC1 loadings (Figure 2) and stages 17-44 have negative PC1 scores (Figure 1). Hence, activin, chordin, derriere, follistatin, HNF-3beta and N-CAM are expressed predominately during the late stages of development. Xbra and Cerberus (green) have positive PC2 loadings (Figure 2) and stages 11 and 15 have positive PC2 scores. Hence, Xbra and Cerberus are expressed during the mid blastula transition.

The PC1 vs. PC2 plots show the samples and the genes in a reduced space of 2-dimensions. Although the space is optimised for information, some is missing. The amount of information (technically, the amount of explained variance) contained in a PC1 vs. PC2 scatter plot is obtained from the eigen values. For the unscaled, mean centred and autoscaled data it is 96%, 90% and 77%. The amount of information

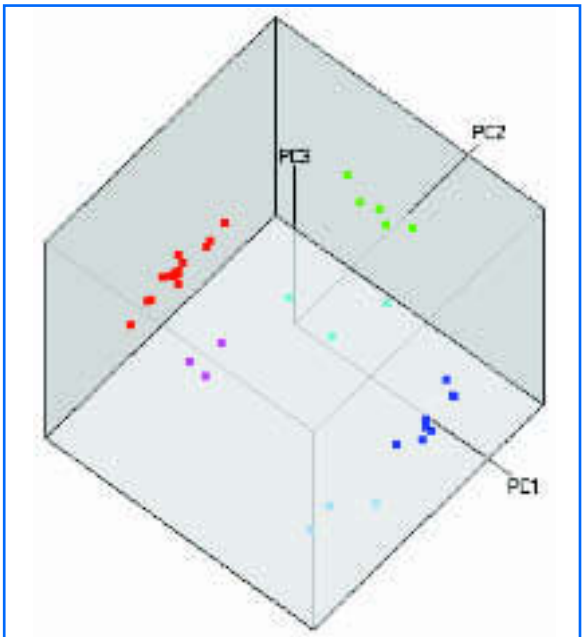


Figure 4: PC1 vs. PC2 vs. PC3 loadings plot of the genes. Same colour codes as in Figure 2

decreases with increasing degree of scaling. More information is represented when using raw data, but the information is biased to the more expressed and variable genes. Using the autoscaled data 23% of the information is missing. This is not necessarily a concern because the missing information has low correlation, since the importance of the PCs decreases with increasing index. Still, if more information is needed it can be shown in a PC1 vs. PC2 vs. PC3 scatterplot (Figure 4). This plot accounts for 85% of the information in the data and reveals subgroups in two of the original three groups.

Cluster analysis

The data were also classified by hierarchical cluster analysis using the unweighted pair method and the Euclidean distance.[6] Results obtained using raw, mean centred and autoscaled data are shown in Figure 5 (samples) and Figure 6 (genes). The dendrograms obtained from raw and mean centred data are identical, since mean centring does not change the relative distances between sample points in a multidimensional expression space. The dendrogram of the

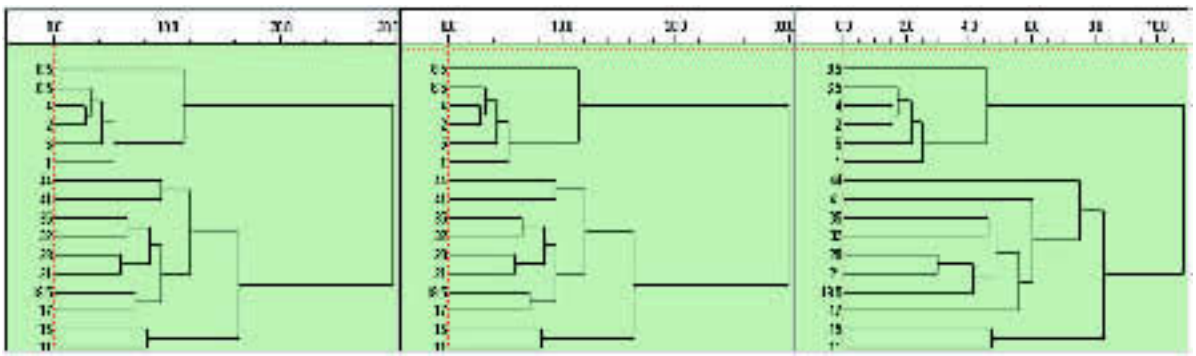


Figure 5: Hierarchical clustering of the stages. From left to right: raw, mean centred and autoscaled data

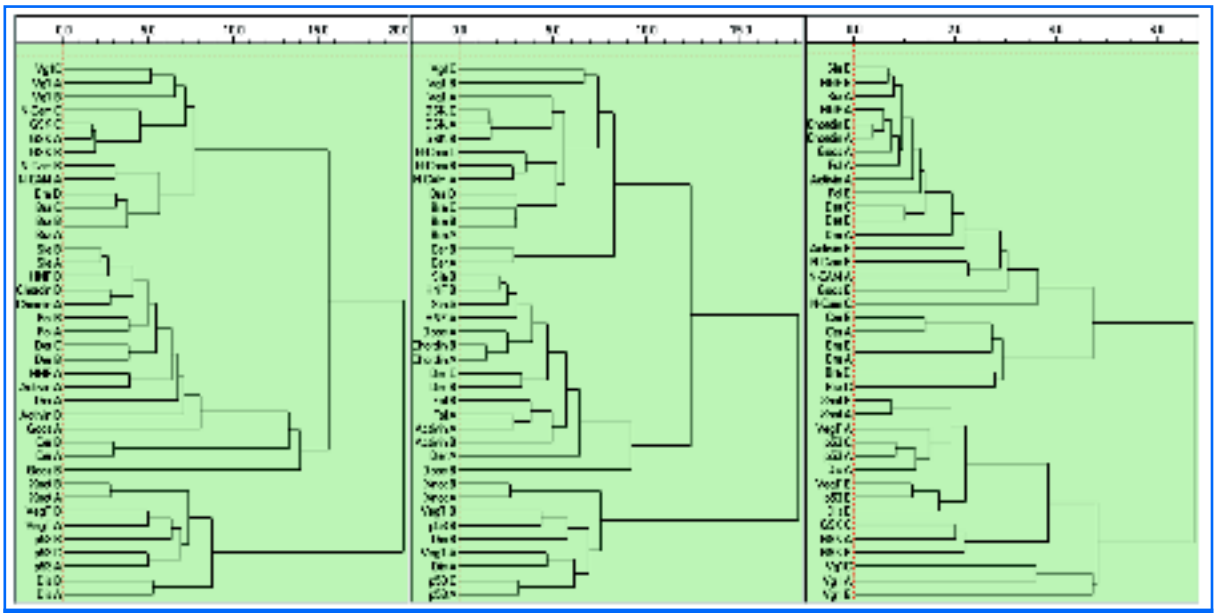


Figure 6: Hierarchical clustering of the genes. From left to right: raw, mean centered and autoscaled data

autoscaled data is different but has the same main features. Stages 1-8.5 form a group, where stages 1-6.5 are very similar. Stages 11 and 15 form the second group and stages 17-44 a third. Clustering of the genes is influenced only slightly by mean centring, while autoscaling has a larger effect (Figure 6). Still, the clusters reveal the same groups as found by the PCA.

Self-organising map

Finally, a rather new methodology was used to verify the findings above. It is based on a branch of mathematical techniques that do not require formal equations, but use rules to organise the data through a series of random events. One such technique is the Kohonen's self-organising map (SOM).[7] An example of a SOM based on the autoscaled data is shown in Figure 7. It clearly separates the six groups of genes supporting the conclusion that they have distinct expression profiles. □

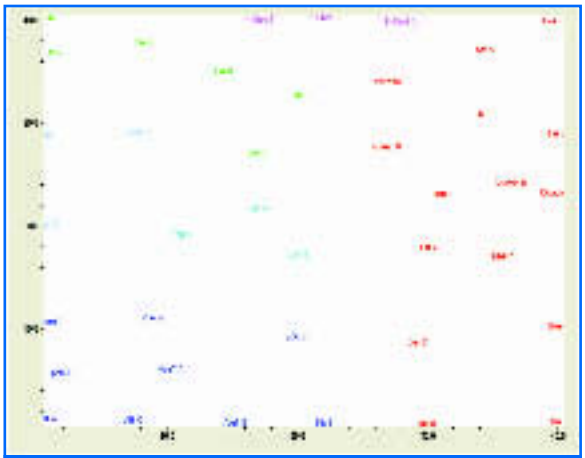


Figure 7: SOM trained with autoscaled data. The net is 40 x 40. It was trained 10000 epochs, with $\alpha = 0.1$. Colours are as in Figure 2

Acknowledgements

MK acknowledges a grant from the Spanish Ministry of Education and Science (SAB2005-0162). Part of this project was supported by the Carl Trygger foundation, European FP6 project SMARTHEALTH (FP6-2004-IST-NMP-2-016817), the grant agency of the AS CR no. B500520601 and by the project no. AVOZ 50520514 awarded by the AS CR.

References

1. *The Real-Time Polymerase Chain Reaction, M. Kubista, J.M. Andrade, M. Bengtsson, A. Forootan, J. Jonak, K. Lind, R. Sindelka, R. Sjöback, B. Sjögreen, L. Strömbom, A. Ståhlberg, N. Zoric, Molecular Aspects of Medicine (2006) 27, 95-125*

2. *Editorial, NATURE BIOTECHNOLOGY VOLUME 24 NUMBER 9 SEPTEMBER 2006, 1039.*

3. *Developmental expression profiles of Xenopus laevis reference genes. Sindelka R, Ferjentsik Z, Jonak J.,Dev Dyn. 2006 Mar;235(3):754-8.*

4. *Normal table of Xenopus laevis. Nieuwkoop PD and Faber J. 1994. Garland Publishing, Inc. New York & London*

5. *Principal Component Analysis, Jolliffe I. T., Series: Springer Series in Statistics. 2nd ed., 2002 (ISBN: 978-0-387-95442-4)*

6. *Andrew Moore: "K-means and Hierarchical Clustering – Tutorial Slides": http://www.autonlab.org/tutorials/kmeans.html*

7. *Self-Organizing Maps. Teuvo Kohonen, Series: Springer Series in Information Sciences Vol. 30, 3rd ed., 2001, (ISBN: 978-3-540-67921-9)*

Paper IV

M. Kubista, R. Sindelka, A. Tichopad, A. Bergkvist, D. Lindh, A. Forootan, The Prime Technique. Real-time PCR data analysis. G.I.T. Laboratory Journal, 2007. 9-10: p. 33-35.

The Prime Technique

Real-time PCR Data Analysis

For measuring gene expression there is only one technique: PCR. But how can it be used with maximum efficiency? This article tries to give the answer to that question.

Mikael Kubista, Institute of Molecular Genetics and TATAA Biocenter, Sweden

Radek Sindelka, Institute of Molecular Genetics, Czech Republic

Quantitative real-time PCR (QPCR) is today the prime technique to measure gene expression. [1] When properly used it offers unprecedented sensitivity, accuracy and reproducibility. But there are caveats. The target is mRNA, which must be extracted and converted to cDNA in a reverse transcription process that can proceed at highly variable yield depending on protocol. [2, 3] RNA is further rapidly degraded by nucleases abundant in biological samples. [4] To account for processing variation the expression of marker genes is normalised with appropriate endogenous control genes and technical repeats are performed to reduce confounding variance.

