

**The changing roles of *Hox3* genes in insect evolution:
characterizing the *zen* paralogues in the beetle
*Tribolium castaneum***

In a u g u r a l - D i s s e r t a t i o n

zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln



vorgelegt von

Daniela Gurská

aus Košice

Köln, 2017

Berichterstatter:

Dr. Kristen A. Panfilio

Prof. Dr. Siegfried Roth

Vorsitzender der Prüfungskommission:

Prof. Dr. Thomas Wiehe

Tag der mündlichen Prüfung:

14.12.2017

TABLE OF CONTENTS

ABSTRACT	7
1 INTRODUCTION	9
1.1 The evolution and origin of <i>zen</i>	9
1.1.1 Changes during the switch from Hox3 (embryonic) to Zen (extraembryonic).....	9
1.1.2 When did the switch from Hox3 (embryonic) to Zen (extraembryonic) occur?.....	13
1.2 Extraembryonic membranes: evolution and function.....	14
1.2.1 Functions of <i>zen</i> in the extraembryonic membranes	16
1.3 Two possible evolutionary scenarios lying behind diversification of <i>zen</i> functions.....	17
1.4 Functional diversification of paralogues during insect evolution	17
1.5 Correlation of extraembryonic membranes evolution with the evolution of <i>zen</i>.....	18
1.6 Insect model organism <i>Tribolium castaneum</i>	21
1.6.1 Extraembryonic development during embryogenesis of <i>Tribolium castaneum</i>	21
1.7 Aims of the study	24
2 MATERIAL AND METHODS	25
2.1 <i>Tribolium castaneum</i> stock maintenance	25
2.1.1 <i>Tribolium castaneum</i> husbandry	25
2.1.2 Egg collection.....	25
2.1.3 Egg dechoriation.....	25
2.1.4 Fast freeze of eggs.....	26
2.1.5 Fixation and devitellination.....	26
2.2 Basic molecular methods	26
2.2.1 RNA extraction	26
2.2.2 cDNA synthesis.....	27
2.2.3 Primer design.....	27
2.2.4 Polymerase chain reaction.....	30
2.2.5 TA cloning.....	31
2.3 Quantitative reverse transcription PCR.....	31
2.3.1 RT-qPCR data analysis	32
2.4 <i>In situ</i> hybridization	33
2.4.1 Probe synthesis	33
2.4.2 Probe hybridization	33
2.4.3 Digoxigenin antibody incubation	33
2.4.4 Colorimetric detection.....	34
2.5 Gene expression silencing	34
2.5.1 Double stranded RNA synthesis.....	34
2.5.2 Parental RNA interference	35
2.6 Phenotypic scoring after RNA interference	35
2.6.1 Nuclear staining.....	35
2.6.2 Cuticle preparation	36
2.6.3 Serosal cuticle integrity determination	36
2.7 Protein expression	36
2.7.1 Protein extraction	36

2.7.2	Protein concentration measurement	37
2.7.3	SDS-PAGE.....	37
2.7.4	Western blotting	38
2.7.5	Cryo-sectioning	39
2.7.6	Antibody staining	39
2.8	Visualization of specimens	40
2.8.1	Sample mounting for microscopy	40
2.8.2	Microscopy and picture processing	40
2.9	<i>In silico</i> analysis	40
2.9.1	Sequence alignment.....	40
2.9.2	Identification of conserved non-coding regions	41
2.10	RNA-sequencing after RNA interference.....	41
2.10.1	RNA-sequencing	41
2.10.2	Sample preparation for RNA-sequencing experiments	42
2.11	Generating pipeline for RNA-sequencing data analysis	43
2.11.1	Joining the files	43
2.11.2	Ranking test.....	44
2.12	RNA-sequencing data analysis	44
2.12.1	Quality control.....	44
2.12.2	Trimming.....	44
2.12.3	Filtering overrepresented sequences of ribosomal and mitochondrial RNA	45
2.12.4	Mapping	45
2.12.4.1	Mapping to official gene set	45
2.12.4.2	Mapping to the genome	46
2.12.5	Feature counting	46
2.12.6	Filtering out genes with low read count	47
2.12.7	Differential expression analysis	47
2.12.8	Principal component analysis	47
2.12.9	Analysis of shared targets.....	47
2.12.10	Gene ontology term analysis	48
3	METHOD DEVELOPMENT.....	53
3.1	Generating a pipeline for RNA-sequencing and differential expression data analysis	53
3.1.1	Quality control, trimming and filtering	53
3.1.2	Mapping RNA-sequencing data to the genome and to the official gene set.....	54
3.1.3	Differential expression analysis	55
3.1.3.1	Influence of the read length	55
3.1.3.2	Influence of the program used for the differential expression analysis.....	57
4	RESULTS.....	60
4.1	Hox3 locus sequence conservation	60
4.1.1	Investigation of conserved non-coding regions between <i>Tc-zen1</i> and <i>Tc-zen2</i>	60
4.1.2	Investigation of conserved non-coding regions in Hox3 locus of four closely related <i>Tribolium</i> species.....	62
4.2	<i>Tc-zen1</i> and <i>Tc-zen2</i> wild type expression dynamics.....	66
4.2.1	Expression domains of <i>Tc-zen1</i> and <i>Tc-zen2</i> during early embryogenesis	66
4.2.2	Expression profile of <i>Tc-zen1</i> and <i>Tc-zen2</i> transcript during early embryogenesis	69
4.2.3	Expression profile of <i>Tc-zen1</i> and <i>Tc-zen2</i> transcript during late embryogenesis.....	71
4.2.4	Spatial and temporal protein expression profiles of Tc-Zen1 and Tc-Zen2	72

