



## UvA-DARE (Digital Academic Repository)

### Transcriptome dynamics in early zebrafish embryogenesis

Rauwerda, J.

**Publication date**

2017

**Document Version**

Other version

**License**

Other

[Link to publication](#)

**Citation for published version (APA):**

Rauwerda, J. (2017). *Transcriptome dynamics in early zebrafish embryogenesis*.

**General rights**

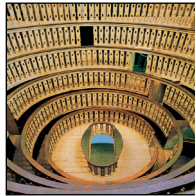
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

---

# Summary





## *INTRODUCTION*

The aim of this thesis is defined as to obtain a deeper insight in transcriptome dynamics, by investigating at an omics scale the changing gene-expression during biological processes. The approach taken is to carry out a number of non-perturbing gene-expression studies in which zebrafish eggs and embryos are used as model system. Early zebrafish embryogenesis is introduced. Specifically, the material that is maternally collected in the egg, maternal mRNA characteristics, maternal mRNA silencing, the organization of maternal mRNA, the silencing of the oocyte transcription, clearance of maternal mRNA and zygotic transcription initiation are briefly discussed. A technical perspective on omics technologies required to study transcriptome dynamics is given. Furthermore, some bioinformatics aspects of this study are introduced. Finally, an outline of the thesis is given.

## *CHAPTER 1 RNA ISOLATION METHOD FOR SINGLE EMBRYO TRANSCRIPTOME ANALYSIS IN ZEBRAFISH*

A zebrafish egg is small: 700  $\mu\text{m}$  in diameter excluding the egg shell (chorion). The zebrafish embryo, including the yolk, does not substantially increase in size until the somite stages (from 11 hours post fertilization (hpf) onward). In this chapter, the challenge was met how to obtain enough RNA from a single egg, which is a single cell, or from an individual embryo that can be used as input for transcriptomics experimentation. In recent years, several methods have been developed for single-cell transcription profiling. These all heavily rely on amplification steps, either by a PCR reaction or via in vitro transcription (IVT). Both approaches suffer from biases. Here a protocol is presented by which we optimize the yield of the RNA isolation in order to allow for a lower number of amplification cycles and thus limiting the technical bias in transcriptomics studies using low amounts of input material. The technical variability of this method was tested and its robustness was demonstrated. The RNA isolation method developed here was subsequently used in all experiments that were performed in this thesis.

## *CHAPTER 2 INTEGRATING HETEROGENEOUS SEQUENCE INFORMATION FOR TRANSCRIPTOME-WIDE MICROARRAY DESIGN; A ZEBRAFISH EXAMPLE*

There are a number of resources that identify or predict, and characterize genomic elements, such as Ensembl, Vega, Refseq and Unigene. These resources still are incomplete and their information is continuously being updated. In each resource, a procedure is followed that implements a different set of rules. The results are gene and/or transcript catalogues that only partly overlap and to which individual elements different levels of confidence must be attached. However, in our studies into transcriptome dynamics, we preferably should be able to detect, e.g. by microarrays all genomic sequences that can

be expressed as RNA in an organism. In this chapter, we present a strategy to integrate heterogeneous sequence information that can be used as input for an up-to-date microarray design. We show it is possible to use a set of heterogeneous transcript resources for microarray design, reduce the number of candidate target sequences on which the design is based and reduce redundancy. By changing the parameters in the procedure it is possible to control the similarity and the amount of candidate sequences for the design. This procedure was used to design the microarrays that were used in Chapters 3 and 4 of this thesis.