Typical QPCR experimental design is shown in figure 1. Studying the effect of treatment samples are collected of control and treated subjects. Each sample is divided into replicate RT reactions, which are split into replicates for QPCR. [2]

Each QPCR measurement generates a CT value, which is the number of amplification cycles required to reach a certain threshold signal level. CT values are inversely proportional to the logarithm of the initial number of target copies present in the sample. Figure 2 shows CT values in a spreadsheet. Expression of one marker gene (MG) and one reference gene (RG) was measured in 6 control and 6 treated subjects, with RT triplicates and QPCR duplicates. This gave a total of 72 samples. 36 were analyzed per run together with interplate calibrators. The treatment and the replicates are indexed in classification columns identified with labels that begin with a hatch. Another classification column indexes the sample amounts used. QPCR data are pre-processed as follows:

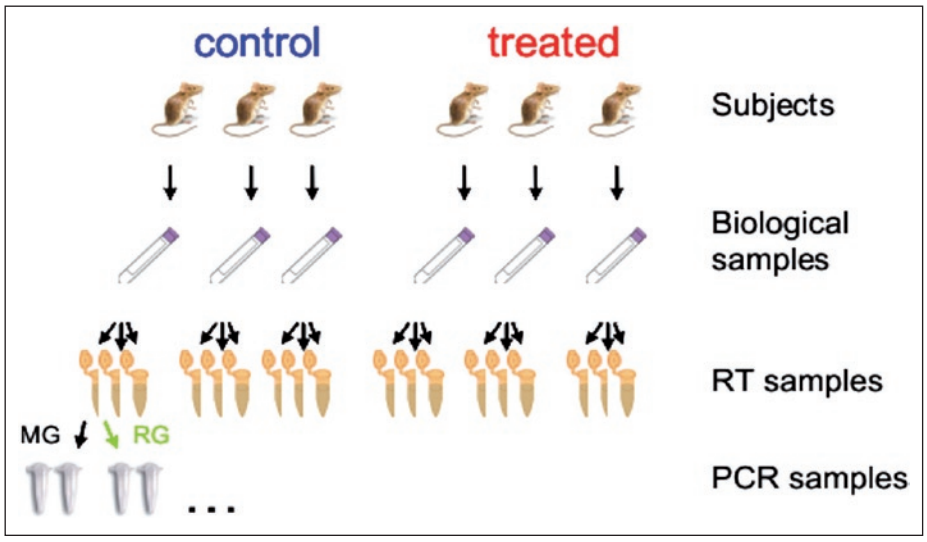


Fig. 1: Nested QPCR experimental design

Correct for Off-scale Measurements

Occasionally amplification response curves never reach threshold. Sometimes signal reaches threshold but is due to formation of aberrant products, such as primer-dimers. In either case we don't have reliable CT reading. Experiments containing such off-scale data should be analyzed with non-parametric methods that do not assume data are Normal distributed. [5] Non-parametric methods, however, are weaker than parametric in the sense that more replicates are needed to reach reliable conclusions. An alternative is to replace off-scale data with fictive CT values and use parametric tests. Fictive CT values are set to the highest CT observed for a truly positive sample, assumed to be the level of detection (LOD), plus 1. [1] This corresponds to assigning a concentration that is half of LOD to the off-scale samples. This is no more erroneous than assigning zero concentration to these samples, because we do not know that they are blank; we only know that they contain less amount of target than we are able to detect. If we are uncertain about the correction we repeat the analysis assigning CT(LOD) + 2 to the off-scale samples. If the result is virtually the same, we can be confident that the correction has negligible effect.

Efficiency Correction

PCR efficiency ($0 \leq E \leq 1$) depends on both the assay used and on the sample matrix. Estimations of PCR efficiency can be more or less advanced. [1] Once estimated, the measured CT values are corrected as:

$$CT_{E=100\%} = CT_E \frac{\log(1+E)}{\log(2)}$$

Variations between Runs

When all samples cannot be run in one plate correction is needed. This is done by including an identical sample, called interplate calibrator (IC), in all runs, which is analyzed for all genes.

$$CT_{plate\ norm} = CT_r - CT_r^{IC} + \frac{1}{m}\sum_{r=1}^{m} CT_{IC}$$

Sample Amount

For samples based on different starting amounts CT values are corrected as:

$$CT_{conc=1} = CT_{conc} - \log_2(conc)$$

Here is a typo! Correct equation has a plus sign. See GenEx manual (www.multid.se)

QPCR Technical Repeats

QPCR technical repeats are averaged before normalisation with reference genes.

$$CT_{QPCR\_average} = \frac{1}{n}\sum_{i=1}^{n} CT_{QPCR\_repeats}$$

Reference Genes

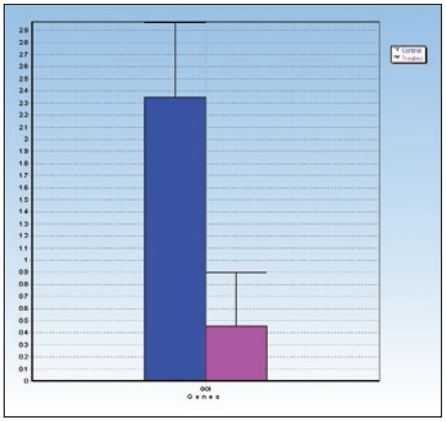
Expression of marker genes is normalised to that of reference genes:

$$CT_{MG,norm} = CT_{MG} - \frac{1}{n}\sum_{i=1}^{n} CT_{RG}$$

RT Technical Repeats

RT technical repeats are averaged after normalisation with reference genes.

$$CT_{RT\_average} = \frac{1}{n}\sum_{i=1}^{n} CT_{RT\_repeats}$$



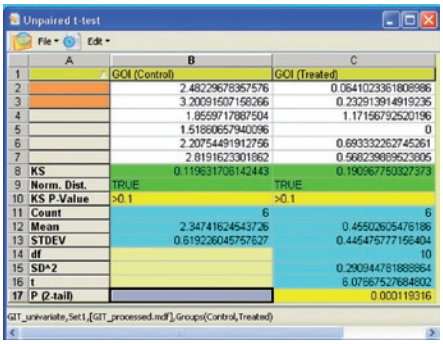
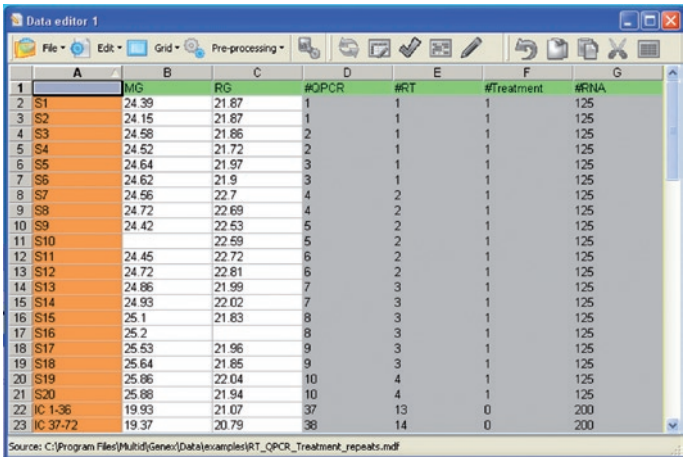


Fig. 3: Top: Bar graph showing mean with 95 % confidence interval for control and treated samples. Bottom: Result of unpaired 2-tailed t-test.

Fig. 2: Data prepared for analysis. Classification columns, identified by a hatch, index technical and biological replicates and also contain additional information needed for data pre-processing. IC is interplate calibrator.



Relative Quantities

Relative expression among samples is calculated as:

$$RQ = 2^{CT_{ref} - CT}$$

Log Scale

Data shall be Normal distributed for analysis with tests such as the t-test, linear regression, and ANOVA. Gene expression data are usually not Normal distributed when expressed as relative quantities, but often become Normal distributed by logarithmic transformation to fold differences (FD). Traditionally log base 2 is used:

$$FD = \log_2(RQ)$$

Data in figure 1 were pre-processed assuming 90 % PCR efficiency for the marker gene and 95 % efficiency for the reference gene. When averaging the technical repeats missing data were accounted for.

After pre-processing the FD's of the biological repeats were averaged and are shown in figure 3 with 95 % confidence interval. They were confirmed to be Normal distributed by the Kolmogorov-Smirnov's test, and the means were compared with the unpaired 2-sided t-test (fig. 3). This gave P = 0.00012 so we conclude that treatment most likely has effect on the expression of the marker gene. Analyzing the same data with the non-parametric Mann-Whitney's test we still find the difference significant, but with lower confidence (P_{MW} = 0.005).

Multiple Genes

The procedure can be used to compare the effect of treatment on several genes. However, such multiple testing is statistically unsound. The P-value is the probability to observe a difference that is at least as large as the measured in absence of treatment effect. It is up to the investigator to decide how low a P-value shall be taken as indicator of treatment effect. Often investigators work with 95 % confidence, which for a single test translates to P ≤ 0.05. With this criterion one out of 20 studies with no effect will turn out as a false positive. Usually this is acceptable error rate. But what happens if multiple genes are compared? If 10 genes are compared in two identical samples probability that all 10 show a differences below threshold of P = 0.05 is 0.95^10 = 0.60, or 60 %. Hence, the probability that at least one out of them gives P ≤ 0.05 is 40 %! This problem of multiple testing becomes more and more important with increasing number of genes, and has led to serious criticism of microarray gene-expression profiling studies. [6]

Multivariate Expression Profiling

The proper way to analyze data based on expression of multiple genes is with multivariate methods. [1, 7] Most common are Principal

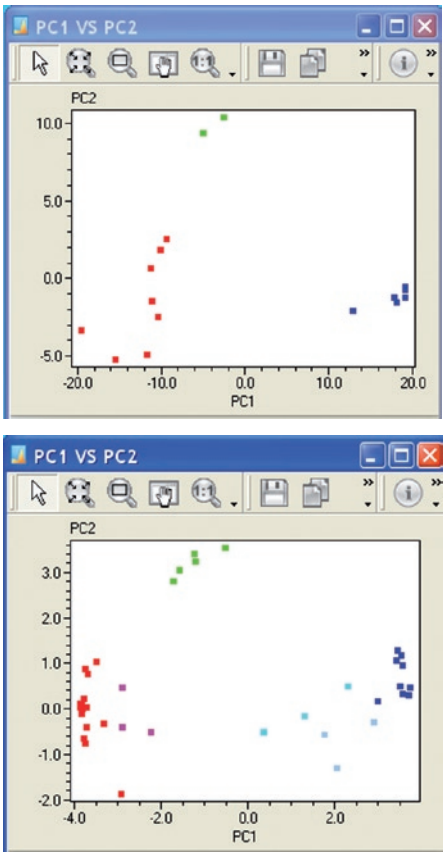


Fig. 4: PCA classification of *Xenopus laevis* expression data presented in PC1 vs. PC2 scatter plots. Top panel shows samples and bottom panel shows genes. Biological replicates are shown as independent symbols in the left panel.

Component Analysis (PCA), Hierarchical clustering, and the Self-Organised Map. With these methods multiple samples, each studied by measuring the expression of multiple genes, can be readily analyzed to identify those samples and those genes that have common expression behaviour. The data are preprocessed as above, with one additional step.

Mean Centering/Autoscaling

To remove the effect of the overall expression of the different genes the mean expression of every gene is subtracted.

$$FD_{MC} = FD - \overline{FD}$$

Hence, for mean centered data a certain increase in expression has the same significance independently of the absolute expression level of that gene. Mean-centred data are mainly used for the classification of samples.

To remove also the effect of the magnitude of the change, the data are further divided with the standard deviation:

$$FD_{AS} = \left(FD - \overline{FD}\right) / SD = FD_{MC} / SD$$

The autoscaled data are mainly used to classify genes.

Expression of sixteen genes was measured during sixteen developmental stages, ranging from the oocyte to the tadpole, of the frog *Xenopus laevis.* The data were processed as described above, though without normalisation with reference genes, since no genes with constant expression during *Xenopus laevis* development has been identified. [8] This is possible when data are scaled. [1] Top panel in figure 4 shows classification of the developmental stages based on PCA of mean-centered data and bottom panel shows classification of the genes based on autoscaled data. Both panels reveal that development goes through three distinct stages.

Acknowledgement

Material used is from TATAA Biocenter (www.tataa.com) biostatistics courses in QPCR. Data were analyzed with GenEx software from MultiD Analyses (www.multid.se). The example data are available on: www.multid.se/downloadspage.html.

References

[1] Kubista M. *et al.:* Molecular Aspects of Medicine 27, 95–125 (2006)

[2] Ståhlberg A. *et al.:* Clin. Chem. 50, 509 (2004)

[3] Ståhlberg A. *et al.:* Clin. Chem. 50, 1679 (2004)

[4] Fleige S. and Pfaffl M.W.: Molecular Aspects of Medicine 271, 26–139 (2006)

[5] Intuituve Biostatistics, Harvey Motulsky, Oxford University Press, New York, ISBN: 0-19-50-8607-4, 1995

[6] Ioannidis J.P.: Lancet 365, 454–455 (2005)

[7] Kubista M. *et al.:* European Pharmaceutical Review 56, 1 (2007)

[8] Sindelka R. *et al.:* Dev Dyn.235, 754–758 (2006)

Authors:

Mikael Kubista, Laboratory of Gene Expression, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic and TATAA Biocenter AB

Radek Sindelka, Laboratory of Gene Expression, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Prague, Czech Republic

Ales Tichopad, Physiologie Weihenstephan, Technical University Munich, Germany

Anders Bergkvist, MultiD Analyses AB, Gothenburg, Sweden

Daniel Lindh, MultiD Analyses AB, Gothenburg, Sweden

Amin Forootan, MultiD Analyses AB, Gothenburg, Sweden

▶ Contact

Mikael Kubista
TATAA Biocenter AB
Göteborg, Sweden
mikael.kubista@tataa.com

Paper V

R. Sindelka, J. Jonák, R. Hands, SA. Bustin, M. Kubista, Intracellular expression profiles measured by real-time PCR tomography in the *Xenopus laevis* oocyte. Nucleic Acids Res, 2008. 36(2), 387-92.