4.2.5	Cellular localization of Tc-Zen2 transcription factor	76
4.2.6	Extraembryonic Tc-Zen2 protein expression	79
4.3	Characterization of <i>Tc-zen1</i> and <i>Tc-zen2</i> phenotypes after parental RNA interference	82
4.3.1	Detailed characterization of <i>Tc-zen1</i> knockdown.....	82
4.3.2	Knockdown strength and phenotypic penetrance after <i>Tc-zen1</i> ^{RNAi}	83
4.3.3	Detailed characterization of <i>Tc-zen2</i> knockdown.....	84
4.3.4	Knockdown strength and phenotypic penetrance after <i>Tc-zen2</i> ^{RNAi}	87
4.3.5	Potential regulatory interactions between <i>Tc-zen1</i> and <i>Tc-zen2</i>	90
4.3.6	Possible off target knockdown effects of <i>Tc-zen1</i> long dsRNA fragment on <i>Tc-zen2</i> expression	92
4.4	Global evaluation of <i>Tc-zen</i> genes' targets by RNA-sequencing after RNA interference.....	95
4.4.1	Variance between wild type and knockdown samples of early developmental stages	95
4.4.2	Identification of candidate target genes of <i>Tc-zen1</i> and <i>Tc-zen2</i> during early embryogenesis.....	96
4.4.3	Evaluation of potential target genes of <i>Tc-zen1</i> and <i>Tc-zen2</i>	96
4.4.4	<i>Tc-zen1</i> candidate target genes	97
4.4.5	<i>Tc-zen2</i> candidate target genes	103
4.4.6	Does <i>Tc-zen2</i> copy <i>Tc-zen1</i> function during early embryogenesis?.....	109
4.4.7	Variance between wild type and knockdown samples of late developmental stages	112
4.4.8	Identification of <i>Tc-zen2</i> candidate target genes during late embryogenesis	113
4.4.9	Functional profile of <i>Tc-zen2</i> candidate target genes of late development.....	113
4.4.10	Evaluating differential expression of <i>Tc-zen2</i> candidate targets in late development	121
5	DISCUSSION.....	124
5.1	Conservation levels of non-coding regions between <i>zen</i> genes	124
5.1.1	Promoters of <i>Tc-zen1</i> and <i>Tc-zen2</i> differ in sequence	124
5.1.2	Several conserved non-coding regions between four <i>Tribolium</i> species were identified	125
5.2	Transcriptional and translational regulation differ between the <i>Tc-zen</i> paralogues	125
5.2.1	<i>Tc-zen1</i> is transiently expressed in embryo	125
5.2.2	<i>Tc-zen1</i> and <i>Tc-zen2</i> are not maternally supplied.....	126
5.2.3	<i>Tc-zen1</i> and <i>Tc-zen2</i> expression peak is only during early embryogenesis.....	127
5.2.4	Minimal <i>Tc-zen2</i> expression is sufficient for protein turnover through late embryogenesis	127
5.2.5	Tc-Zen2 is localized only to nucleus during early and late embryogenesis	127
5.2.6	Tc-Zen2 is expressed exclusively in the serosa during entire lifespan of extraembryonic membranes	128
5.3	Knockdowns of the <i>Tc-zen</i> paralogues differ in strength and cause different phenotypes.....	128
5.3.1	<i>Tc-zen1</i> and <i>Tc-zen2</i> knockdowns result in two distinct phenotypes	128
5.3.2	Knockdown strength of <i>Tc-zen1</i> ^{RNAi} is higher than the one of <i>Tc-zen2</i> ^{RNAi}	129
5.3.3	<i>Tc-zen</i> paralogues are mutual downstream targets	130
5.4	<i>Tc-zen</i>'s functions are separated to early and late development	130
5.4.1	Knockdown of <i>Tc-zen2</i> in early stages does not have a robust transcriptional effect	130
5.4.2	<i>Tc-zen2</i> has subtle early regulatory role	131
5.4.3	Defects in extraembryonic development were not observed after knockdown of <i>Tc-zen1</i> and <i>Tc-zen2</i> candidate targets of early embryogenesis	131
5.4.4	<i>Tc-zen2</i> does not copy <i>Tc-zen1</i> 's function during early embryogenesis	132
5.4.5	Functional profile of <i>Tc-zen2</i> candidate target genes of late development was retrieved	133
5.4.6	RT-qPCR miniscreen confirmed results obtained from differential expression analysis	135
5.5	Distinct <i>Tribolium zen</i> functions most likely arose through sub-functionalization	135
5.6	Changes in protein sequence and features enabled switch from Hox3 to Zen.....	136
5.7	Conclusion.....	138

6	OUTLOOK	139
7	REFERENCES	141
8	SUPPLEMENT	150
	ZUSAMMENFASSUNG	153
	ACKNOWLEDGEMENTS	155
	ERKLÄRUNG	157
	CURRICULUM VITAE	158

ABSTRACT

Hox genes encode transcription factors responsible for the determination of axial patterning of all bilaterian embryos. However, insect *Hox3* orthologues, named *zerknüllt* (*zen*), have changed their function multiple times, which led to the abandonment of the canonical *Hox* function, and subsequent switch of their functional domain from embryonic to extraembryonic tissue. To date, in fact, all described *zen* genes play role in extraembryonic membranes (EEMs). The EEMs protect the embryos from the insults of the outer environment and their formation allowed insects to oviposit in various niches, ultimately allowing them to colonize land. The evolution of the EEMs is tightly linked to the evolution of *Hox3/zen*. Concurrently with the origin of EEMs, *Hox3* has gradually switched from embryonic role to *zen*'s function in the EEMs. However, it is only within winged insect that the complete transition from *Hox3* to *zen* and complete EEMs are observed. Further, besides switching to extraembryonic tissue, in this new domain, *zen* genes have also acquired two different functions: one in early tissue specification and the other in late morphogenesis. However, little is known about the causes triggering the switch from *Hox3* to *zen*, and the subsequent functional divergence of *zen*. Here, in order to get insight into what has triggered the functional divergence of *zen*, I focused on the holometabolous beetle *Tribolium castaneum*, as two functionally diverged paralogues were described: one with the function during early embryogenesis (*Tc-zen1*) and the second one with the function during late embryogenesis (*Tc-zen2*).

In order to decipher how the two diverged functions of *Tc-zen1* and *Tc-zen2* were acquired, I investigated transcriptional and translation regulation of both *Tc-zen* genes during early and late embryogenesis. I showed that, although the early function was described only for *Tc-zen1*, both paralogues reach their expression peak during early embryogenesis. To reveal the degree of divergence in transcriptional targets between the paralogues during early development, I knocked down (via parental RNA interference, pRNAi) the *Tc-zen* genes and performed RNA-sequencing (RNA-seq). Differential expression (DE) analysis and the subsequent comparative analysis of the identified targets of *Tc-zen1* and *Tc-zen2* suggest that the paralogues do not share substantial number of transcriptional targets during early embryogenesis. Additionally, principal component analysis revealed that despite the early expression of both paralogues, the impact of *Tc-zen2* knockdown on early transcriptional control was significantly lower than for *Tc-zen1*, which is consistent with *Tc-zen2* having a late function. Nonetheless, the analysis of expression levels of each *zen* gene in knockdown samples of its paralogue revealed a subtle regulatory function of *Tc-zen2* during early embryogenesis, particularly in repression of *Tc-zen1*.

