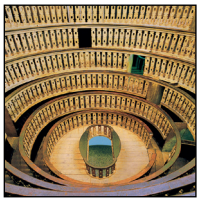
Transcriptome dynamics in early zebrafish embryogenesis

Rauwerda, J.

Publication date
2017
Document Version
Other version
License
Other

Concluding Remarks



Introduction

The research for this thesis has been carried out in part-time over a period of seven years. As a consequence, I have had the opportunity to follow, longer than in a regular PhD experience the advance in scientific fields such as bioinformatics, genomics and (zebrafish) embryology. During these years, in all of these fields progress was spectacular; a large amount of knowledge has been acquired, many new technologies both in the laboratory (wet-lab) as in silico (dry-lab) have been developed and have matured. Yet, many questions on early embryonic development remain, such as on the total or partial clearance of maternal RNAs or, and this is a much broader question, on a precise characterization of the molecular mechanisms that lead to the onset of gastrulation. Also, with the advancement of knowledge many new questions have arisen, such as questions on the biological function of the sub-cellular structures that can be observed at the onset of gastrulation [1] or on the biological function of different 3'UTRs that are used during development [2,3].

The aim we set at the start of our research was to get a deeper understanding of transcriptome dynamics; we wanted to know how much individual cells, such as eggs that are in the same physiological stage, differ. Also, we wanted to understand how, during a short period of development and measured at a resolution of minutes rather than hours, gene expression changes and how much individual zebrafish embryos differ from one another in that respect. In the course of this thesis' research we have at least partially answered these questions for two very specific phases in the development of zebrafish: the egg and gastrulation. The question how generalizable these answers are to other cellular and life's processes is open for debate. Undoubtedly, with the advancement of technology it will become easier to include aspects of dynamics and individual signature in biological research. In our opinion these aspects should, wherever feasible, be included.

Research questions on transcriptome dynamics carry a large technological component. In our studies we used new technologies, such as the small-RNA sequencing methods applied in Chapter 5 and made technological contributions ourselves, e.g. by the development of the RNA-isolation method presented in Chapter 1 [4]. Advancements in technology, such as techniques in the area of microscopy [1], applications of sequencing technology [5–10], single-cell analysis [11] and RNA tomography [12] make it possible to study the developing embryo at an ever higher resolution in time, in space and with respect to molecular signatures. Yet, our own research, as well as the aforementioned studies reveal a fundamental problem in biological research: the way life is organized makes it necessary to find ways to study interconnected (i.e. in a tissue, organ or organism) and more or less autonomous units (e.g. cells). On submission of our study on single zebrafish embryos (Chapter 4) to a scientific journal one reviewer commented that the presented gene-expression types hold limited informative value, because in our study we looked at populations of cells, i.e. one whole embryo at the time. Of course he/she is right to note that a zebrafish embryo contains many cells and cell populations; however, in comparison to existing studies [5,6,9,10] we profiled single

individual embryos rather than pools of individuals and, to our knowledge we did this for the first time ever. When studying fundamental biological processes, it can indeed be fruitful to investigate these on the level of an individual cell. But by profiling only one or at most a few cells, the understanding of functioning at higher levels is not automatically generated. Moreover, single cells are not the most fundamental unit at which biology can be studied. In Chapter 3 we studied single eggs which are single cells (if we ignore the small polar bodies). Eggs are highly complex and dynamic structures with many subcellular compartments containing different transcripts [13,14]. Locati et al. (personal communication) identified a specific maternal type of 5S and 45S rRNA in zebrafish eggs that probably are processed in different ways than their maternal counterparts, possibly in different compartments. Hence, if we want to study the processing of these different rRNA types, single cell analyses are not sufficient. The further we zoom in, the more we learn about the underlying fundamental processes, but this is at the cost of the overall picture. Investigating transcription dynamics on the level of subcellular structures, cells, tissues and so on will help us to understand and model transcription and transcription regulation, as reviewed by [15]. To do this we need temporal and spatial data of unsurpassed detail. New techniques such as RNA tomography [12] and RNA tracking in live cells [16], as well as the advancements in next-generation sequencing will help us with this and hopefully will enable us to eventually understand how life starts. Questions like the earlier stated one on the precise characterization of molecular mechanisms that lead to the onset of gastrulation are but a tiny part of this large endeavor.