### *CHAPTER 3A MOTHER-SPECIFIC SIGNATURE IN THE MATERNAL TRANSCRIPTOME COMPOSITION OF MATURE, UNFERTILIZED ZEBRAFISH EGGS*

Maternal mRNA present in mature oocytes plays an important role in the proper development of the early embryo. As the composition of the maternal transcriptome almost without exception has been studied with pooled mature eggs, potential differences between individual eggs are unknown. In this chapter we presented a transcriptome study on individual zebrafish eggs from clutches of five mothers in which we focus on the differences in maternal mRNA abundance per gene between and within clutches. To minimize technical interference, we used mature, unfertilized eggs from siblings. About half of the number of analyzed genes was found to be expressed as maternal RNA. The expressed and non-expressed genes showed that maternal mRNA accumulation is a non-random process, as it is related to specific biological pathways and processes relevant in early embryogenesis. Moreover, it turned out that overall the composition of the maternal transcriptome is tightly regulated. Expression differences observed in maternally expressed genes can for the far majority of genes be traced back to differences between mothers, whereas only a limited variability is observed between eggs from the same mother. Linking chromosome location, transcription factor binding sites, and 3' UTR miRNA target sites to clusters of genes that were grouped by a distinct and unique mother-specific gene-expression, suggests biological relevance of the mother-specific signatures in the maternal transcriptome composition. Altogether, the maternal transcriptome composition of mature zebrafish oocytes seems to be tightly regulated with a distinct mother-specific signature.

### *CHAPTER 3B TRANSCRIPTOME DATA ON MATERNAL RNA OF 24 INDIVIDUAL ZEBRAFISH EGGS FROM FIVE SIBLING MOTHERS*

The data that is produced and the scripts that were developed for the study in Chapter 3A are made available in an annotated form. A Bayesian procedure is presented that exploits the observation that the log variance of log intensities is bimodally distributed. This allows for the calculation of the conditional probability of a gene to be in the high variance distribution, i.e. 'expressed', given a certain intensity. This procedure is used in the analysis of the data from Chapters 3 and 4.

---

#### CHAPTER 4 *TRANSCRIPTOME DYNAMICS OF EARLY ZEBRAFISH EMBRYOGENESIS DETERMINED BY HIGH-RESOLUTION TIME SERIES ANALYSIS*

Recently, many discoveries have been made in the field of gene-expression in early embryogenesis, such as by studies mentioned in the Introduction in the paragraph on the Biological Perspective. However, the behavior of transcriptomes in individual embryos has hardly been studied yet. In this chapter we present a high-resolution gene-expression time series with 180 individual zebrafish embryos, obtained from nine different spawns, developmentally ordered and profiled from late blastula to mid-gastrula stage. On average one embryo per minute was analyzed. The focus is on identification and description of the transcriptome dynamics of the 6,734 expressed genes in early embryogenesis, rather than to biologically interpret profiles in cellular processes and pathways. The existence of a tightly regulated embryonic transcriptome program, even when individuals from different spawns are compared is shown. In general, gene expression shows low variability and during this developmental period changes rather gradually. Ten types of dynamic behavior were identified and the exact expression starting and stopping points of several genes during this developmental period could be pinpointed. Starting and stopping genes appeared to have the steepest fold changes observed in this study. Apart from four truly oscillating genes, no genes were observed with a peaking expression. Additionally, several genes showed expression at two or three distinct levels that strongly related to the spawn an embryo originated from. This spawn-specific transcriptome effect is likely the accumulated outcome of the many epigenetic factors in gene-expression regulation.

#### CHAPTER 5 *CELLULAR FACTORS INFLUENCING TRANSCRIPTOME DYNAMICS IN EARLY ZEBRAFISH EMBRYOGENESIS*

In this chapter, the gene-expression types that were defined in Chapter 4 are further analyzed, together with the set of genes that show distinct spawn specificity. These categories were analyzed by investigating the genome location of genes in these types, their gene size, their gene promoter regions for specific transcription factor binding sites, their 3' UTRs for cytoplasmic polyadenylation motifs, and their 3' UTRs for miRNA targets. We also checked for overrepresentation of specific pathways and gene ontology categories in these gene expression types. Furthermore, we investigated the actual miRNA expression of individual embryos at approximately 50% epiboly in a sequencing experiment using the same spawns as in the experiment of Chapter 4. The miRNA expression of individual embryos spanning the entire zebrafish embryogenesis including unfertilized eggs and adult zebrafish was established in another sequencing experiment. We observe that genes in gene-expression types are active in specific pathways. However, rather than to characterize early gastrulation as to be regulated by a handful of mechanisms, we find an overwhelming amount of processes involved in regulation of this developmental stage. Not only the resulting gene expression is tightly regulated, also at least some of the regulators themselves are tightly controlled. This is illustrated by the behaviour of miR-430. We conclude that even the gene-expression types that unfold in this short