To further investigate transcriptional regulation by *Tc-zen2* during late embryogenesis, I have first showed that after the *Tc-zen2* expression peak is reached during early embryogenesis, the low transcript expression persists until the late development. Consistent with the timing of the transcript expression, I showed that Tc-Zen2 protein is present until the late developmental stage, where its function takes place. To identify transcriptional targets of *Tc-zen2* during late embryogenesis, I performed the second RNA-seq after pRNAi experiment. DE analysis revealed much higher impact of *Tc-zen2* on transcriptional control

during late embryogenesis than during early embryogenesis. The functional profile of candidate target genes of *Tc-zen2* during late embryogenesis was obtained by thorough gene ontology (GO) term analysis. Consistent with the phenotypic manifestation of the morphogenesis function of *Tc-zen2* during late embryogenesis, many of the identified candidate targets were assigned to GO terms with function in epithelial morphogenesis.

In conclusion, the results obtained within the presented project suggest that acquisition of the two distinct functions of *Tc-zen* paralogues might be partially explained by two different transcriptional signatures they attained. While the function of *Tc-zen1* temporally correlates with its expression peak and transcriptional regulation of its downstream targets, *Tc-zen2*, although expressed early, has very low impact on the downstream transcriptional regulation during early embryogenesis. Moreover, the fact that *Tc-zen* paralogues share very few targets during early embryogenesis suggests only subtle early regulatory roles of *Tc-zen2* and separation of its morphogenesis function to late embryogenesis. This result was further endorsed by observation of *Tc-zen2* transcript and protein expression throughout embryogenesis until the *Tc-zen2* late function takes place. In addition, I identified a much higher number of *Tc-zen2* candidate transcriptional targets during late embryogenesis, of which many likely play roles in epithelial morphogenesis. These diverse lines of evidence suggest that the diverged functions of *Tc-zen1* and *Tc-zen2* might have been acquired by regulation of different downstream transcriptional targets, which could have ultimately allow for separation of *Tc-zen* paralogues functions to early and late development.

1 INTRODUCTION

1.1 The evolution and origin of *zen*

Evolutionary developmental studies of vertebrates and invertebrates indicate that the same genetic toolkit is used repeatedly for the construction of diverse animal body plans. During evolution, the genes are recycled and their regulation is altered to an extent that new developmental roles are acquired. But how is it possible that a gene loses an important developmental role without drastically altering embryogenesis?

The most common textbook example of a conserved genetic toolkit is the example of homeobox (Hox) genes. Hox genes have emerged as master regulators of development as they encode transcription factors responsible for the determination of the diverse body plans of all developing bilaterian embryos. Hox genes are organized on chromosomes into clusters and are expressed in a colinear fashion, which means that their position on the chromosome corresponds to segment identity within the embryo along the anterior-posterior axis (Lewis, 1978). Hox genes have been highly conserved throughout evolution. Not only do they share high sequence similarity, but Hox transcription factors also share the same protein features: e.g. each of the Hox genes possesses a 180 bp long homeobox, which encodes a 60 amino acid long homeodomain forming three α -helices, through which the Hox transcription factors bind DNA (Scott et al., 1989). The common organization of Hox genes on chromosome, their colinear expression, the sequence conservation and the same protein features suggest that the rise of Hox gene classes and the distinct subclasses dates back before the insect and vertebrate lineages split. This assumption is well supported by the functional equivalence studies, which show that some of the human (McGinnis et al., 1990) and mouse (Malicki et al., 1990; Zhao et al., 1993) Hox genes are able to functionally substitute the Hox cognates in *Drosophila*.

Nonetheless, despite the high conservation of developmental role in axial patterning, class 3 *Hox* genes have evolved so dramatically during insect evolution that they lost the canonical function, and, during embryogenesis, they show neither embryonic nor colinear expression. Instead, insect *Hox3* orthologues, known as *zerknüllt* (*zen*), have acquired new functional domain in extraembryonic tissue, specifically in extraembryonic membranes (EEMs) (Hughes and Kaufman, 2002b). What exactly triggered the change of such a conserved function and when exactly did this switch occur?

1.1.1 Changes during the switch from *Hox3* (embryonic) to *Zen* (extraembryonic)

Like canonical Hox genes, the insect *Hox3* orthologue *zen* possesses the homeobox, which encodes the homeodomain responsible for binding to the DNA. The position that *zen* occupies within the Hox cluster (*Hox3* locus) serves as evidence that, although *zen* diverged in function, it indeed derived from *Hox3* gene. Despite this fact, the alignment of the *Hox3* and *Zen* protein sequences of different bilaterian species is not possible outside of their homeodomain. Even the alignment outside of the homeodomain of *Zen* proteins themselves does not show well conserved sequence motifs (Panfilio and Akam, 2007).

During insect evolution, *zen* has undergone multiple rounds of independent lineage specific duplications generating two or more copies of *zen* genes. In the *Drosophila* lineage, *zen* has undergone two rounds of duplication, resulting in additional two *Hox3* orthologues: *zen2* and *bicoid* (Rushlow et al., 1987). Nonetheless, even protein sequences of the duplicates of *Drosophila melanogaster* are not alignable outside of the homeodomain (Panfilio et al., 2006). The only exception so far described is the one of *Tribolium castaneum* Zen paralogues, where a high level of amino acid sequence conservation outside of the homeodomain is observed. This is due to the fact that *T. castaneum zen* paralogues derived from a recent tandem duplication (Panfilio et al., 2006).

Moreover, even within the homeodomain itself, the sequences of Hox3/Zen proteins are not strongly conserved. The alignment of amino acid sequence of the homeodomains across bilaterian species shows that the insect Zen homeodomain sequences differ from one another considerably more than the Hox3 homeodomain sequences among vertebrates (Fig. 1.1) (Falciani et al., 1996; Panfilio et al., 2006; Panfilio and Akam, 2007).