The elusive role of cytoplasmic polyadenylation

During oogenesis the oocyte transcriptome is produced and probably acquires, while in the nucleus a long poly(A) tail (~200-250 nucleotides) [17]. On export to the cytoplasm a complex containing the cytoplasmic polyadenylation element-binding protein (CPEB) and several other factors shorten or remove the poly(A) tail. This has been shown for Xenopus [18], and indirectly also for zebrafish [2,5]. Prior to the maternal to zygotic transition (MZT) most transcripts contain no or only short poly(A) tails [19] but do contain elements in their 3'UTR that allow for binding into a stabilizing conformation [18]. CPEB is also involved in the (re)-polyadenylation and does so by binding onto cytoplasmic polyadenylation elements in the 3'UTR of a transcript [17]. While maternal mRNAs are enriched in their 3'UTRs for miR-430 targets, miR-430 genes are among the first zygotic genes that are expressed from the 128-cell stage onward [8]. MiR-430 induces rapid deadenylation of targeted genes [20]. Estimates on the scale of clearance of maternal transcripts exist and often assume a total clearance of maternal RNAs., Walser and Lipshitz [21] mention, based on the data of [22] 36% clearance at fertilization, 42% after the MZT and 23% degraded later on in embryogenesis. A problem with this approach is that the authors implicitly exclude the possibility of cytoplasmic deadenylation and polyadenylation after fertilization in their calculations because they base their estimates on data produced with an oligo(dT) protocol. This protocol is strongly biased against short or absent poly(A) tails. In Chapter 4 we learned that expression changes

on average are below 1.3-fold change per hour. Extrapolation of the majority of the gene expression profiles that either stay level or increase during the early gastrula stage (Types 5 and 7), would position the starting of gene expression prior to fertilization (Chapter 4). Moreover, by a comparison of the results from Chapter 3 and Chapter 4, it became clear that 87% of the expressed genes that are expressed from late blastula to mid-gastrula are also expressed in unfertilized egg. This finding is in line with the finding of Lee et al. [23]. In our opinion it is unlikely that from the MZT onward until early gastrula, a period of approximately 2 hours, a massive expression of zygotic genes takes place at a rate that is generally not observed in the post-blastula embryos while simultaneously nearly identical maternal transcripts are cleared. Although it is a fact that with the expression of miR-430 clearance of maternal transcripts takes place and that zygotic transcripts can differ from their maternal counterparts by a different and in general longer 3'UTR [2,3], we think it is highly likely that maternal transcripts are used and re-used in the late blastula stage and beyond.

Oversimplifying complex biological processes

A total clearance of maternal RNA by early gastrula however is conveyed by many of the visualizations on the concept of the maternal to zygotic transition (MZT), e.g. Figure 1 in [24] or Figure 1A in [5]. Another concept implied by these visualizations is that a change in maternal transcript abundance is simultaneously counteracted by an increase in zygotic transcript abundance, thus suggesting a constant transcript abundance. In general this is not true, at MZT a peak in the amount of polyadenylated transcripts in an embryo is observed [9]. Although biological cartoons are a useful means to convey biological concepts, it is also important to be aware of their pitfalls, such as the one observed in the visualization of the MZT by the aforementioned two examples. Hierarchical decomposition calls for categorization, which in turn demands neglecting certain sample specifics e.g. by using smoothing procedures or by the use of cut-off values. Aanes et al. provide, based on a subset of developmentally expressed genes seven clusters of early developmental genes [5]. These clusters of genes later were used in studies on other aspects of early development such as in [2]. We have looked at the profiles of transcripts in these gene clusters in our data and observed a high variability within them. The seven cluster profiles only represent the behavior of individual transcripts to some extent and as such give a limited view on the transcriptome dynamics of these categories. We noticed that especially in the study of transcriptome dynamics it is a problem to generalize observations into concepts. This is one of the reasons why in Chapter 4 we rather give examples of genes that are part of a gene expression type than to use e.g. average or eigen-gene profiles of the entire type. Nevertheless, and aware of these pitfalls, we presented the generalization into 10 gene expression types during late blastula and early gastrulation. The message that is conveyed here is also about the gene-expression types that are not found during this period such as genes that show a half period or one entire period. These generally gradual changes in gene expression during gastrulation, a period in which the embryo differentiates from a single layered to a three layered form, has surprised us.