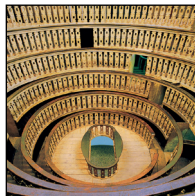
developmental period are the integrated results of many regulating processes and, to further unravel these processes, we need to zoom in on an even higher resolution, specifically in space and with respect to the molecular signature of the transcripts.

#### *CONCLUDING REMARKS*

Finally, our findings are put into perspective; how did our efforts to integrate genomic information work out; what is the fate of maternal RNA; what is the significance of the tight regulation of gene expression found in Chapters 3 to 5; how can we interpret the relative gradual changes in gene expression we found? We discuss our results, specifically with respect to cytoplasmic polyadenylation, conceptualization, starting and stopping points of gene expression, tight regulation of gene expression over time and when individuals are compared, and the individual signature on gene expression. We discuss our results in their relation to the studies of others and give our opinion on possible approaches to deepen our insight into transcriptome dynamics.

---

# Samenvatting







## INLEIDING

Het doel van de experimenten die in het kader van dit proefschrift uitgevoerd zijn wordt allereerst toegelicht: de studies richten zich op het verkrijgen van een dieper inzicht in transcriptoom dynamica door op een omics-schaal veranderingen in genexpressie tijdens biologische processen te bestuderen. Met het transcriptoom wordt het ensemble van alle aanwezige RNA-moleculen bedoeld. Deze steady state is het resultaat van transcriptie en van afbraakprocessen. Een studie op omics-schaal richt zich in principe op alle genen, c.q. aanwezige RNA-moleculen. In de hier gekozen aanpak worden een aantal genexpressie studies uitgevoerd met zebraavis eitjes en embryo's als modelsysteem. Hierbij hebben we niet ingegrepen in de ontwikkeling van eitjes of embryo's.

Na het beschrijven van de aanpak wordt de vroege embryogenese in zebraavis besproken en in context geplaatst. Hierbij wordt het materiaal dat de moeder doorgeeft in het eitje, de karakteristieken van het maternale mRNA (boodschapper RNA), het stilleggen van translatie van maternaal mRNA, het stilleggen van de oocyt transcriptie, de verwijdering van maternaal mRNA tijdens de embryogenese en de initiatie van de transcriptie van de zygote kort besproken. De technische voorwaarden welke vervuld moeten worden om een omics-studie op dit materiaal uit te voeren worden bediscussieerd, samen met de bioinformatica aspecten die hierbij van belang zijn. Tenslotte wordt een samenvatting van het proefschrift gegeven.

## HOOFDSTUK I EEN RNA-ISOLATIEMETHODE VOOR DE TRANSCRIPTOOM ANALYSE VAN EEN INDIVIDUEEL ZEBRAVIS EMBRYO

Een zebraavis eitje is klein: zonder de eischaal (chorion) heeft het een diameter van 700  $\mu\text{m}$ . Een zebraavis embryo, inclusief het dooiermateriaal blijft tot het somite stadium (ongeveer 11 uur na bevruchting, hpf) ongeveer even groot. In dit hoofdstuk wordt een nieuwe methode beschreven die het mogelijk maakt voldoende RNA uit één eitje - een eitje is één enkele cel - of uit één individueel embryo te verkrijgen om een transcriptoom experiment uit te kunnen voeren. In de afgelopen jaren zijn er verschillende methodes ontwikkeld voor single-cell transcriptoom studies. Deze maken echter alle gebruik van amplificatie stappen, hetzij door middel van PCR, hetzij met in vitro transcriptie (IVT) methodes. Het nadeel van deze technieken is dat ze bias introduceren. In hoofdstuk I wordt een protocol gepresenteerd dat de opbrengst van de RNA-isolatie optimaliseert en het daarmee mogelijk maakt de transcriptoom experimenten met een lager aantal amplificatiestappen uit te voeren. Hierdoor kan, ondanks de lage hoeveelheid uitgangsmateriaal de technische bias beperkt worden. In dit hoofdstuk is de technische variabiliteit van de ontwikkelde methode getest en de robuustheid ervan aangetoond. Deze RNA-isolatiemethode is in alle experimenten die in dit proefschrift zijn uitgevoerd gebruikt.