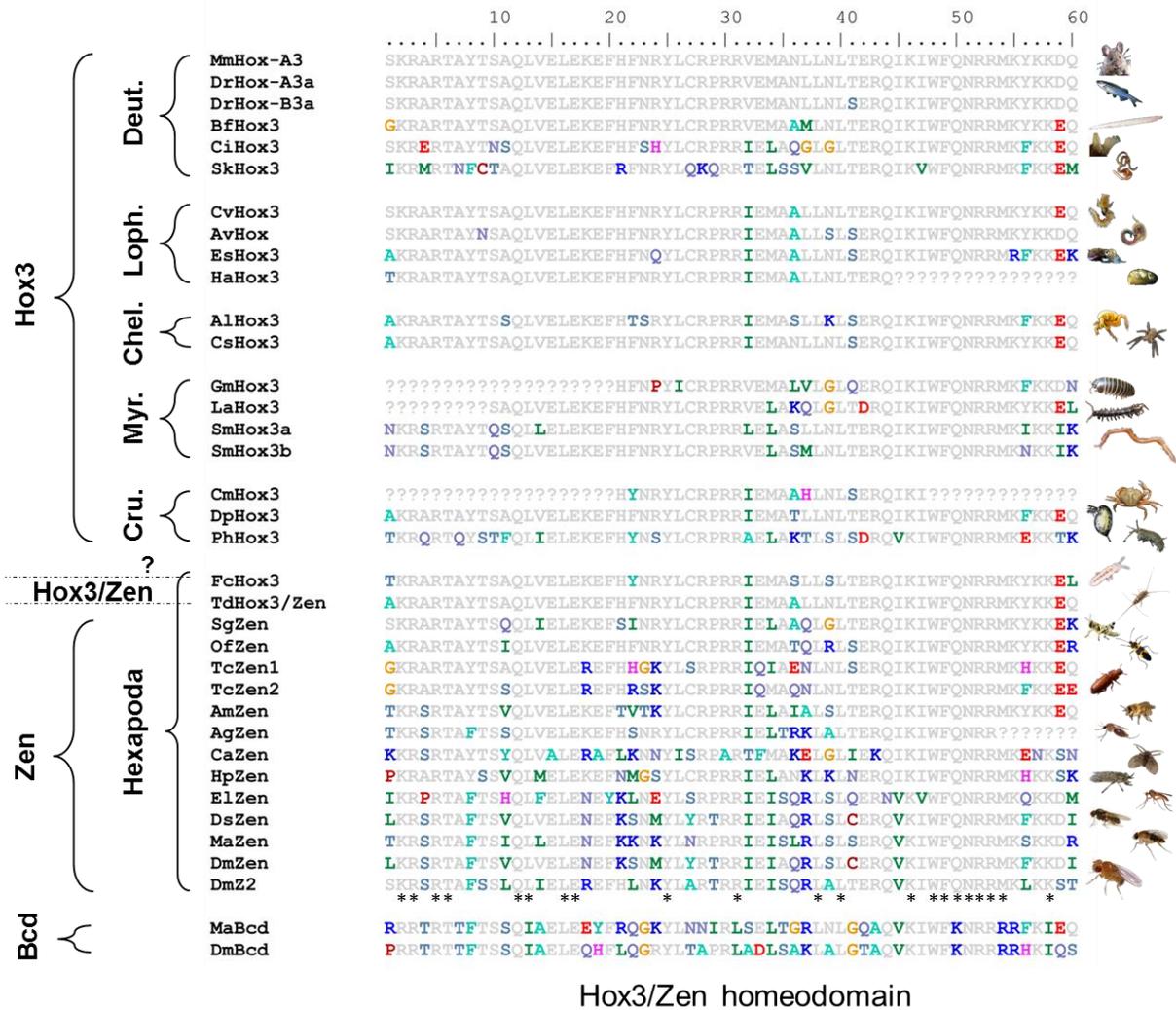


Figure 1.1. Alignment of amino acid sequences of class 3 Hox transcription factors homeodomains. Genes are grouped based on expression data (*Hox3* - embryonic vs. *zen* - extraembryonic) and according to the taxonomy order; gene types are assumed for *Strigamia maritima* and unknown for *Folsomia candida*. Question marks in the amino acid sequences indicate no data availability for the particular species. Amino acids, which differ from the *Mus musculus* homeodomain sequence, are highlighted in colors. Amino acids, which correspond to the *Mus musculus* homeodomain sequence, are represented by grey color. The homeodomain sequences of insect Zen proteins differ from one another more than those of deuterostome (Deut.) species: *Mm* - *Mus musculus*, *Dr* - *Danio rerio*, *Bf* - *Branchiostoma floridae*, *Ci* - *Ciona intestinalis*, *Sk* - *Saccoglossus kowalevskii*. Lophotrochozoa (Loph.): *Chaetopterus variopedatus*, *Av* - *Alita virens*, *Es* - *Euprymna scolopes*, *Ha* - *Haliotis asinina*. Chelicerata: *Al* - *Archezogetes longisetosus*, *Cs* - *Cupiennius salei*. Myriapoda (Myr.): *Gm* - *Glomeris marginata*, *La* - *Lithobius atkinsoni*, *Sm* - *Strigamia maritima*. Crustacea (Cru.): *Cm* - *Carcinus maenas*, *Dp* - *Daphnia pulex*, *Ph* - *Parhyale hawaiiensis*. Hexapoda: *Fc* - *Folsomia candida*, *Td* - *Thermobia domestica*, *Sg* - *Schistocerca gregaria*, *Of* - *Oncopeltus fasciatus*, *Tc* - *Tribolium castaneum*, *Am* - *Apis mellifera*, *Ag* - *Anopheles gambiae*, *Ca* - *Clogmia albipunctata*, *Hp* - *Haematopota pluvialis*, *El* - *Empis livida*, *Ds* - *Drosophila subobscura*, *Ma* - *Megaselia abdita*, *Dm* - *Drosophila melanogaster* (modified from Panfilio and Akam, 2007).