In Chapter 5 many cellular factors were identified that play a role in the gene expression within the gene expression types, but many of them were only affecting a subset of genes within the type. Hence, the similar behavior in gene-expression types is the result of many factors that are at work and that differ from gene to gene. To understand the biological processes that are at work, it is necessary to be able to look at the gene expression of individual genes. For this reason we set up the website (http://genseq-h0.science.uva.nl/shiny/Dr-Browser-v2) that provides this information for the studies in this thesis.

The importance of knowing the starting and stopping points of gene expression

One of the assets of the study presented in Chapter 4 is the identification of genes which expression starts (Type 8, Type 9) or stops (Type 1, Type 2) during late blastula and early gastrula. For many genes we were able to identify the exact gene-expression starting and stopping time points. This is important, because in principle it would enable to study pathways not as a monolithic block of genes, but to approach it as processes that enfold in time. Such approaches in which changes of gene expression are evaluated by their meaningfulness in the pathway (e.g. with respect to order and movement, up or down or tissue in which the pathway is deployed) have been proposed earlier [25], but the resolution of the data until now has been such that success of these approaches is still limited. During our studies we tried to combine our gene-expression results from Chapter 4 with the recently published RNA tomography dataset [12]. With the high resolution spatial data, it is possible to pinpoint gene expression to a location in the embryo. Together with our information on starting and stopping genes, in theory it should be possible to infer the dynamics of a pathway in time. Although we were successful in linking some KEGG pathways, spatial expression clusters and our data, we have not been able to infer a 'pathway unfolding in time'. We think this is due to a lack of sequencing depth; in 3D tomography the depth of the sequencing in 503 voxels is determined by the 50 + 50 + 50 tomography slices that are sequenced. Nevertheless, the combination of RNA tomography experiments sequenced at an ultra-high depth together with high time-resolution data is a very attractive approach for pathway reconstruction as it does not require any modification in the biological system under study (cf. the tracking of RNA in live cells with CRISPR/Cas9 [16]).

Tight regulation in embryogenesis

One of the biggest surprises in our study was to discover that the regulation of the zebrafish egg transcriptomes as well as the embryonic transcriptomes is extremely tight. When we compare eggs from different mothers the maternal message appears to be highly similar. Even in embryos that stem from different spawns, and that under the microscope markedly display individual morphological features (cf. Introduction, Figure 1), the far majority of genes around early gastrula shows a strict regulation. Biological

processes are sometimes characterized by an hourglass or bow-tie architecture, defined as 'a feature of multi-layered networks in which the intermediate layer has significantly fewer components than the input and output layers' [26]. We have observed three aspects in the gene regulation in early gastrula: i) transcription levels are regulated in a very tight manner; ii) the number of output layers is large (cf. the number of processes unfolding in early gastrulation, Chapter 5, Table 2), and iii) the number of input layers also appears to be large, one of the conclusions in Chapter 5 being that 'an overwhelming amount of processes [is] involved in regulation at this developmental stage. Gene expression therefore can be considered as the knot of the bow-tie. A remarkable observation in Chapter 5 was that at least one of the many regulators that control the abundance of transcripts, miR-430 was in its turn again tightly regulated (Chapter 5, Figure 5): another knot in another bow-tie architecture. It is not unlikely that this is a more generic principle throughout early embryogenesis. Such bow-ties do not only explain the robustness in development, but identification of them can also help to predict fragilities in the system that in general can be found in the knots [27].