Published online 26 November 2007

Nucleic Acids Research, 2008, Vol. 36, No. 2 ***387–392***
doi:10.1093/nar/gkm1024

Intracellular expression profiles measured by real-time PCR tomography in the *Xenopus laevis* oocyte

Radek Sindelka[1], Jiri Jonák[1,4], Rebecca Hands[2], Stephen A. Bustin[2] and Mikael Kubista[1,3,*]

[1]Laboratory of Gene Expression, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Videnska 1083, 14220 Prague 4, Czech Republic, [2]Barts and the London Queen Mary's School of Medicine and Dentistry, Academic Surgery, London, E1 1BB, UK, [3]TATAA Biocenter, Medicinargatan 8A, 413 46 Göteborg, Sweden and [4]Institute of Medical Biochemistry, First Medical Faculty, Charles University, Katerinska 1083, 12000 Prague 2, Czech Republic

Received October 9, 2007; Revised October 27, 2007; Accepted October 28, 2007

ABSTRACT

Real-time PCR tomography is a novel, quantitative method for measuring localized RNA expression profiles within single cells. We demonstrate its usefulness by dissecting an oocyte from *Xenopus laevis* into slices along its animal–vegetal axis, extracting its RNA and measuring the levels of 18 selected mRNAs by real-time RT-PCR. This identified two classes of mRNA, one preferentially located towards the animal, the other towards the vegetal pole. mRNAs within each group show comparable intracellular gradients, suggesting they are produced by similar mechanisms. The polarization is substantial, though not extreme, with around 5% of vegetal gene mRNA molecules detected at the animal pole, and around 50% of the molecules in the far most vegetal section. Most animal pole mRNAs were found in the second section from the animal pole and in the central section, which is where the nucleus is located. mRNA expression profiles did not change following *in vitro* fertilization and we conclude that the cortical rotation that follows fertilization has no detectable effect on intracellular mRNA gradients.

INTRODUCTION

A single egg contains all the information required for its proliferation and differentiation into a complete organism and accurate spatial distribution of maternal factors is a critical issue for early development, cell determination, differentiation and germ layers formation (1). All mRNAs translated during the initial stages of development originate from the mother as transcription of new zygotic mRNA is initiated only after 12 cell divisions during what is called the midblastula transition (MBT).

The cellular distribution of maternal factors and their functions are usually studied in model organisms such as *Drosophila melanogaster*, *Caenorhabditis elegans* and *Mus musculus*. These studies are hampered by the very small amounts of RNA (~pg of total RNA) in invertebrate and mammalian cells. In contrast, the egg from the African clawed frog *Xenopus laevis* contains a microgram of total RNA. Furthermore, two differently coloured hemispheres can easily be distinguished in *Xenopus* eggs. The colouration difference identifies the first developmental animal–vegetal, A–V, axis, which is formed during mid- and late stages of oogenesis. The dark pigmented animal hemisphere derives its colour from the pigmented melanosomes and contains the egg nucleus (2), whereas the opposite light vegetal hemisphere contains yolk platelets. During early development, the animal hemisphere is transformed into ectodermal cells with epidermis and neural fate. The vegetal hemisphere follows endodermal fate (gut) and the marginal zone forms mesoderm layer with blood, bone and muscle cell types.

Maternal factors are distributed along the A–V axis during oogenesis and have many different roles in *Xenopus* early development. Some are transcription factors, while others are signalling factors or regulators of activity of signalling molecules (3). Two groups of mRNA molecules have been reported to localize in the vegetal hemisphere during oogenesis. Germ cell determinants such as Xcad2 (Nanos related, Zn finger protein), Xpat (unknown function), DeadSouth (RNA helicase) and mRNAs for the Wnt11 (Wnt family member) gene are vegetally localized in early stages 1 and 2 by the METRO

*To whom correspondence should be addressed. Tel: +46 31 741 18 00; Fax: +46 31 741 17 01; Email: mikael.kubista@img.cas.cz

(messenger transport organizer) pathway (2–8). A second group of vegetal genes includes VegT (T-box transcription factor), Otx1 (a homeobox gene) and Vg1 (TGF-beta family member). These localize vegetally by cytoskeletal-based transportation during later stages of oogenesis (8,9). Other genes, such as Oct60 (transcription factor, POU family), An1 (Ubiquitin like fusion protein), An2 (Mitochondrial ATPase subunit), Ets1 and Ets2 (transcription factors, ETS family members) and XPar-1 (serine/threonine kinase) have been found localized to the animal pole (3,9–11). The mechanism behind this specific localization is not well understood.

During fertilization, *Xenopus* sperm enters the egg through the animal hemisphere and the point of entry can be distinguished by a change in cortex cytoskeleton structure that leads to a local change in pigmentation. The process, called cortical rotation, occurs some 25 min after fertilization. The cortex rotates by about 30° and alters A–V organization through cytoskeletal and cytoplasmic rearrangements (12,13). The cortex movement induces local redistribution of β-catenin protein and the β-catenin stabilizing agent to a site opposite to the sperm's entrance (14,15). Accumulated β-catenin proteins then induce local gene expression of some zygotic factors including siamois and Xnr3. These are important for the formation of the organizer, which defines the future dorsal site of the embryo.

We have previously shown that there is substantial variation in gene expression among seemingly homogeneous cells (16). In this work we describe a novel approach to study intracellular expression profiles by real-time PCR tomography (17,18).

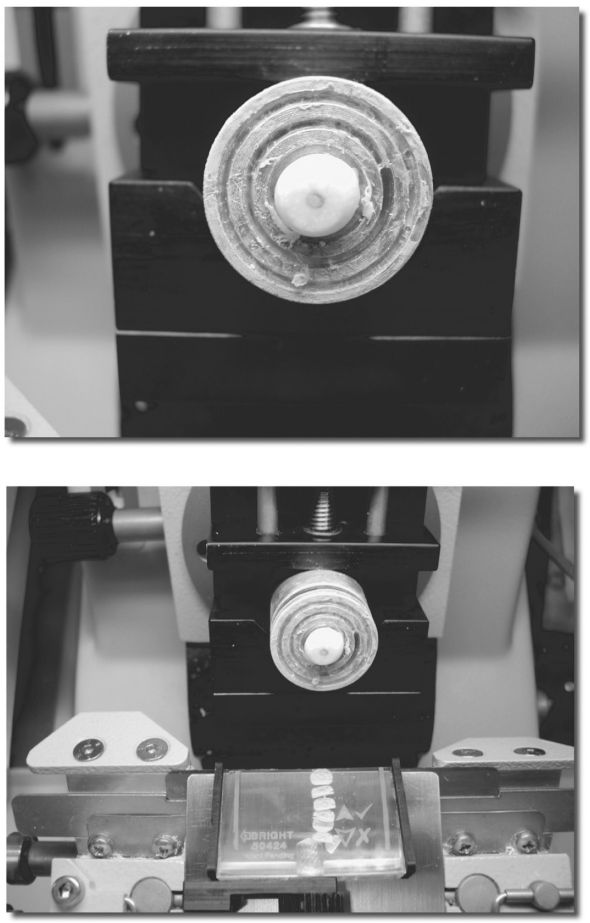


Figure 1. Photographs showing the process of preparing material for real-time PCR tomography. (**A**) The *Xenopus laevis* oocyte imbedded in OCT is mounted in a cryostat. (**B**) The material is sliced for subsequent analysis by real-time RT-PCR.

MATERIALS AND METHODS

Xenopus laevis females were stimulated by hCG (human chorionic gonadotrophin) injection and *in vitro* fertilized (IVF) eggs were incubated at 22°C. The eggs were not treated with cystein, which is common procedure, because the treatment compromises manipulation of the eggs and RNA stability after defrosting of the material. Only eggs that turned round with the animal pole on the top were harvested for sectioning. More than 90% of the turned eggs divided within 90 min following IVF. Four types of eggs were collected: unfertilized eggs and eggs collected at 25, 50 and 85 min post- IVF. First cell division occurs after 90 min. The collected eggs were frozen at –70°C and stored.

For analysis the eggs were embedded in optimum cutting temperature (OCT) compound and dissected into 35 slices (30 μm) across the A–V axis (Figure 1). Consecutive slices were pooled into five groups with seven slices in each. From each group, 200–500 ng of total RNA was extracted using RNeasy Micro kit (Qiagen). RNA concentrations were determined with the Nanodrop® ND1000 quantification system (Nanodrop Inc.) and RNA quality was assessed with the 2100 Bioanalyzer using the RNA Pico Chip (Agilent). In general RNA quality was very high. Total RNA was reverse transcribed (High Capacity cDNA Archive Kit- Applied Biosystems) using 100 ng total RNA with 2.5 μl of random primers in water in a total volume of 16.5 μl. The mixture was incubated for 10 min at 72°C. After cooling to room temperature, 1 μl of dNTPs (25×), 2.5 μl of 10× reverse transcription buffer and 2.5 μl of MultiScribeReverse Transriptase (50 U/μl) were added. The mixture was incubated for 2 h at 37°C. cDNA was diluted to a final volume of 100 μl. Real-time PCR assays had a final volume of 25 μl and contained 3 μl of cDNA, 1U SureStart *Taq* DNA polymerase (Stratagene-Europe), 2.5 μl of reaction buffer (10×), 3 mM $MgCl_2$, 0.4 mM dNTP mix, 50 000-fold diluted SYBRGreen I (Molecular Probes), 25 000-fold diluted ROX reference dye and 0.3 mM primers. PCR was run in a Mx3005P (Stratagene) with cycling conditions: 95°C for 12 min, 45 cycles at 95°C for 20 s, 60°C for 25 s, 72°C for 30 s. After cycling the samples were heated to 95°C for 1 min, and melting curve was recorded between 65 and 95°C.

Gene-expression data were analyzed using GenEx software from MultiD Analysis (www.multid.se) and Prism4 from Graphpad (www.graphpad.com). It was not possible to use any internal reference genes for normalization, since this is the first time intracellular mRNA

levels are being quantitated and there is no information on what mRNAs might be homogeneously distributed within the cell. Consequently, we normalized individual mRNAs against the total amount of RNA used for reverse transcription, essentially measuring gene-expression levels relative to the total amount of RNA in each section. Since RNA yield is rather uniform, we assume that total RNA is distributed homogenously in the cell. Consequently, normalization is against the volume of the segments, thus accounting for the differences in segment sizes due to the spherical shape of the cell. Although the data are perfectly comparable within each section, there may be bias across sections due to variations in the density of total RNA and in reverse transcription yields. These are caused by the sample matrix, which is quite different in the animal and the vegetal poles of the oocyte. The real-time PCR CT values were converted to relative quantities assuming 100% PCR efficiency, and the amounts of transcripts in the five egg sections were expressed as the fractions of the mRNA molecules found in each of five segments along the A–V axis in the *Xenopus* oocyte:

$$x_j = \frac{2^{-\mathrm{CT}_j}}{\sum_{i=1}^{5} 2^{-\mathrm{CT}_i}}$$

CT_j is the CT determined for section 'j' of the oocyte and x_j is the fraction of the mRNA found in this section. Since the amounts of mRNAs in the five sections are of the same order of magnitude, the assumption of 100% PCR efficiency will have little effect on the calculated intracellular mRNA profiles. The initial normalization against total RNA ensures that the profiles reflect true variations in the levels of the mRNAs along the A–V axis of the cell.

The conventional real-time PCR results were confirmed for selected genes with digital PCR using the BIOMARK digital array from Fluidigm (www.fluidigm.com). The array is designed to accept 12 sample mixtures, which each is partitioned into a different 765-chamber grid. One step RT–qPCR was performed directly on the chip. Ten-microliter reaction mix was loaded onto the chip, containing 3.4 µl of total RNA, 0.5 µl SuperScript RT/Taq (CellsDirect qPCR-RT kit, Invitrogen), 1 µl buffer containing ROX, 1 µl of primers (9 µM) and FAM-labelled TaqMan probe (2 µM) and 0.1 µl of Tween (10%). The input amount of total RNA was tuned to produce less cDNA molecules than the number of chambers. The mixture was distributed into the 765 chambers, incubated for 15 min at 50°C for reverse transcription and then analyzed by PCR, starting with HotStart activation at 95C° for 2 min followed by 45 PCR cycles at 95°C for 15 s and 60°C for 30 s. FAM/ROX fluorescence signal was collected at the end of each cycle, and the number of chambers that gave positive fluorescence signal after 40 cycles was registered. Assuming Poisson distribution of the cDNA molecules in the chambers, the average number of cDNA molecules per chambers is given by $\mathrm{In}\{[1-P(x\geq 1)]^{-1}\}$, where $P(x\geq 1)$ is the fraction of positive PCR reactions. A sample distributed into 765 chambers thus contained a total of $765\times \mathrm{In}\{[1-P(x\geq 1)]^{-1}\}$ cDNA molecules. The number of mRNA molecules in the sample can then be grossly estimated assuming 80% cDNA synthesis yield in the reverse transcription reaction (19).