Apart from the changes within the protein sequence of Hox3/Zen, several changes in protein features were described. One feature, common for most of the Hox transcription factors, is the presence of a hexapeptide -YPWM- motif (with four strictly conserved amino acids) upstream of the homeodomain, through which Hox transcription factors bind Extradenticle, a TALE (three amino acid loop extension) family homeodomain cofactor (Passner et al., 1999; Rieden et al., 2004). This hexapeptide motif seems to be lost in Zen proteins. Moreover, the presence or absence of the hexapeptide motif correlates with an embryonic (*Hox3*-like) or an extraembryonic (*zen*-like) expression, respectively (Falciani et al., 1996; Panfilio et al., 2006; Panfilio and Akam, 2007). Further, Zen proteins are noticeably smaller and their homeodomains are positioned closer to the N-terminus, which correlates with the loss of the hexapeptide motif (Panfilio and Akam, 2007).

In addition, changes in the gene structure between *Hox3* and *zen* have been observed as well. It seems that possessing two introns is a common feature for *zen* genes and that *Hox3* genes rather have only one intron (Fig. 1.2) (Panfilio and Akam, 2007).

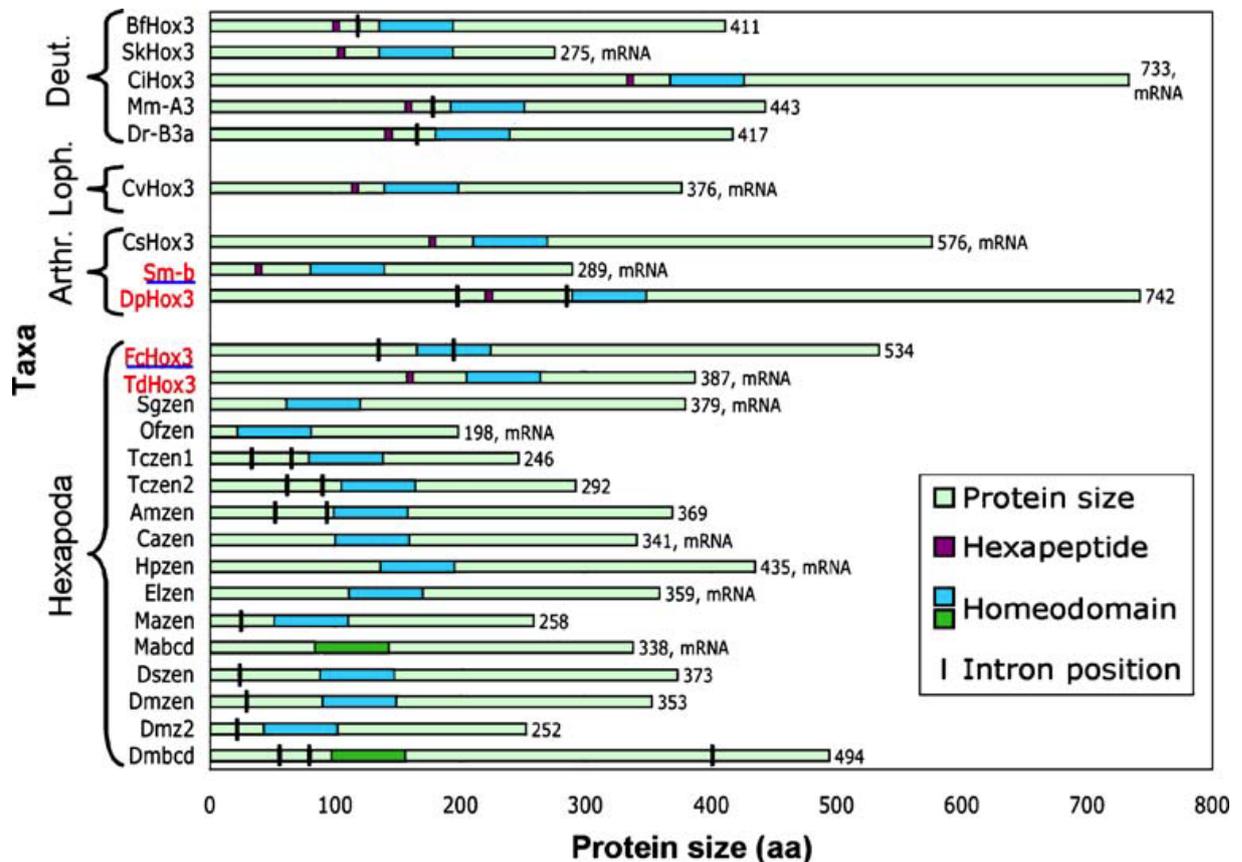


Figure 1.2. Comparison of complete proteins of Hox3/Zen/Bcd orthologues. The size and the position of the protein (light green), hexapeptide (purple), homeodomain (Zen-blue, Bcd-green) and intron position (black bar) are indicated. Zen proteins are noticeably shorter than Hox3 proteins and they lost the hexapeptide motif. Acquisition of extra intron seems to be a feature of Zen proteins. When no genomic DNA data were available, mRNA sequence was considered. For species underlined with blue no expression data are available and their canonical Hox3 function is inferred from the presence of the hexapeptide and/or the position within phylogeny. Taxonomic abbreviations are listed in the figure legend of the Fig. 1.1 (reproduced from Panfilio and Akam, 2007). Arthr.-Arthropoda, Loph.-Lophotrochozoa, Deut.-Deuterostomia.

1.1.2 When did the switch from Hox3 (embryonic) to Zen (extraembryonic) occur?

Based on the information about protein and gene structure changes between Hox3 to Zen available in the species described above, the following progression of the changes during arthropod evolution was proposed by Panfilio and Akam, 2007: (1) the second intron, N-terminal to the hexapeptide motif, was already acquired in crustaceans; (2) loss of the hexapeptide motif, shift of the homeodomain towards N-terminus and overall shortening of the size happened in neopterans (winged insects); (3) one of the introns was lost in dipterans (true flies); and (4) acquisition of various introns and change in protein size during the divergence of *bcd* from *zen* occurred within dipterans (true flies).

In summary, expression data available from species of the arthropod subphyla Chelicerata, Myriapoda and Crustacea suggest that these species express *Hox3* gene in typical *Hox*-like expression pattern in the embryo, whereas in the last arthropod subphylum Insecta, extraembryonic (EE) expression is observed (references cited in Panfilio et al., 2006; Papillon and Telford, 2007). A fingerprint of *Hox3/zen* evolution was left in the basal wingless insect, the firebrat *Thermobia domestica* (Hexapoda). During early embryogenesis, in the firebrat, *Hox3*-like expression is observed in the mouthparts of the embryo, but later on, the expression is apparent in the layer of cells partially covering embryo on the posterior side. This cell layer is in fact mature amnion, one of the EEMs. Thus, *T. domestica* expresses its *Hox3/zen* gene in both embryonic and EE tissue, representing a transition stage from Hox3 to Zen (Hughes et al., 2004). Consistent with the proposed progression of changes during evolution, the change from Hox3 to Zen must have at least partially occurred in insect lineage before the divergence of winged insect species (Fig. 1.3) (Hughes et al., 2004; Panfilio and Akam, 2007).