**HOOFDSTUK 2** *INTEGRATIE VAN HETEROGENE (DNA EN RNA) SEQUENTIE INFORMATIE VOOR EEN TRANSCRIPTOOM-BREED MICROARRAY ONTWERP; EEN ZEBRAVIS VOORBEELD*

Er bestaan een aantal bronnen die genomische elementen indentificeren, voorspellen en/of karakteriseren. Voorbeelden daarvan zijn Ensembl, Vega, Refseq en Unigene. Deze bronnen leveren nog steeds geen uitputtende en definitieve beschrijving van deze elementen en worden voortdurend aangepast. Elke bron volgt een eigen procedure en past een set bron-specifieke regels toe. Het resultaat is dat de verschillende gen- en transcriptcatalogi slechts gedeeltelijk overlappen. Bovendien zijn aan genen en transcripten een verschillend niveau van betrouwbaarheid toegekend. We verwachten echter dat de detector in onze studie naar transcriptoom dynamica, d.w.z. bijvoorbeeld een microarray, in staat is om alle genomische sequenties die als RNA in een organisme tot expressie kunnen komen kan waarnemen. In dit hoofdstuk wordt een strategie gepresenteerd om heterogene sequentie informatie op een zodanig wijze te integreren dat deze gebruikt kan worden als input voor een up-to-date microarray ontwerp. We tonen aan dat het mogelijk is een verzameling heterogene transcript bronnen te gebruiken om een microarray te ontwerpen, om het aantal kandidaat doel sequenties op welk het ontwerp gebaseerd is beperkt te houden en om redundantie te verminderen. Door de parameters in de procedure aan te passen is het bovendien mogelijk het oplossend vermogen van het array en het aantal kandidaat sequenties voor het ontwerp te reguleren. Het in hoofdstuk 2 beschreven algoritme is gebruikt bij het ontwerp van de microarrays die zijn gebruikt in de experimenten in de hoofdstukken 3 en 4 van deze studie.

**HOOFDSTUK 3A** *EEN MOEDER SPECIFIEKE SIGNAATUUR IN DE SAMENSTELLING VAN HET MATERNALE TRANSCRIPTOOM VAN RIJPE ONBEVRUCHTE ZEBRAVIS EITJES*

Het in rijpe oocyten aanwezige maternale mRNA speelt een belangrijke rol in de ontwikkeling van het vroege embryo. De samenstelling van het maternale transcriptoom is tot nu toe bijna zonder uitzondering bestudeerd door in een experiment een hoeveelheid eitjes uit één of meer legsels samen te nemen (te poolen). In zo'n experimentele opzet kunnen mogelijke verschillen in het transcriptoom van individuele eitjes niet worden waargenomen. In dit hoofdstuk wordt een transcriptoom studie op individuele eitjes van legsels van vijf moeders gepresenteerd, waarin de verschillen in de aanwezigheid van maternale mRNA's per gen tussen en binnen legsels bestudeerd wordt. Om de invloed van mogelijke genomische verschillen te minimaliseren werden daarbij rijpe onbevuchte eitjes van zusters gebruikt. Ongeveer de helft van het aantal gemeten genen bleek tot expressie te zijn gebracht. De verdeling van tot expressie gebrachte en niet tot expressie gebrachte genen laat zien dat de accumulatie van maternale RNA's gerelateerd is aan specifieke biologische pathways en aan processen die met vroege embryogenese verband houden. Dit toont aan dat het een niet-random proces is. Bovendien bleek dat over het geheel genomen de samenstelling van het maternale transcriptoom zeer strak gereguleerd is. Waargenomen verschillen in expressie van deze maternale genen kunnen in de overgrote meerderheid van de gevallen toegewezen worden aan moeders;