RESULTS

Expression levels of mRNAs specified by the Wnt11, FoxH1, VegT, Vg1, Oct60, GSK-3β, dishevelled, elongation factor-1α (EF-1α), Xdazl, Xmam, Tcf-3, GAPDH, β-catenin, Xcad2, Otx1, XPar-1, Deadsouth and Stat3 genes were all characterized by distinct and reproducible intracellular gradients. As an example, Figure 2A and B shows Vg1 and Oct60 intracellular mRNA gradients measured on eggs from four different females. Oct60 is predominantly found at the animal pool, while Vg1 is preferably found at the vegetal pool. Although there is variation among individual cells, the intracellular gradients are clearly observed against the biological variation of the females, as reflected by the standard errors of the means. Figure 2C and D also shows mRNA intracellular distributions for Vg1 and Oct60 prior to IVF, and at 20, 55 and 85 min after IVF. Statistical analysis using two-way ANOVA with Bonferroni post-test on a pairwise comparison of the profile of the unfertilized oocyte with mRNA profiles collected at different time points after fertilization revealed that the correlation between segment and mRNA level is extremely significant ($P < 0.0001$), but that there is no effect of fertilization and time following fertilization ($P \approx 1$).

The mRNA profiles of 15 genes characterized in at least six eggs are shown in Figure 3A. The profiles fall into two distinct classes, and are characteristic of animal and vegetal locations, respectively. The mRNAs located preferentially at the animal pole are FoxH1, Oct60, GSK-3β, dishevelled, EF-1alpha, Xmam, Tcf-3, GAPDH, β-catenin and XPar-1. Those located at the vegetal pole are VegT, Vg1, Xdazl, Wnt11 and Otx1. In addition, Stat3 was measured in four cells and found to be located in the animal hemisphere, while Xcad2 (measured in four cells) and Deadsouth (measured in three cells) were vegetally located (data not shown). For Oct60 (animal) and Wnt11 (vegetal), the intracellular expression profiles measured by QPCR tomography were confirmed with digital PCR (Figure 4). Oct60 shows highest expression in Sections 2 and 3 from the animal pole, while Wnt11 expression is largest in Section 5, which is closest to the vegetal pole. Qualitatively, this is in agreement with the real-time PCR results in Figure 3. Assuming there are no important differences, we calculated the average vegetal and animal mRNA profiles also shown in Figure 3B. The data are based on 117 measured vegetal profiles and 166 measured animal profiles. The error bars represent ±1 SD, within which 68% of the measured values should be found. The standard errors of the means were insignificant, and the average values shown by the symbols have negligible errors.

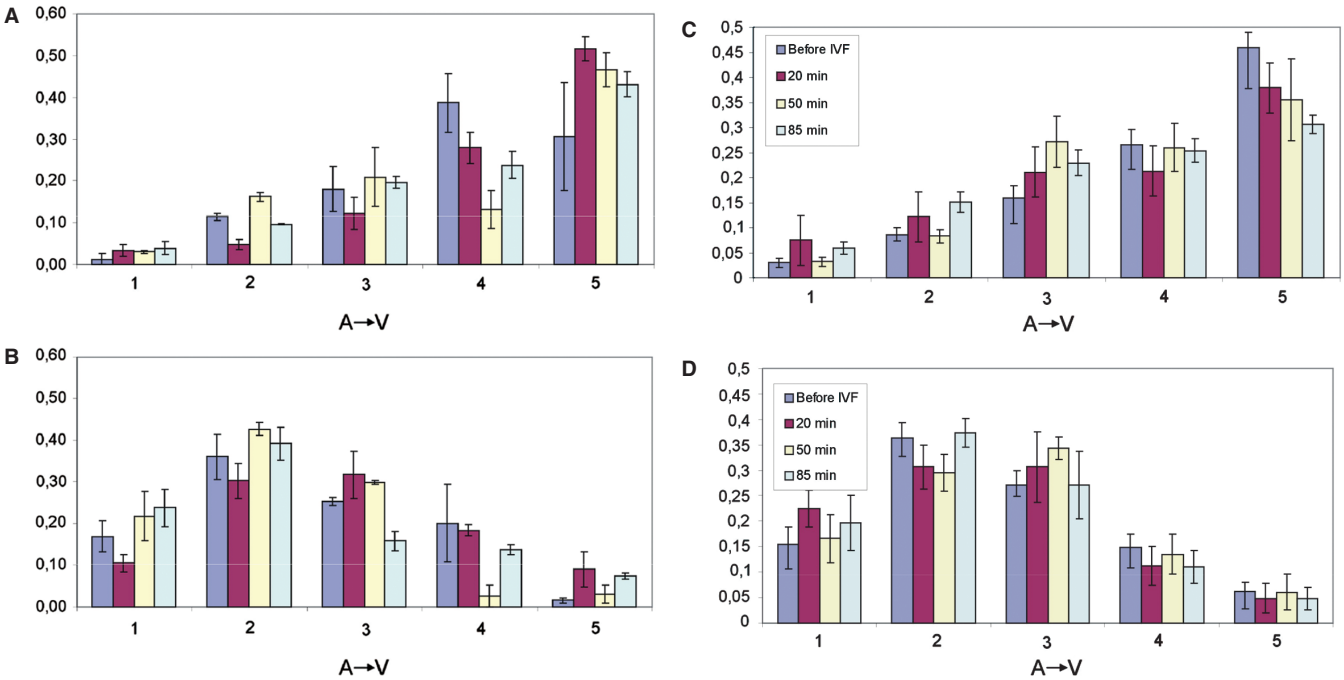


Figure 2. Intracellular gradients (A→V) of mRNA levels in *Xenopus laevis* oocytes. Distribution of (**A**) Vg1 and (**B**) Oct60 expression density along the oocyte animal–vegetal axis. The RNA was prepared from two to three individual eggs (standard error of the means indicated by error bars) from four different females (indicated by regular bars). Effect of fertilization. Distribution of (**C**) Vg1 and (**D**) Oct60 along the animal–vegetal axis. RNA was prepared from at least six eggs before IVF and at 20, 50 and 85 min after fertilization. Error bars indicate standard error of the means.

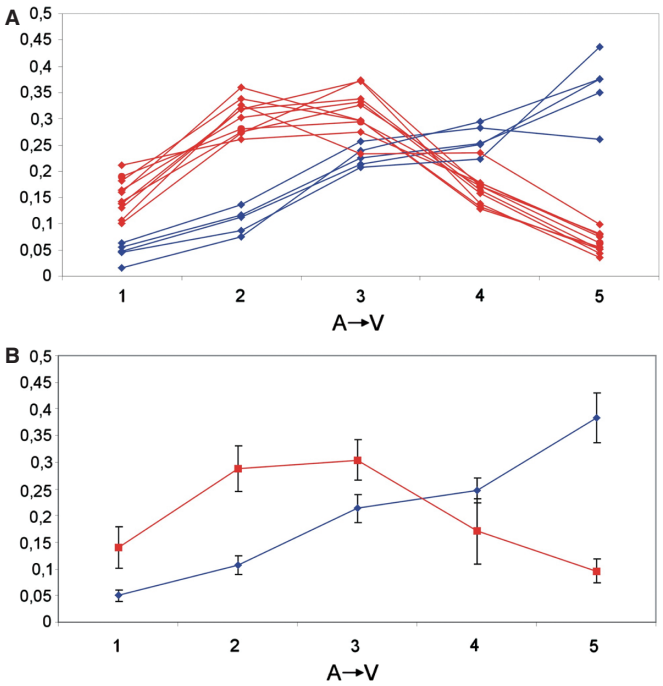


Figure 3. (**A**) Averaged intracellular mRNA concentration profiles (A→V) for genes studied in at least six eggs. Animal genes (FoxH1, Oct60, GSK-3β, dishevelled, EF-1alpha, Xmam, Tcf-3, GAPDH, β-catenin and XPar-1) are shown in red and vegetal genes (VegT, Vg1, Xdazl, Wnt11 and Otx1) are shown in blue. (**B**) Average expression profiles of all vegetal (red) and all animal (blue) genes. The error bars indicate 1 SD.

DISCUSSION

This is the first report of sub-cellular expression profiling and quantification of mRNA within a single cell. Using real-time PCR, which is currently the most sensitive and reliable technique for quantitative mRNA analysis, we measured the intracellular profiles of selected developmental mRNAs within the *X. laevis* oocyte. Our results reveal the existence of characteristic expression gradients, and demonstrate that real-time PCR tomography is highly suitable for measuring them quantitatively. Out of the 18 genes studied, 11 were found preferentially located at the animal pole (animal genes), while seven were preferentially located at the vegetal pole (vegetal genes). The 'animal genes' were FoxH1, Oct60, GSK-3β, dishevelled, EF-1α, Xmam, Tcf-3, GAPDH, β-catenin, XPar-1 and Stat3. Oct60 has previously been found located at the animal pole by *in situ* hybridization (10). EF-1α and GAPDH have been ascribed housekeeping functions and used as reference genes (20). However, they show clear animal location. Interestingly, APC, β-catenin, Fz7, GSK-3β, dishevelled and Tcf-3, which specify components of the Wnt pathway are animal genes, whereas Wnt11 itself shows vegetal location. Xmam and FoxH1 have not been localized previously. The genes found to have vegetal location were VegT, Vg1, Xdazl, Wnt11, Otx1, Deadsouth and Xcad2.

Within the resolution of our technique all genes contained in each of the two groups had comparable profiles. The animal genes were preferentially found in the

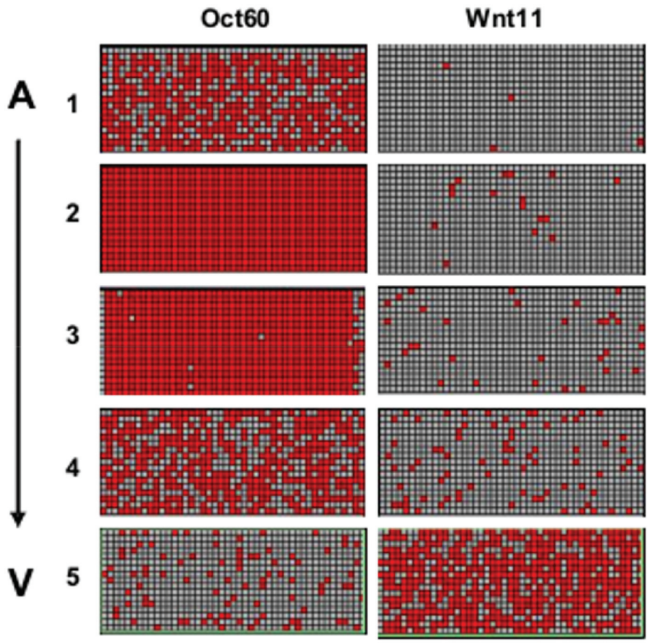


Figure 4. Digital PCR of Wnt11 and Oct60 showing abundance of transcripts in the five oocyte segments (A→V). mRNA from each segment was distributed into 765 chambers that were analyzed by RT-PCR. Red indicates positive PCR for targeted product.

second and third (central) sections of the oocytes, while the vegetal genes were found preferentially in the fourth and fifth sections. However, although the polarization of the vegetal genes is much stronger than the (opposite) polarization of the animal genes, that polarization is not total. About 5% of mRNA molecules of the vegetal genes were found in the first section taken from the opposite pole and another 10% in the second section (Figure 3B). Hence, the extreme polarization of both animal and vegetal genes seen by *in situ* hybridization techniques, where virtually all genes are located at either pole (9), is not supported by our observations. Instead our data suggest that although there is a distinct bias to the location of the mRNA, it is distributed more evenly. The reason for this discrepancy is unclear; however, we note that the cell nucleus is expected to be located in sections two and three, which is where we find the animal genes to be most abundant. Perhaps most of the animal mRNAs are still located within the nucleus and are held there until their translation is required. Interestingly, fertilization of the oocytes and the cortical rotation that follows has no detectable effect on the intracellular mRNA gradients.