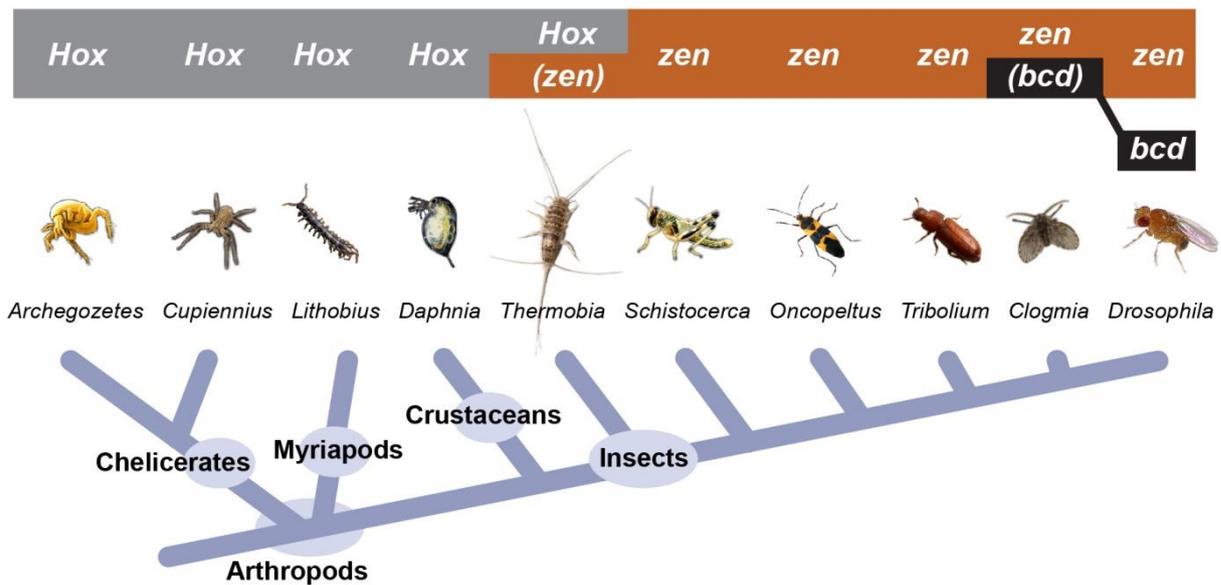
Hox / zen / bcd

Figure 1.3. Changes in the functions of *Hox3/zen/bcd* during arthropod evolution. The figure illustrates the model of the evolution of *Hox3/zen/bcd* based on available expression data from different arthropod species. *Hox3* gene of the mite, the spider (Chelicerata), the centipede (Myriapoda) and *Daphnia* (Crustacea) display canonical *Hox3* gene expression in the embryo during development (Damen and Tautz, 1998; Telford and Thomas, 1998b; Hughes and Kaufman, 2002a; Papillon and Telford, 2007). The basal wingless insect firebrat (*Thermobia*) shows both embryonic and EE expression, which is considered to be a transition stage between canonical *Hox3* and insect *zen* (Hughes et al., 2004). The grasshopper, the bug and the beetle show only EE expression of *zen* gene (Falciani et al., 1996; Dearden et al., 2000; van der Zee et al., 2005; Panfilio et al., 2006). In non-cyclorrhaphan flies, like *Clogmia*, the expression of a single gene has both *zen*- and *bcd*-like character (Stauber et al., 2002). In *Drosophila* the expression of *zen* is in EE tissue, while *bcd* displays again embryonic expression (Rushlow and Levine, 1990; Stauber et al., 1999) (modified from Hughes et al., 2004).

1.2 Extraembryonic membranes: evolution and function

EEMs are present in numerous arthropod eggs. However, during insect evolution, EEMs became more complex and in species within the winged insect lineage, complete EEMs have evolved. Most winged insect species possess two separate EEMs: amnion and serosa. The amnion covers the embryo on the ventral side, forming a yolk-free cavity, which is filled with fluid. The serosa lines the vitelline membrane and in this way covers embryo, amnion and yolk (Fig. 1.4, “most insects” schematic) (references cited in Panfilio, 2008).

Cells of the EEMs do not form parts of the embryonic tissues. Given the facts that we find EEMs in most insect species and that they do not contribute to the embryo itself raises the question regarding the significance of their temporary existence. As the EEMs surround the embryo, the most apparent function is in protecting the embryo from the impacts of the outer environment. Since serosa secretes a chitin based cuticle (Panfilio, 2008), it provides mechanical support of the egg as well as protection against mechanical damage (Panfilio et al., 2013; Farnesi et al., 2015). Moreover, serosal cuticle of the flour beetle (Jacobs et al., 2013) and tropical mosquito (Rezende et al., 2008) has been shown to protect the embryo also

from desiccation. Apart from the protective function of the serosal cuticle, serosa itself provides the embryo with innate immunity and protects the embryo from pathogen infection (Chen et al., 2000; Jacobs et al., 2014).

Determination of the amniotic function is more complicated. The fluid-filled amniotic cavity could evolve to potentially serve as mechanical protection, but the fact that amnion has been reduced during insect evolution leading to the dipteran lineage, to only cover the most dorsal part of the yolk (Fig. 1.4, “reduction of amnion” schematic), suggests that amnion, most likely, has other functions as well.

The formation of EEMs ultimately contributed to the eminent evolutionary success of insects (Grimaldi and Engel, 2005). Due to the protective function of EEMs and the secreted cuticle, insects acquired the ability to oviposit in dry environment and became one of the earliest land animals (Zeh et al., 1989). Thus, the acquisition of EEMs enabled winged insects to lay their eggs in new ecological niches and to colonize land.