The signature of the individual

One of the reasons to focus on the profiling of individual eggs and individual embryos was that we could not answer the very basic question on the admissibility of using pools of individuals in transcriptomics experimentation. Since the use of pools has been and is quite common in biological research, this is an important question. For zebrafish eggs and zebrafish early embryogenesis we can now answer that question. In general, by using pools of samples for most genes not a large error is introduced. In eggs we found that although almost all genes show a significant 'mother' effect, this contribution is generally small. Thus, depending on the research question, if the use of pools is unavoidable, it might be wise to use pools of eggs from the same mother. In gastrulation some genes show a spawn specific effect, but this effect is more intricate and is less pronounced than in the maternal message, the latter observation also made by [9]. In general, if we want to study genes which expression is depending on the individual, we should not use sample pools. Moreover, in statistical terms it is more informative to have 10 separate sequencing runs on 10 individuals than a 10 times as deep sequencing run on the pool of these individuals. At least having data on individuals available will help us to identify genes that are expressed at an individual level. On the importance of these genes we don't have a definitive answer. Fact is that these genes are outside the knot in the bow-tie and as a group are involved in many different processes.

Conclusion

During this PhD study, we learned a lot about zebrafish embryogenesis, the associated gene-expression and transcriptome dynamics. Nevertheless, we are left with more questions than we started with, which renders the uneasy feeling to know less now about the subject than at the beginning of this endeavor. This might be an unavoidable aspect

of science, certainly when you start to dissect a biological system, as complicated as embryogenesis. On the other hand, it is an exciting prospect that there are so many questions still that need to be answered. After all, answering them demands entering unmapped territory in technology as well as in biology.

References

1. Keller PJ, Schmidt AD, Wittbrodt J, Stelzer EHK. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. Science. American Association for the Advancement of Science; 2008;322: 1065–9. Available: http://science.sciencemag.org/content/322/5904/1065.abstract
2. Ulitsky I, Shkumatava A, Jan CH, Subtelny AO, Koppstein D, Bell GW, et al. Extensive alternative polyadenylation during zebrafish development. Genome Res. 2012;22: 2054–66. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3460199&tool=pmcentrez&rendertype=abstract
3. Li Y, Sun Y, Fu Y, Li M, Huang G, Zhang C, et al. Dynamic landscape of tandem 3' UTRs during zebrafish development. Genome Res. 2012;22: 1899–906. doi:10.1101/gr.128488.111
4. de Jong M, Rauwerda H, Bruning O, Verkooijen J, Spaink HP, Breit TM. RNA isolation method for single embryo transcriptome analysis in zebrafish. BMC Res Notes. 2010;3: 73. doi:10.1186/1756-0500-3-73
5. Aanes H, Winata CL, Lin CH, Chen JP, Srinivasan KG, Lee SGP, et al. Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. Genome Res. 2011;21: 1328–38. doi:10.1101/gr.116012.110
6. Vesterlund L, Jiao H, Unneberg P, Hovatta O, Kere J. The zebrafish transcriptome during early development. BMC Dev Biol. 2011;11: 30. Available: http://www.biomedcentral.com/1471-213X/11/30
7. Harvey SA, Sealy I, Kettleborough R, Fenyes F, White R, Stemple D, et al. Identification of the zebrafish maternal and paternal transcriptomes. Development. 2013;140: 2703–10. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3678340&tool=pmcentrez&rendertype=abstract
8. Heyn P, Kircher M, Dahl A, Kelso J, Tomancak P, Kalinka AT, et al. The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep. The Authors; 2014;6: 285–292. doi:10.1016/j.celrep.2013.12.030
9. Owens NDL, Blitz IL, Lane MA, Patrushev I, Overton JD, Gilchrist MJ, et al. Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. Cell Rep. 2016;14: 632–647. doi:10.1016/j.celrep.2015.12.050
10. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. Genome Res. 2012;22: 577–91. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3290793&tool=pmcentrez&rendertype=abstract
11. Tang F, Lao K, Surani MA. Development and applications of single-cell transcriptome analysis. Nat Methods. Nature Research; 2011;8. doi:10.1038/nmeth.1557
12. Junker JP, Noë ES, Guryev V, Peterson K a, Shah G, Huisken J, et al. Genome-wide RNA Tomography in the Zebrafish Embryo. Xell. 2014;159: 662–675. doi:10.1016/j.cell.2014.09.038
13. Wester P, Vos J. Identification of Endocrine Disruptive Effects in the Aquatic Environment-a Partial Life Cycle Study in Zebrafish [Internet]. RIVM teport 640920001/2003. 2003. Available: http://rivm.openrepository.com/rivm/handle/10029/9047
14. Leung CF, Webb SE, Miller AL. On the mechanism of ooplasmic segregation in single-cell zebrafish