In summary, real-time PCR tomography can measure intracellular mRNA gradients more sensitively and with greater resolution than traditional *in situ* hybridization. In the present work, each cell was cut into 35 30 µm slices, yielding up to 75 ng of RNA per slice. This is not close to any limit, since a regular cryostat can easily cut slices of 10 µm, yielding some 100 slices from a single *X. laevis* oocyte. This would allow the generation of mRNA profiles with much higher resolution, the only potential constraint being the amount of RNA extracted from each slice. However, the use of appropriate multiplexing and/or pre-amplification techniques should help overcome this limitation. Other applications of real-time PCR tomography are readily envisaged: the localization of nuclei through genomic DNA, of mitochondria through mitochondrial DNA and of translationally active sites through ribosomal RNA. The techniques can also be used to localize viruses and bacteria in tissue sections.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by grant No. B500520601 project No. AVOZ 50520514 awarded by the AS CR, grant from the Swedish Research Council and by Carl Tryggers foundation. We are grateful to the charity Bowel and Cancer Research for support. We also thank Martin Pieprzyk for valuable help with digital PCR measurements. Funding to pay the Open Access publication charges for this article was provided by the Institutional Research Concept No. AV0Z50520701 of the Academy of Sciences of the Czech Republic.

Conflict of interest statement. None declared.

REFERENCES

1. Shav-Tal,Y. and Singer,R.H. (2005) RNA localization. *J. Cell. Sci.*, **118**(Pt 18), 4077–4081.
2. Danilchik,M.V. and Gerhart,J.C. (1987) Differentiation of the animal-vegetal axis in Xenopus laevis oocytes. I. Polarized intracellular translocation of platelets establishes the yolk gradient. *Dev. Biol.*, **122**, 101–112.
3. Heasman,J. (2006) Maternal determinants of embryonic cell fate. *Semin. Cell Dev. Biol.*, **17**, 93–98.
4. Gerhart,J.C., Vincent,J.P., Scharf,S.R., Black,S.D., Gimlich,R.L. and Danilchik,M. (1984) Localization and induction in early development of Xenopus. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **307**, 319–330.
5. MacArthur,H., Bubunenko,M., Houston,D.W. and King,M.L. (1999) Xcat2 RNA is a translationally sequestered germ plasm component in Xenopus. *Mech. Dev.*, **84**, 75–88.
6. MacArthur,H., Houston,D.W., Bubunenko,M., Mosquera,L. and King,M.L. (2000) DEADSouth is a germ plasm specific DEAD-box RNA helicase in Xenopus related to eIF4A. *Mech. Dev.*, **95**, 291–295.
7. Chang,P., Perez-Mongiovi,D. and Houliston,E. (1999) Organisation of Xenopus oocyte and egg cortices. *Microsc. Res. Tech.*, **44**, 415–429.
8. Zhou,Y. and King,M.L. (2004) Sending RNAs into the future: RNA localization and germ cell fate. *Iubmb Life*, **56**, 19–27.
9. King,M.L., Messitt,T.J. and Mowry,K.L. (2005) Putting RNAs in the right place at the right time: RNA localization in the frog oocyte. *Biol. Cell*, **97**, 19–33.
10. Mowry,K.L. and Cote,C.A. (1999) RNA sorting in Xenopus oocytes and embryos. *FASEB J.*, **13**, 435–445.
11. Ossipova,O., He,X. and Green,J. (2002) Molecular cloning and developmental expression of Par-1/MARK homologues XPar-1A and XPar-1B from Xenopus laevis. *Mech. Dev.*, **119**(Suppl. 1), S143–S148.
12. Gerhart,J., Danilchik,M., Doniach,T., Roberts,S., Rowning,B. and Stewart,R. (1989) Cortical rotation of the Xenopus egg:

consequences for the anteroposterior pattern of embryonic dorsal development. *Development*, **107**(Suppl), 37–51.
13. Vincent,J.P. and Gerhart,J.C. (1987) Subcortical rotation in Xenopus eggs: an early step in embryonic axis specification. *Dev. Biol.*, **123**, 526–539.
14. Heasman,J., Kofron,M. and Wylie,C. (2000) Beta-catenin signaling activity dissected in the early Xenopus embryo: a novel antisense approach. *Dev. Biol.*, **222**, 124–134.
15. Wylie,C., Kofron,M., Payne,C., Anderson,R., Hosobuchi,M., Joseph,E. and Heasman,J. (1996) Maternal beta-catenin establishes a 'dorsal signal' in early Xenopus embryos. *Development*, **122**, 2987–2996.
16. Bengtsson,M., Stahlberg,A., Rorsman,P. and Kubista,M. (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res.*, **15**, 1388–1392.
17. Kubista,M., Andrade,J.M., Bengtsson,M., Forootan,A., Jonak,J., Lind,K., Sindelka,R., Sjoback,R., Sjogreen,B. *et al.* (2006) The real-time polymerase chain reaction. *Mol. Aspects Med.*, **27**, 95–125.
18. Nolan,T., Hands,R.E. and Bustin,S.A. (2006) Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.*, **1**, 1559–1582.
19. Ståhlberg,A., Kubista,M. and Pfaffl,M. (2004) Comparison of reverse transcriptases in gene expression analysis. *Clin. Chem.*, **50**, 1678–1680.
20. Sindelka,R., Ferjentsik,Z. and Jonak,J. (2006) Developmental expression profiles of Xenopus laevis reference genes. *Dev. Dyn.*, **235**, 754–758.

Paper VI

Z. Ferjentsik, R. Sindelka, G. Lin, J. Jonák, Expression patterns of Src-family tyrosine kinases during *Xenopus laevis* development, Int. J. Dev. Biol., 2008.

Expression patterns of Src-family tyrosine kinases during *Xenopus laevis* development

Zoltán Ferjentsik[1, 2], Radek Šindelka[1], Jiří Jonák[1, 3*]

[1]Laboratory of Gene Expression and Laboratory of Bacteriology, Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Vídeňská 1083, Prague 4, 14220, Czech Republic

[2]Division of Cell and Developmental Biology, University of Dundee, Dow Street, DD1 5EH Dundee, UK

[3]Institute of Medical Biochemistry, First Medical Faculty, Charles University, Kateřinská 32, Prague 2, 12000, Czech Republic

[*]**Corresponding author:**
Jiří Jonák, Institute of Molecular Genetics, Vídeňská 1083, Prague 4, 14220, Czech Republic, tel.: +420 241063273, fax: +420 224310955, e-mail: jjon@img.cas.cz

Keywords: Src tyrosine kinases, *Xenopus laevis*, quantitative real-time PCR, whole-mount *in situ* hybridisation, early development

Running title: Developmental Expression Pattern

Email addresses:

Zoltán Ferjentsik: z.ferjentsik@dundee.ac.uk
Radek Šindelka: sindelka@img.cas.cz
Jiří Jonák: jjon@img.cas.cz

Abstract

Src family tyrosine kinases (SFKs) play important roles in cell morphology, differentiation, motility and proliferation. Elevated expression and/or specific activity of Src kinases are characteristic for several types of human cancer. However, little information is available about the role and spatio-temporal expression of SFKs in early embryonic development. In this study we determined, in *Xenopus laevis*, expression patterns of five SFK genes *src, fyn, yes, lyn* and *laloo* as well as of the *csk* gene, a negative regulator of SFKs, using RT-qPCR and *in situ* hybridisation. We found that transcripts of all SFKs and *csk* were already detectable in one-cell embryos and their levels similarly oscillated during subsequent development. First, after stage 8, the levels of SFK and *csk* mRNAs began to decrease, reached minimum between stages 10 and 28 and increased again. In the later stages (33-45) the levels of *fyn, yes* and *csk* mRNAs returned to approximately maternal ones, whereas the *src, laloo* and *lyn* mRNAs transcripts exceeded, up to about 3.5-6-fold, their maternal levels. *In situ* hybridisation analysis located the SFK and *csk* transcripts in the animal hemisphere of *Xenopus* embryos. Subsequent gastrula stages showed signals in ectodermal cells, mid-neurula stage embryos at neural folds, and the tailbud stages showed strong signals in the brain and neural tube. RT-qPCR concentration profiling along the animal-vegetal axis proved in blastula and gastrula the preferential localisation of *yes, src, lyn* and *csk* transcripts towards the animal pole in a gradient-like manner. However, *laloo* and *fyn* displayed a vegetal pole preference.

1. Results and discussion:

Src and Src-family protein-tyrosine kinases (SFKs) are proto-oncogenes and represent one of the nine presently recognised classes of non-receptor tyrosine-kinases (Pellicena and Miller, 2002). As documented in a great many reports the members of the SFK family participate in a variety of signalling pathways that control cell behavior, including differentiation and transformation (for a review see e.g. Blume-Jensen and Hunter, 2001). On the other hand, the role of SFK members in developmental processes has been examined much less extensively. The experiments carried out on mice demonstrated that Src/Fyn and Src/Yes-double knockouts die perinatally (Stein *et al*., 1994) and Src/Fyn/Yes-triple knockouts at an early stage of embryonic development (Klinghoffer *et al*., 1999).

In frogs *Xenopus laevis*, Steele and co-workers reported that *src, yes, and fyn* transcripts were already present in the maternal RNA pool (Steele, 1985; Steele *et al*. 1989, 1990). In contrast to mice, studies on frogs allowed *in vitro* direct and continuous observation and examination of developing embryos from the time of fertilization. We found that overexpression of Src kinase over a certain threshold resulted either in defective gastrulation and death, or in the development of malformed embryos characterized by a depressed level of cadherin and α-, β- and γ-catenins in their tissues (Takáč *et al*., 1998, Jonák, 2000, Dvořákova, 2000). Curiously, SFK downregulation had similar effects. Injection of antisense RNAs against Src, Fyn and Yes led to the failure of *X. laevis* blastopore and neural tube closure, to shortening of the anterior-posterior axis or other defects in embryogenesis (Denoyelle *et al*., 2001). This demonstrated that the level of Src must be strictly kept within certain boundaries; both too high and too low level of Src has deleterious effects on the development. Src does not seem to have a role in mesoderm differentiation (Denoyelle *et al*., 2001), but Fyn and Laloo could induce mesoderm formation in *Xenopus* animal caps assays (Weinstein *et al*., 2001). *In situ* hybridisation analysis of *Xenopus* embryos demonstrated nervous system-specific expression of *src* mRNA (Collett and Steele, 1992) and *fyn* mRNA (Saito *et al*., 2001), suggesting that *src* and *fyn* may play a role in elaboration of the nervous system.

RT-PCR analysis of *csk* and *laloo* expression during early *Xenopus* development showed maternally present transcripts, a greatly diminished expression by mid-blastula stages and a rise in expression again in late neurula stages (Song *et al*., 2001).

The spatial and temporal changes in gene expression are a key mechanism in embryo development. However, as summarized above, such information about SFK members is only fragmentary. Therefore, in the current study, we have utilized a quantitative real-time PCR (RT-qPCR) protocol to examine and compare expression levels of five SFK genes *src, fyn, yes, lyn and laloo*, and of *csk*, the negative regulator of SFKs, during the period of early development of *X. laevis*. In addition, we have also detailed the spatio-temporal expression patterns of *csk* and all five SFK genes using the whole-mount *in situ* hybridisation and determined intraembryonal distribution of their transcripts along the animal-vegetal axis by RT-qPCR .

1.1. Quantitative real-time PCR

The temporal expression patterns of *src, fyn, yes, lyn, laloo* and csk mRNAs determined by RT-qPCR are shown in Fig. 1. Previously, we demonstrated that RT-qPCR normalisation of mRNA expression patterns to reference genes such as ODC, GADPH, EF-1α, H4 or L8, widely used in *Xenopus* RT-qPCR experiments, is not particularly suitable, because their levels vary during *Xenopus* development. We found that normalisation to total RNA is more appropriate (Šindelka *et al*., 2006). Therefore, the mRNA expression profiles of *src, fyn, yes, lyn, laloo* and *csk* in stage series were normalised to total RNA of each embryonic stage and to stage one, and are presented in arbitrary units.

All five examined SFK mRNAs as well as the *csk* mRNA were already detectable in *Xenopus* one-cell embryos, indicating their maternal origin. This confirmed the results from Steele´s laboratory (Steele, 1985; Steele *et al*. 1989, 1990) obtained for *src*, *yes* and *fyn* transcripts as well as for *laloo*

and *csk* transcripts described in Song *et al*. (2001). Following fertilization, the levels of all examined mRNAs stayed stable up to about stage 8 (*src, fyn, yes, lyn and csk*) except for *laloo*, the level of which increased about 3 times to this stage. Then, the levels of all the transcripts began to decrease to a minimum. It was reached at stages 10.5 (*src*), 10–22 (*fyn*), 16-22 (*yes, laloo, csk*), or 10-16 (*lyn*). The levels of all transcripts then started to rise again obviously as a result of zygotic expression. The *fyn*, *yes* and *csk* mRNA levels roughly returned to maternal levels at around stage 41, 37 and 33, respectively, whereas *src*, *laloo* and *lyn* mRNAs reached, in more than one of the later stages (33-45), levels about 3.5, 5 and 6-fold higher, respectively, than was their level in the fertilized one-cell embryos. The earliest onset of zygotic expression was detected for *src* mRNA, confirming the results of Collett and Steele (1992). The *laloo* temporal expression profile presented here complements the partial *laloo* expression data previously published by Weinstein *et al*. (1998) and also correlates very well with the RT-PCR expression analysis described in Song *et al*. (2001).
In order to quantitatively compare the expression efficiency of individual SFK and *csk* genes, we normalised their expression levels to that of *lyn* mRNA (Fig. 2). While the separate expression patterns of the examined genes (Fig. 1) particularly highlight similarity in their expression profiles in the course of early development, the *lyn* mRNA normalised patterns mainly visualise quantitative differences among individual SFK in the expression of their genes. The Fig. 2 shows that the levels of *fyn* and *laloo* mRNAs are kept about 10 times lower throughout early development than are the levels of *yes, lyn* or *src*. Interestingly, the level of *csk* mRNA is maintained higher or approximately comparable with the highest levels of other examined SFK throughout all developmental stages (Fig. 2). This should not be surprising as Csk regulates activity of all SFK members. Indeed, its concentration profile follows most closely that of the most strongly expressed SFK member, the *yes* mRNA.