in eitjes van dezelfde moeder wordt slechts een beperkte variabiliteit in maternale genexpressie waargenomen. We hebben vervolgens door middel van hiërarchische clustering genen verzameld met een unieke en verschillende moeder-specifieke genexpressie. De vergelijking van deze gen sets met chromosoom lokatie, transcriptie factor bindingsplaatsen en 3' UTR miRNA doel sequenties suggereert dat de gevonden moeder specifieke signatuur in de de maternale transcriptoom samenstelling biologisch relevant is. Samenvattend wijst dit experiment op een strakke regulatie van de maternale transcriptoom samenstelling in rijpe zebraavis oocyten samen met een unieke en verschillende moeder specifieke signatuur.

### *HOOFDSTUK 3B TRANSCRIPTOOM DATA VAN MATERNAAL RNA VAN 24 INDIVIDUELE ZEBRAVIS EITJES VAN VIJF MOEDERS (ZUSTERS)*

De data en de scripts die geproduceerd zijn voor het experiment dat in hoofdstuk 3A beschreven is worden hier in een geannoteerde vorm beschikbaar gemaakt. Hier wordt de Bayesiaanse procedure gepresenteerd die is ontwikkeld om een tot expressie komend transcript van een niet tot expressie komend transcript te kunnen onderscheiden. Deze procedure gebruikt de waarneming dat de log variantie van de intensiteit sterk bimodaal verdeeld is. Dit maakt de berekening van de voorwaardelijke waarschijnlijkheid mogelijk dat een gen zich in de hoge variantie modaliteit bevindt. Deze per array bepaalde grens wordt geïnterpreteerd als het signaal niveau waarboven genen betrouwbaar gemeten kunnen worden en waarvoor daarmee een expressieniveau vastgesteld kan worden. De hier gegeven procedure is gebruikt in de analyse van de data van hoofdstuk 3 en 4.

### *HOOFDSTUK 4 TRANSCRIPTOOM DYNAMICA VAN VROEGE ZEBRAVIS EMBRYOGENESE BEPAALD IN EEN HOGE-RESOLUTIE TIJDREKES*

In de afgelopen jaren zijn vele ontdekkingen gedaan op het gebied van genexpressie in vroege embryogenese. In de Biological Perspective paragraaf van de introductie zijn verschillende van deze genoemd. Omtrent het gedrag van het transcriptoom in individuele embryo's zijn echter nog nauwelijks studies uitgevoerd. In dit hoofdstuk wordt een hoge resolutie genexpressie studie gepresenteerd met 180 individuele zebraavis embryo's, afkomstig van negen verschillende legfels (verschillende moeders en vaders), geordend naar ontwikkeling in de periode van late blastula tot midden gastrula. Gedurende deze periode van drie uur zijn embryo's willekeurig bemonsterd zodat gemiddeld per minuut ontwikkelingsstijd een embryo is geanalyseerd. Meer dan naar de biologische relevantie van gevonden profielen in cellulaire processen of pathways richt dit hoofdstuk zich op het identificeren van de transcriptoom dynamica van de 6,734 in vroege embryogenese tot expressie gebrachte genen. Het bestaan van een strak gereguleerd embryonaal transcriptoom programma wordt aangetoond, zelf wanneer individuen afkomstig van verschillende legfels worden vergeleken. In het algemeen vertoont gen expressie gedurende deze ontwikkelingsperiode een lage variabiliteit; veranderingen zijn geleidelijk. Er worden tien typen dynamisch gedrag geïdentificeerd. Bovendien worden voor een