However, EEMs did not evolve only to serve the protective function of the embryo. In addition, morphogenetic movements of these simple squamous epithelia are essential for the progression of proper embryonic development. In fact, the precise morphogenetic movements of the EEMs in bug and beetle species have been described in detail (e.g.: Panfilio and Roth, 2010; Panfilio et al., 2013; Hilbrant et al., 2016). In simplicity, the EEMs have first to form, later rupture, contract and evert, and eventually undergo apoptosis in order to accompany and guide the embryo through correct progression of its development. EE development of *T. castaneum* and morphogenetic movements of the EEMs will be described in detail in the section 1.6.1.

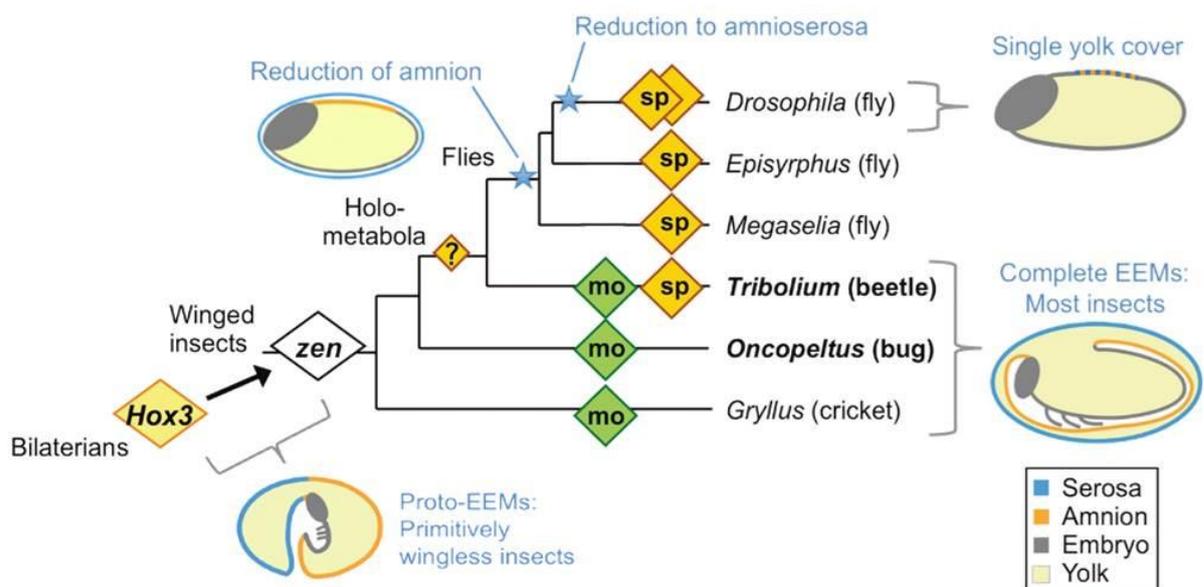


Figure 1.4. Correlation of Zen functions and the anatomical innovation (EEMs). Diamonds are representing *zen* gene and its either early specification function (sp - orange), or late morphogenesis function (mo - green) in the context of phylogenetic positions of hemimetabolous (incomplete metamorphosis) and holometabolous (complete metamorphosis) insect species. Schematic cartoons show evolution of the EEMs; blue text and stars highlight their secondary reduction (adopted and modified from Horn et al., 2015; Panfilio, K.A., Nakamura, T., Mito, T., and Noji, S., unpublished data).

1.2.1 Functions of *zen* in the extraembryonic membranes

The evolution of EEMs is tightly linked to the evolution of *Hox3/zen*. *Zen* has undergone multiple rounds of functional divergence, which resulted in two distinct functions of *Zen* in EEMs: early specification of serosal tissue identity and morphogenesis of matured EEMs during late developmental stages in the process of membrane rupture (Fig. 1.4) (reviewed in Panfilio, 2008; Horn et al., 2015).

The morphogenesis function of *zen* was described for basally branching hemimetabolous species (insects with incomplete metamorphosis) (Fig. 1.4). In cricket (*Gryllus bimaculatus*) (Panfilio, K.A., Nakamura, T., Mito, T., and Noji, S., unpublished data), robust and successful EEM withdrawal is either blocked or partially impaired in the absence of *zen* function. In bug (*Oncopeltus fasciatus*) (Panfilio, 2009), silencing *zen* function through parental RNA interference (pRNAi) causes failure of the rupture of EEMs, whereas establishment of the serosal tissue identity was not affected and no structural defects of the serosa were observed.

As previously mentioned, during the winged insect evolution, *zen* has undergone lineage specific duplications. In the beetle *T. castaneum*, the tandem duplication generated two copies of *zen* gene (*Tc-zen1* and *Tc-zen2*) each fulfilling one of the functions described for *zen*. After pRNAi of *Tc-zen1*, serosal tissue identity is completely lost, and after pRNAi of *Tc-zen2* EEMs either fail to withdraw, or the direction of the withdrawal is altered (van der Zee et al., 2005).

In flies, only the specification function of *zen* was described so far (Fig. 1.4). This includes lower cyclorrhaphan flies (*Megaselia abdida* and *Episyrphus balteatus*), where knockdown of *zen* led to the loss of the serosal tissue (Rafiqi et al., 2008). In *D. melanogaster*, *zen* has been shown to fulfill specification function, when the knockout of this gene led to the loss of the EE tissue and ultimately to a lethal phenotype (Wakimoto et al., 1984). However, in *D. melanogaster*, *zen* has undergone two rounds of duplication. While the first round of duplication generated *zen2* copy, which has been shown not to be essential during embryonic development (Pultz et al., 1988; Rushlow and Levine, 1990), the second round of duplication generated functionally divergent *bicoid* (*bcd*) (Stauber et al., 1999). *Bcd* is maternally localized to the anterior pole of the embryo and the translation of its mRNA results in anterior-posterior concentration gradient essential for the head and thorax development (St Johnston et al., 1989; Rushlow and Levine, 1990; Dearden and Akam, 1999). *Hox3/zen/bcd* evolution represents an interesting case where a single gene has changed its developmental role at least twice during evolution.