1.2. Whole-mount *in situ* hybridisation

To analyse the spatio-temporal expression patterns of SFKs and *csk*, we performed the whole-mount *in situ* hybridisation on *Xenopus laevis* embryos using digoxigenin-labeled antisense RNA probes. The sense RNA probes were used as negative controls and they did not show any signals. *myoD* digoxigenin-labeled antisense RNA probe was used as a positive control .
Expression of SFKs and *csk* was analysed from the one-cell stage to the tadpole stage 43 (Fig. 3), similarly as in the RT-qPCR experiments. In early cleavage stage embryos, SFKs and *csk* transcripts were detectable in the animal hemispheres but they were absent from the vegetal hemisphere (Fig. 3 A). At gastrula stages, the transcripts were detected only in the ectoderm layer (Fig. 3 B). In mid-neurula stage embryos, expression of SFKs and *csk* was detected at neural folds (Fig. 3 C).
At stages 28 and 31 (Fig. 3 D, E), strong signals were found in the brain region of the neural tube and the expression of SFKs and *csk* also appeared in the eyes, branchial arches and otic vesicles. *Fyn* (Fig. 4 B) and *yes* (Fig. 4 D) expression at tailbud stage persisted in the forebrain, midbrain, hindbrain, neural tube, eyes, otic vesicles and branchial arches. The expression of SFKs and *csk* became ubiquitous from tadpole stage 41 (not shown).
We conclude that SFKs as well as their negative regulator *csk* are in *Xenopus* expressed maternally, their levels oscillate; decline around stage 10.5 and resume again in later developmental stages. Quantitatively, the levels of *yes*, *lyn* and *src* mRNAs detected in early embryos are several times higher than those of *fyn* and *laloo* and comparable with the level of *csk* mRNA. Expression of SFKs and *csk* is transiently nervous system-specific as documented in the whole-mount *in situ* hybridisation experiments.

1.3. Detailed RT-qPCR analysis of blastula and gastrula stage

To verify, by RT-qPCR analysis, the early ectodermal pattern of some SFKs and *csk* expression described above, we determined concentration profiles of transcripts of all kinases along the animal-vegetal (A/V) axis in blastula and gastrula stage embryos. Blastula embryos (stage 8.5) were dissected into three portions (animal pole, marginal zone, and vegetal pole), gastrula embryos (stage 11) into four portions (animal pole, dorsal and ventral marginal zone and vegetal pole; see scheme in Fig. 5) and the portions were subjected to RT-qPCR analysis. The amount of transcripts in each section was expressed in per cent of the overall content of individual transcripts present in the whole embryo.

This RT-qPCR "intraembryonal" analysis detected majority of *yes*, *src*, *lyn* and *csk* transcripts in the animal pole portion, less in the marginal zone portion and still less in the vegetal pole portion of the embryos. *Yes*, *src*, *lyn* and *csk* transcripts formed concentration gradients along the embryonic *X. laevis* animal-vegetal axis (Fig. 5). Approximately 50% of transcripts of these kinases were present in the animal pole portion, approximately 30-40% in the marginal zone portion and less than 20% of transcripts were present in vegetal pole portion of blastula stage embryos (Fig. 5 A). Very similar concentration ratios among the animal, marginal and vegetal transcripts of *yes*, *src*, *lyn* and *csk* were also detected in gastrula stage embryos (Fig. 5 B). No significant differences in expression levels of these kinases in dorsal and ventral marginal zones were detected. Altogether, these findings are in good agreement with the data obtained by *in situ* hybridisation (Fig. 3 A, B) and y*es*, *src*, *lyn* and *csk* could be considered as "animal genes".

On the other hand, expression patterns of *fyn* and *laloo* were found to be different. The amount of *fyn* transcripts was distributed almost identically among all three A/V sections of blastula stage embryos (Fig 5 A), and at gastrula stage it increased towards the vegetal pole (Fig. 5 B). The *laloo* mRNA expression pattern turned out to be quite opposite to that of *yes*, *src*, *lyn* and *csk* mRNAs. More than 80% of *laloo* transcripts at the blastula stage and almost 60% at the gastrula stage were found to be located in the vegetal pole portion and could be considered as a "vegetal gene". Dorsal and ventral marginal zones of gastrula did not differ in expression of either *fyn* or *laloo* (5 B).

The expression profile of *laloo* transcripts determined by RT-qPCR was not consistent with the *in situ* hybridisation results (Fig. 3 A, B). Possible reason could be a weaker penetration of RNA probes to cells of vegetal pole of *Xenopus* embryos as well as a lower concentration of both *fyn* and *laloo* transcripts in embryos in comparison to all other kinases examined here (Fig. 2). We believe that RT-qPCR can measure intraembryonal mRNA gradients with greater resolution and sensitivity than traditional *in situ* hybridisation. (Compare also with RT-qPCR "tomography" carried out on *X. leavis* oocytes; Šindelka, R., Jonák, J., Hands, R., Bustin, S., Kubista, M., manuscript in press).

2. Experimental procedures:

2.1. Embryos and explants

Xenopus laevis embryos were obtained by *in vitro* fertilization and staged according to the Nieuwkoop and Faber tables (Nieuwkoop and Faber, 1967). Embryos were dejellied in 2% cysteine (Sigma) (pH 8) and then cultured in 0.1xNAM (Slack and Forman, 1980).
To analyse the distribution of mRNA transcripts along the animal-vegetal axis in blastula (8.5) and gastrula (11) stage *Xenopus* embryos they were manually dissected into three (animal, marginal, vegetal) and four, respectively, portions (see Fig. 5 for details) and subjected to real-time RT–qPCR expression analysis of SFKs and *csk*. The mRNA expression levels were normalised to total RNA and *gapdh* expression level.

2.2. RNA probes

Digoxygenin sense and antisense transcripts were synthesized using SP6/T7 RNA labelling kit (Roche) following the manufacturer's instructions.
Xenopus src, fyn, yes, lyn, laloo, csk coding sequences were cloned into the pCS2- vector (kind gift of D. Turner) after PCR amplification using following primer pairs: *src*: sense, 5'-ccgctcgagatgggtgccactaaaagtaag-3'; antisense, 5'-gctctagattaaaggttgtccccaggc-3'; *fyn*: sense, 5'-ggaattccatgggctgtgtgcaatgcaag-3'; antisense, 5'-ccgctcgagttacaggttgtctccaggctg-3'; *yes*: sense, 5'-cgggatcctcatgggctgtataaaaagtaagg-3'; antisense, 5'-ggaattccttatagattgtccccaggctggt-3'; *lyn*: sense, 5'-cgggatcctcatgggatgtataaaatcaaaaac-3'; antisense, 5'-ggaattccctaaggctgttgttgatattg-3'; *laloo*: sense, 5'-ggaattccatgggctgcatcaagtcaaag-3'; antisense, 5'-ccgctcgagttaaggttgtgcctggtactg-3'; *csk*: sense, 5'-ccgctcgagccatgtcggtggtacaggccc-3'; antisense, 5'-gctctagatcagtgatacagttccttggc-3'.
Primers were designed using the published sequences for *src* BC110764, *fyn* X52188, *yes* X14377, *lyn* AB003358, *laloo* AF081803 and *csk* AF052430.
For antisense RNA, the pCS2- vectors were linearized with BamHI (*fyn*, *yes*, *lyn*, *laloo*), HindIII (*src*) or ApaI (*csk*). For sense RNA (negative control experiments), the pCS2- vectors were linearized with NotI. Both antisense- and sense-digoxigenin-labeled RNA probes were obtained using T7 or SP6 RNA polymerase (Roche). RNA probes were purified with ProbeQuant G-50 Micro Columns (Amersham) and checked by agarose gel electrophoresis. Synthesized control sense probes gave no staining after whole-mount *in situ* hybridisation (results not shown).

2.3. Whole-mount *in situ* hybridisation and sectioning

Whole-mount *in situ* hybridisation was performed according to the standard protocol (Harland, 1991). The antisense *src*, *fyn*, *yes*, *lyn*, *laloo*, and *csk* probes were designed to hybridise specifically with their unique N-terminal region.

2.4. RNA isolation and cDNA preparation

RNA was isolated from *Xenopus* tissue using the Trizol (Invitrogen) or TriReagent (Sigma) method of extraction. A cDNA pool was generated from total cellular RNA by using random oligonucleotides and MMLV reverse transcriptase as previously described (Šindelka *et al*., 2006).

2.5. Quantitative real-time RT-qPCR

The reaction was accomplished in a total volume of 25 μl. The reaction mixture contained 2 μl of the cDNA template, 0.2 mM dNTPs, 240 nM forward and reverse primers, 1 U Taq polymerase (Promega), 2.5 μl of supplied 10x buffer, 2 mM $MgCl_2$, 2.5 μl of 10,000-fold diluted SYBRGreen solution (Molecular Probes) and 0.25 μl of 50 nM Fluorescein solution (Bio-Rad). The reactions

were measured in iCycler (Bio-Rad) with cycling conditions: 95°C for 5 min, 40 cycles at 95°C for 15 sec and 60°C for 60 sec. Serially diluted PCR fragments (standards), identical with those amplified in the real-time PCR experiment, were prepared to obtain calibration curves. Reaction efficiencies determined from calibration curves for each set of primers were between 85 and 100%. The Cts (threshold cycles) of the samples and standards were analysed with Microsoft Excel program and the number of amplified cDNA copies as PCR products from particular stages of development were determined from calibration curves. The average deviation between Cts in parallel experiments did not exceed about 5% for all tested genes and stages. The expression profiles were derived from three independent *X. laevis* serial experiments. Specificity of every amplification reaction was verified by melting curve analysis and gel electrophoresis.

Primers used for real-time PCR were designed by using the Beacon Designer 2.00 program (Premier Biosoft International). Primers used for *src* were: sense, 5'-gcgactgattgaggacaatgagta-3'; antisense, 5'-aggagaattccaaaagaccagaca-3'; for *fyn*: sense, 5'-gccaggcaccatgtctccag-3'; antisense, 5'-ctcctcagacaccacagcgtag-3'; for *yes*: sense, 5'-caccaacaccagtcccttaccc-3'; antisense, 5'-atcttgcttcccaccagtcacc-3'; for *lyn*: sense, 5'-atccagcttctcgtacaccaag-3'; antisense, 5'-tccaccattctccatgctcttc-3'; for *laloo*: sense, 5'-tctaagcaccccagagagga-3'; antisense, 5'-ccgctcgagctttagcaggagatggtccc-3'; for *csk*: sense, 5'-ggcaagctgagcattgacgaag-3'; antisense, 5'-gcggctactgttccctccatc-3'; for *gapdh*: sense, 5'-gccgtgtatgtggtggaatct-3'; antisense, 5'-aagttgtcgttgatgacctttgc-3'.

3. Acknowledgements:

We thank Dr Gufa Lin for his kind assistance during *in situ* hybridisation experiments.
This work was supported by grant 301/02/0408 from the Grant Agency of the Czech Republic (to J.J.) and by project AVOZ 50520514 awarded by the Academy of Sciences of the Czech Republic.
The *in situ* hybridisation experiments were performed by Z.F. in the laboratory of Professor Jonathan Slack at the University of Bath, supported by BSDB and EACR short-time fellowships. We would like to thank all members of his laboratory.

References

BLUME-JENSEN, P., HUNTER, T. (2001). Oncogenic kinase signaling. *Nature* 411(6835): 355-65.

COLLETT, J.W., STEELE, R.E. (1992). Identification and developmental expression of Src+ mRNAs in *Xenopus laevis*. *Dev. Biol.* 152(1): 194-8.