aantal genen de exacte tijdstippen waarop genexpressie start of stopt vastgesteld. Deze startende en stoppende genen produceren in deze studie de wat concentratie betreft snelst veranderende transcripten (hebben de hoogste fold change). Op vier oscillerende genen na worden in deze periode van drie uur geen genen gevonden die een piekwaarde in hun expressie vertonen. Daarnaast vertonen een aantal genen een expressie die zich op twee of drie aparte niveaus afspeelt. Deze niveaus zijn sterk gerelateerd aan het legsel waaruit het embryo afkomstig is. Het sterke vermoeden bestaat dat dit legsel-specifieke transcriptoom effect de resultante is van een verzameling epigenetische factoren die genexpressie regulatie bepalen.

## *HOOFDSTUK 5 CELLULAIRE FACTOREN DIE DE TRANSCRIPTOOM DYNAMICA IN VROEGE ZEBRAVIS EMBRYOGENESE BEÏNVOEDEN*

In dit hoofdstuk worden de genexpressie typen welke in hoofdstuk 4 werden gedefinieerd en de verzameling genen die een legsel specificiteit vertonen verder geanalyseerd. De genexpressie typen worden op de volgende determinanten geanalyseerd: genoom locatie, grootte van de genen, promotor regionen voor specifieke transcriptie factor bindingsplaatsen, 3' UTR's voor cytoplasmatische polyadenylatie motieven en 3'UTRs voor miRNA doel sequenties. De overrepresentatie van specifieke biologische pathways en gen ontologieën werd getest. Bovendien werd de miRNA expressie van individuele embryo's omstreeks 50% epibolie (vroeg gastrula fase) door middel van een sequencing experiment bepaald, waarbij dezelfde legsels als in het experiment van hoofdstuk 4 gebruikt werden. De miRNA expressie van individuele embryonen gedurende de gehele zebrafish embryogenese, inclusief de miRNA expressie in onbevuchte eitjes en volwassen zebrafish werd bepaald in een ander sequencing experiment. We zien dat genen in genexpressie typen in specifieke pathways actief zijn. Vroege gastrulatie kan echter niet beschreven worden als het resultaat van een handvol mechanismen. Eerder is een overweldigende hoeveelheid processen betrokken bij de regulatie die in deze ontwikkelingsfase plaatsvindt. Niet alleen is de resulterende genexpressie strak gereguleerd, dit geldt ten minste ook voor een aantal van de regulatoren zelf. Een en ander wordt geïllustreerd door het gedrag van de miRNA miR-430. We concluderen dat zelfs in deze korte spanne ontwikkelingstijd het gedrag van de genexpressie typen onder controle staat van vele regulerende processen. Om dit gedrag verder te ontrafelen zullen we nog verder moeten inzoomen op het ontwikkelingsproces, in bijzonder wat betreft de locatie waarin de biologische processen zich afspelen en wat betreft het moleculaire signatuur van de transcripten.

### *TOT BESLUIT*

In dit laatste hoofdstuk worden de bevindingen in perspectief geplaatst; met welk resultaat hebben we genomische informatie geïntegreerd; wat is het lot van maternaal RNA; wat is het belang van de strakke regulatie van genexpressie zoals deze in de hoofdstukken 3 en 5 gevonden werd; hoe kunnen we de bescheiden veranderingen in gen-

expressie interpreteren? De discussie van de resultaten wordt gevoerd aan de hand van met name de volgende aspecten: cytoplasmatische polyadenylatie, conceptualisatie, tijdstippen waarop genexpressie start en stopt, de strakke regulatie van genexpressie zowel in tijd als wanneer individuen met elkaar worden vergeleken en de individuele signatuur van genexpressie. Deze discussie wordt gevoerd in de context van studies van anderen. Tenslotte geven we onze mening over mogelijke benaderingen om het inzicht in transcriptoom dynamica verder te verdiepen.