embryos. Dev Growth Differ. 2000;42: 29–40. Available: http://www.ncbi.nlm.nih.gov/pubmed/10831041

15. Coulon A, Chow CC, Singer RH, Larson DR. Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat Publ Gr. 2013;14. doi:10.1038/nrg3484
16. Nelles DA, Fang MY, O'Connell MR, Xu JL, Markmiller SJ, Doudna JA, et al. Programmable RNA Tracking in Live Cells with CRISPR/Cas9. Cell. Elsevier; 2016;165: 488–96. doi:10.1016/j.cell.2016.02.054
17. Richter JD. CPEB: a life in translation. Trends Biochem Sci. 2007;32: 279–85. Available: http://www.ncbi.nlm.nih.gov/pubmed/17481902
18. Simon R, Tassan JP, Richter JD. Translational control by poly(A) elongation during Xenopus development: differential repression and enhancement by a novel cytoplasmic polyadenylation element. Genes Dev. 1992;6: 2580–91. Available: http://www.ncbi.nlm.nih.gov/pubmed/1285126
19. Meijer H a., Bushell M, Hill K, Gant TW, Willis a. E, Jones P, et al. A novel method for poly(A) fractionation reveals a large population of mRNAs with a short poly(A) tail in mammalian cells. Nucleic Acids Res. 2007;35: e132–e132. doi:10.1093/nar/gkm830
20. Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, et al. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science. 2006;312: 75–9. doi:10.1126/science.1122689
21. Walser CB, Lipshitz HD. Transcript clearance during the maternal-to-zygotic transition. Curr Opin Genet Dev. 2011;21: 431–43. Available: http://www.sciencedirect.com/science/article/pii/S0959437X11000621
22. Mathavan S, Lee SG, Mak A, Miller LD, Murthy KR, Govindarajan KR, et al. Transcriptome analysis of zebrafish embryogenesis using microarrays. PLoS Genet. Genome Institute of Singapore, Singapore.; 2005;1: 260–276. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16132083
23. Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES, et al. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature. 2013;503: 360–4. doi:10.1038/nature12632
24. Schier AF. The Maternal-Zygotic Transition: Death and Birth of RNAs. Science (80-). 2007;316: 406–407.
25. Draghici S, Khatri P, Tarca AL, Amin K, Done A, Voichita C, et al. A systems biology approach for pathway level analysis. Genome Res. Cold Spring Harbor Laboratory Press; 2007;17: 1537–45. doi:10.1101/gr.6202607
26. Friedlander T, Mayo AE, Tlusty T, Alon U. Evolution of bow-tie architectures in biology. PLoS Comput Biol. Public Library of Science; 2015;11: e1004055. doi:10.1371/journal.pcbi.1004055
27. Csete M, Doyle J. Bow ties, metabolism and disease. Trends Biotechnol. 2004;22: 446–450. doi:10.1016/j.tibtech.2004.07.007