DENOYELLE, M., VALLES, A.M., LENTZ, D., THIERY, J.P., BOYER, B. (2001). Mesoderm-independent regulation of gastrulation movements by the src tyrosine kinase in *Xenopus* embryo. *Differentiation* 69(1): 38-48.

DVOŘÁKOVÁ, K., HABROVÁ, V., TAKÁČ, M., JONÁK, J. (2000). Depression in the level of cadherin and alpha-, beta-, gamma-catenins in transgenic *Xenopus laevis* highly expressing c-Src. *Folia Biol.* (Praha) 46(1): 3-9.

HARLAND, R.M. (1991). *In situ* hybridization: an improved whole-mount method for *Xenopus* embryos. *Methods Cell Biol.* 36: 685-695.

JONÁK, J. (2000). Sperm-mediated preparation of transgenic *Xenopus laevis* and transmission of transgenic DNA to the next generation. *Mol. Reprod. Dev.* 56(2 Suppl): 298-300.

KLINGHOFFER, R.A., SACHSENMAIER, C., COOPER, J.A., SORIANO, P. (1999). Src family kinases are required for integrin but not PDGFR signal transduction. *EMBO J.* 18(9): 2459-71.

NIEUWKOOP, P.D. and FABER, J. (1967). Normal Table of *Xenopus laevis* (Daudin). Amsterdam: North-Holland.

PELLICENA, P., MILLER, W.T. (2002). Coupling kinase activation to substrate recognition in SRC-family tyrosine kinases. *Front. Biosci.* 7: d256-67.

SAITO, R., FUJITA, N., NAGATA, S. (2001). Overexpression of Fyn tyrosine kinase causes abnormal development of primary sensory neurons in *Xenopus laevis* embryos. *Dev. Growth Differ.* 43(3): 229-38.

ŠINDELKA, R., FERJENTSIK, Z., JONÁK, J. (2006). Developmental expression profiles of *Xenopus laevis* reference genes. *Dev. Dyn.* 235(3): 754-8.

SLACK, J.M., FORMAN, D. (1980). An interaction between dorsal and ventral regions of the marginal zone in early amphibian embryos. *J. Embryol. Exp. Morphol.* 56: 283-99.

SONG, Y., COHLER, A.N., WEINSTEIN, D.C. (2001). Regulation of Laloo by the *Xenopus* C-terminal Src kinase (Xcsk) during early vertebrate development. *Oncogene* 20(37): 5210-4.

STEELE, R.E. (1985). Two divergent cellular src genes are expressed in *Xenopus laevis*. *Nucleic Acids Res.* 13(5): 1747-61.

STEELE, R.E., DENG, J.C., GHOSN, C.R., FERO, J.B. (1990). Structure and expression of fyn genes in *Xenopus laevis*. *Oncogene* 5(3): 369-76.

STEELE, R.E., IRWIN, M.Y., KNUDSEN, C.L., COLLETT, J.W., FERO, J.B. (1989). The yes proto-oncogene is present in amphibians and contributes to the maternal RNA pool in the oocyte. *Oncogene Res.* 4(3): 223-33.

STEIN, P.L., VOGEL, H., SORIANO, P. (1994). Combined deficiencies of Src, Fyn, and Yes tyrosine kinases in mutant mice. *Genes Dev.* 8(17): 1999-2007.

TAKÁČ, M., HABROVÁ, V., MÁCHA, J., ČEŠKOVÁ, N., JONÁK, J. (1998). Development of transgenic *Xenopus laevis* with a high C-src gene expression. *Mol. Reprod. Dev.* 50(4): 410-9.

WEINSTEIN, D.C., HEMMATI-BRIVANLOU, A.A. (2001). Src family kinase function during early *Xenopus* development. *Dev. Dyn.* 220(2): 163-8.

WEINSTEIN, D.C., MARDEN, J., CARNEVALI, F., HEMMATI-BRIVANLOU, A. (1998). FGF-mediated mesoderm induction involves the Src-family kinase Laloo. *Nature* 394(6696): 904-8. Erratum in: Nature 1998 395(6705): 921.

Figure Legends

Fig. 1. The mRNA expression profiles of *Xenopus laevis src, fyn, yes, lyn, laloo* and *csk* genes normalised to total RNA and stage one and expressed in arbitrary units. The numbers on the vertical axis represent the ratio between the average amount of copies of a mRNA at a particular developmental stage and stage one normalised to the same amount of input RNA (means ± SD, n=6 replicates). The numbers on the horizontal axis represent the *Xenopus* developmental stages determined according to Nieuwkoop and Faber (1967).

Fig. 2. The mRNA expression profiles of SFKs and *csk*. mRNA expression profiles are normalised to total RNA and stage one of *lyn* mRNA and expressed in arbitrary units. The numbers on the vertical axis represent the ratio between the average amount of copies of mRNA at a particular developmental stage and stage one normalised to *lyn* mRNA copy amount. The numbers on the horizontal axis represent the *Xenopus* developmental stages determined according to Nieuwkoop and Faber (1967).

Fig. 3. The spatial and temporal expression patterns of *src, fyn, yes, lyn, laloo* and *csk* mRNAs during *Xenopus laevis* development. (A) animal (upper row) and lateral (lower row) view of embryonic stages 1, 2, 3 and 8 (blastula) showing enhanced signals in the animal hemisphere. **(B)** animal (upper row) and vegetal (lower row) view of gastrula stages (10, 10.5, 11) showing no signal in endodermal cells. **(C)** anterior and dorsal view of mid-neurula (stage 16) embryo. SFKs and *csk* are expressed at the neural folds from anterior to caudal end, but absent from the dorsal midline. **(D, E)** lateral view of tailbud (stage 28, 31) embryos. Expression of SFKs and *csk* persists in the developing brain, neural tube, eye, branchial arches and otic vesicle. Bottom panels depict the expression pattern of *myoD* used as a positive control. First two panels show lateral and dorsal view of stage 23 embryos and the other two panels show lateral view of stage 31 embryos with the expression pattern of *myoD*.

Fig. 4. Detailed expression patterns of *Xenopus fyn* and *yes* genes. The *fyn* gene expression patterns **(A, B)**. Anterior view **(A)** of *Xenopus* embryo at neurula stage. The arrow shows the expression of the *fyn* gene in neural folds. Lateral view **(B)** of a tailbud (stage 31) embryo. *Fyn* expression persists in the developing brain, neural tube, eye, branchial arches and otic vesicle.
The *yes* gene expression pattern **(C, D)**. Anterior view **(C)** of *Xenopus* embryo at neurula stage. The arrow shows the expression of *yes* gene in neural folds. Lateral view **(D)** of embryo at stage 31. The arrows show the major expression of *yes* gene in the developing brain, neural tube, eye, branchial arches and otic vesicle.
hb; hindbrain, mb; midbrain, fb; forebrain, ba; branchial arches, ey; eye, ov; otic vesicle.

Fig. 5. "Intraembryonal" expression analysis of *Xenopus* SFKs and *csk* mRNAs at blastula and gastrula stages by RT-qPCR. Blastula (8.5) stage **(A)** and gastrula (11) stage **(B)** embryos were dissected into three (animal, marginal, vegetal) and four (animal, dorsal-marginal, ventral-marginal, vegetal), respectively, sections along the animal-vegetal axis as indicated and analysed as described in Experimental procedures.

Fig 1:

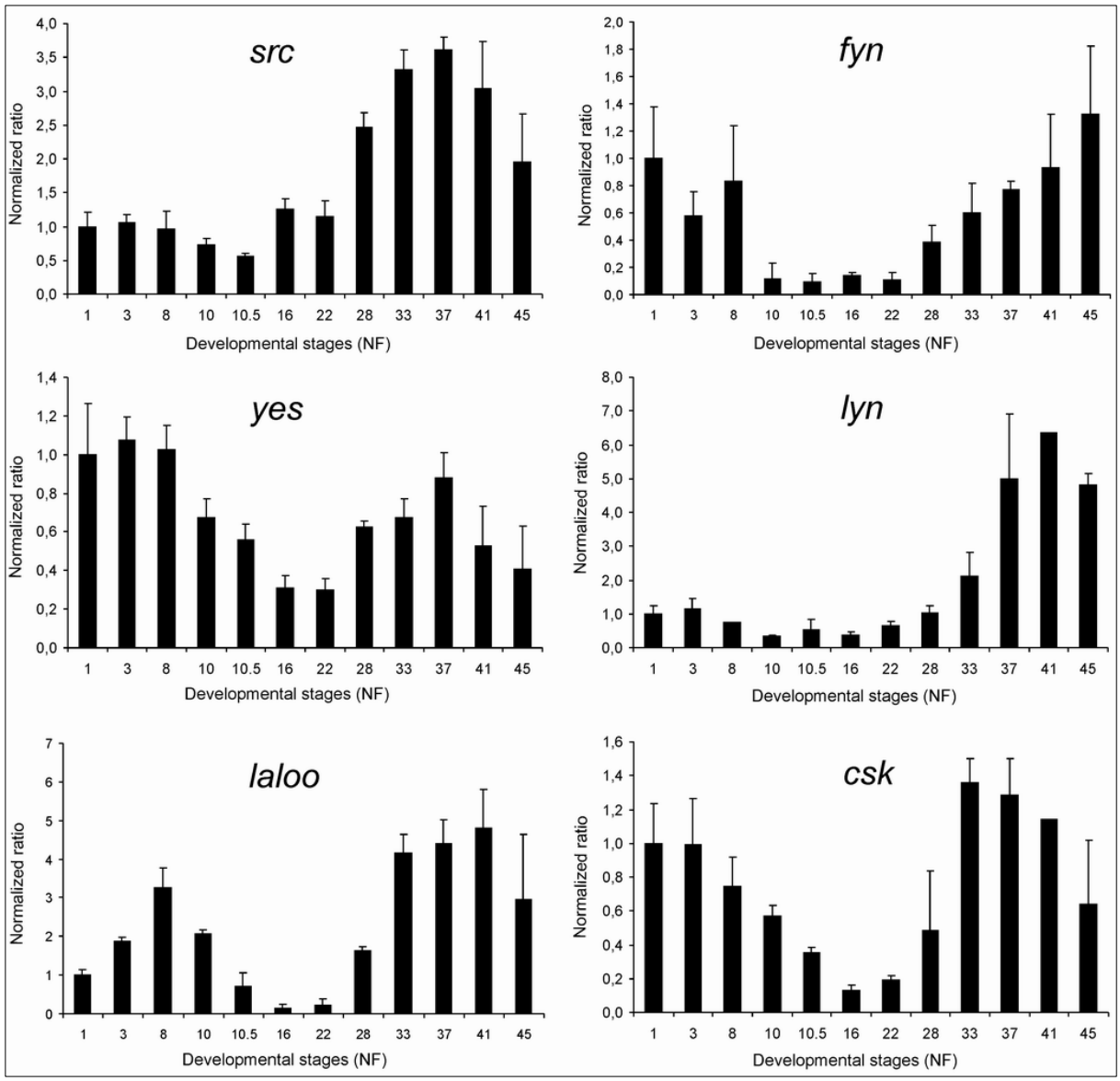

src
fyn
yes
lyn
laloo
csk
Normalized ratio
Developmental stages (NF)
1
3
8
10
10.5
16
22
28
33
37
41
45

Fig 2:

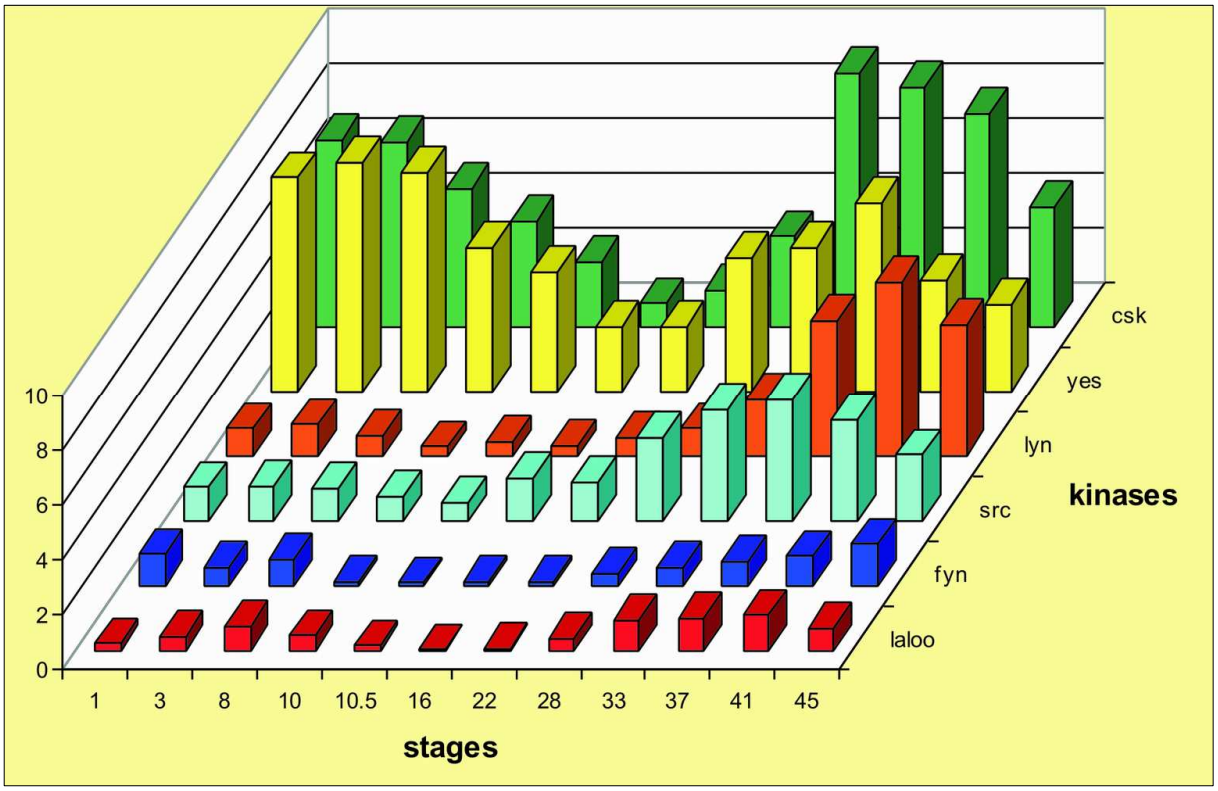

10
8
6
4
2
0
1
3
8
10
10.5
16
22
28
33
37
41
45
stages
csk
yes
lyn
src
fyn
laloo
kinases

Fig 3:

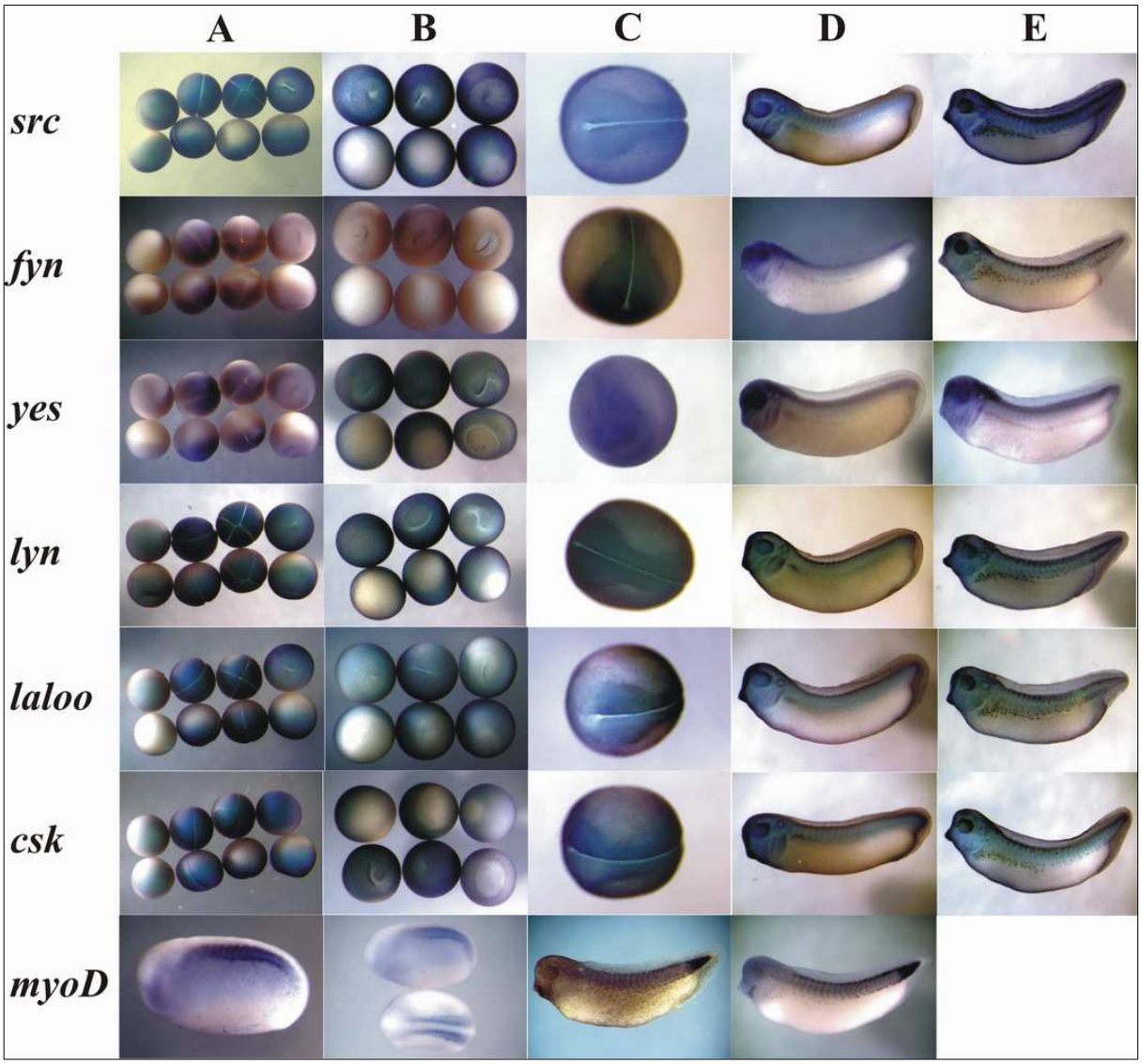


Fig 4:

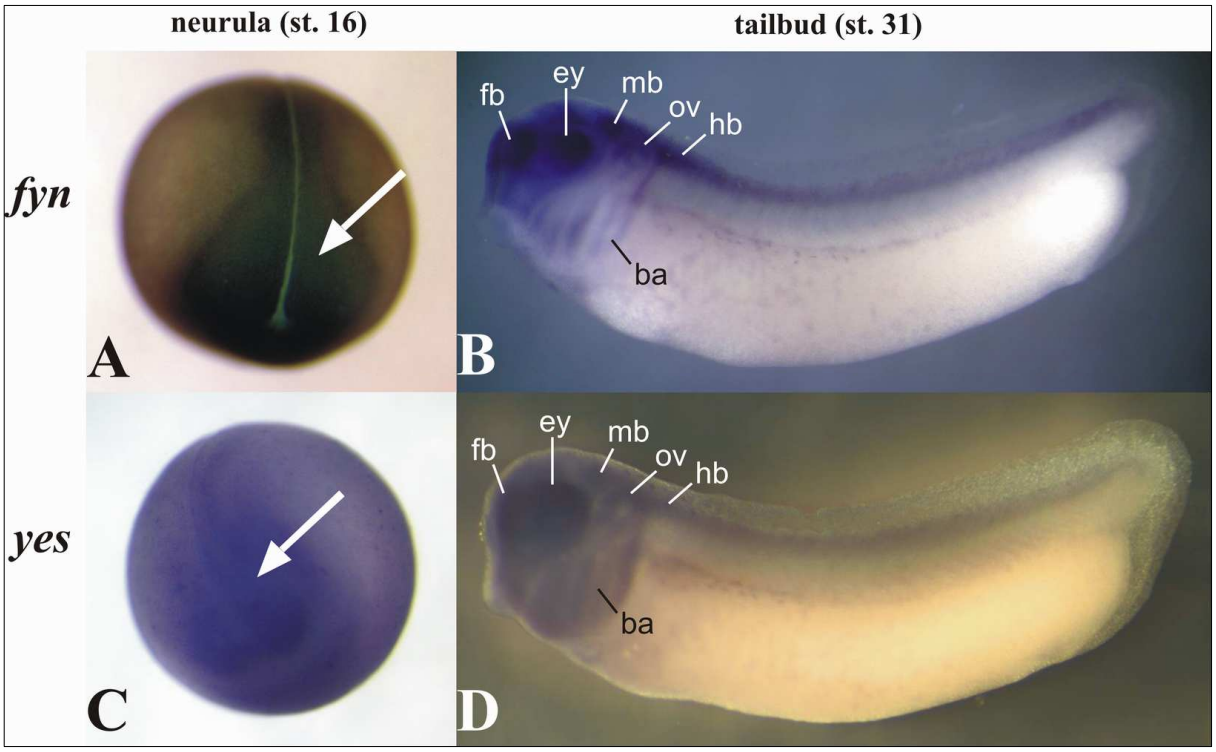
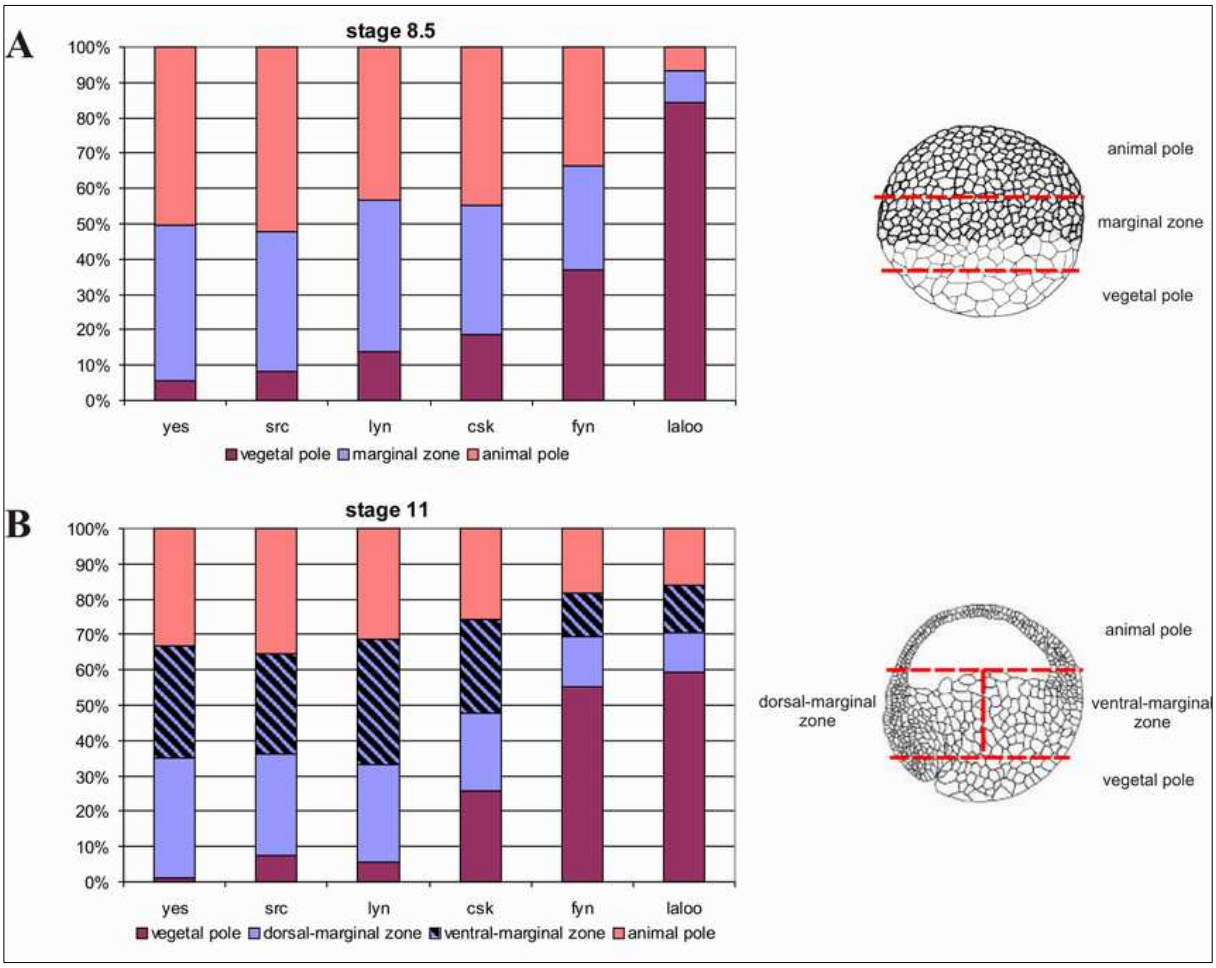


Fig 5:


A
stage 8.5
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
yes
src
lyn
csk
fyn
laloo
vegetal pole
marginal zone
animal pole
animal pole
marginal zone
vegetal pole
B
stage 11
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
yes
src
lyn
csk
fyn
laloo
vegetal pole
dorsal-marginal zone
ventral-marginal zone
animal pole
dorsal-marginal zone
animal pole
ventral-marginal zone
vegetal pole