Duplications in insects *Hox3* locus are not so rare. The recent genomic sequencing of five lepidopteran species revealed that *zen* has undergone multiple rounds of duplications in Dytrisia clade, generating four additional genes besides *zen* (*special homeobox genes A-D*, *shxA-D*). In one particular case of silkworm *Bombyx mori*, 15 copies of different *shx* genes were discovered in its *Hox3* locus. According to the molecular modeling, these *shx* genes have potential to encode the homeodomain and their expression pattern has been described during early oogenesis in the cells of presumptive serosa before the onset of bona fide *zen*

expression (Ferguson et al., 2014). Although functional experiments were not performed, based on these expression data authors suggest that *shx* genes retained an ancestral association with the specification of EEM, while *zen* function might have diverged again.

1.3 Two possible evolutionary scenarios lying behind diversification of *zen* functions

On one hand, the fact that morphogenesis function was described for basally branching species implies that morphogenesis function represents the ancestral and original role of *zen*. In this case *zen* must have undergone two changes: the first from *Hox3* canonical axial patterning function to morphogenesis function in hemimetabolous insects, and the second change back to specification function in holometabolous species, however with the change of the functional domain from embryonic to EE. Since the specification function in EEMs is taking place in the new functional domain, *zen* has acquired new function and therefore, neo-functionalization hypothesis is considered.

On the other hand, in *T. castaneum* two copies of *zen* gene with two distinct functions were described (van der Zee et al., 2005). The morphogenesis function of *Tc-zen2* is more similar to the *zen* function in hemimetabolous species, while the specification function of *Tc-zen1* is more similar to the *zen* function in dipterans. Therefore, we could assume that both of the functions (specification and morphogenesis) are ancestral and the fact that only the morphogenesis function was described for hemimetabolous species suggests that the specification function was simply lost in basally branching species. In this case, the morphogenesis function would have to be lost in dipteran species, and the two *zen* gene copies, each carrying one of the functions, could represent a case of sub-functionalization (Force et al., 1999). Deciphering, which of the evolutionary scenarios (neo- or sub-functionalization) lies behind the changes of *zen* function is one of the underlying motivations for the project presented in this thesis.

1.4 Functional diversification of paralogues during insect evolution

The case of *zen* is not the only example of a gene undergoing duplication with subsequent functional divergence of the paralogues, which occurred during insect evolution. Several intriguing cases of this process have been described. One of the examples is the case of insect β -Catenin orthologue *armadillo* (*arm*). β -Catenin is a scaffolding protein playing multiple important roles in Wnt signaling, cell adhesion and centrosome separation. In the holometabolous beetle *T. castaneum* and the hemimetabolous pea aphid *Acyrtosiphon pisum*, *arm* has undergone independent lineage specific duplications. For both species two copies of *arm* gene have been described (Bao et al., 2012). Detailed sequence analysis of the *arm* paralogues of both species revealed that the second copy of *arm* gene lost α -Catenin binding domain and exceeded the rate of amino acid substitutions of singleton *arm* homologues. The severe sequence alteration might have triggered genetic split of the functions described for β -Catenin. In fact, RNAi experiments in *T. castaneum* have confirmed that the functions have split between the sister paralogues with *Tc-arm1* functioning in cell adhesion and *Tc-arm2*

functioning in centrosome separation. Interestingly, both paralogues retained the function in Wnt signaling, and the absence of one of the paralogues is phenotypically rescued by the other. Therefore, the case of *arm* paralogues in *T. castaneum* is a remarkable example of partial sub-functionalization and partial redundant conservation of an ancestral function (Bao et al., 2012).

Another interesting case of gene duplication and subsequent partial sub-functionalization is an example of engrailed-family genes. *Engrailed* is thought to be a “selector” gene, which by transcriptional regulation of its downstream targets confers posterior compartment identity on the group of cells derived from the same lineage. *Engrailed* genes have duplicated on numerous occasions during metazoan evolution (Gilbert, 2002). In *D. melanogaster*, the subsequent functional divergence obscured the two paralogues with two separate roles in wing patterning. While the *invected* gene is responsible for the determination of anterior cell fate polarities in the wing, *engrailed* plays a crucial role in determining the posterior cell fates. However, the removal of *engrailed* causes only incomplete morphological transformation from posterior to anterior fate in the wing and the complete transformation can only be achieved by simultaneous elimination of both *engrailed* and *invected*. This observation suggests that although the cell fate polarity determination function was split between the paralogues, *invected* partially retained posterior fate specification function (Coleman et al., 1987; Hidalgo, 1994; Guillen et al., 1995; Simmonds et al., 1995; Gustavson et al., 1996). Therefore, the case of *engrailed/invected* paralogues is yet another example of partial sub-functionalization.

The paralogues *engrailed* and *invected* display a particular genomic organization, which seems to be conserved within holometabolous species: they are positioned next to each other and oriented in “tail-to-tail” position with 3′ end in close proximity and with no interposed transcription units between them (Peel et al., 2006). Duplicated engrailed-family genes have been observed across insect species. The fact that at least two copies of *engrailed* (one of them carrying RS-motif typical of *invected*) were reported in basally branching insect species like cockroach (Marie and Bacon, 2000), bug and firebrat (Peterson et al., 1998), locust and even in springtail (Peel et al., 2006) suggest that the duplication, which gave rise to *engrailed* and *invected* paralogues might have predated the radiation of insects.

1.5 Correlation of extraembryonic membranes evolution with the evolution of *zen*

EEMs became less complex during dipteran evolution. Most insect species possess complete serosa and amnion, however the current state of art suggests that amnion underwent two rounds of reduction during the evolution of cyclorrhaphan flies. The embryos of the holometabolous beetle *T. castaneum* are covered by the amnion on the ventral side, which represents the ancestral state of EEMs (Fig. 1.5, “*Tribolium*” schematics). This amniotic topology was described for the embryos of flies from dipteran suborder Brachycera (e.g.: horse fly and dance fly) (Schmidt-Ott, 2000). However, the ventral amnion formation was suppressed at the stem lineage of cyclorrhaphan flies. In well studied lower cyclorrhaphan fly species *M. abdida* and *E. balteatus*, the amnion covers only the dorsal part of the yolk, while

the serosa expands as it covers the egg and afterwards closes ventrally (Fig. 1.5, “*Megaselia*” schematic) (Rafiqi et al., 2008; Rafiqi et al., 2010). Finally, in higher cyclorrhaphan flies (e.g.: *D. melanogaster* and *Themira biloba*), the serosa underwent severe reduction as well, but in addition fused with the amnion to form a single homogenate tissue covering the dorsal side of the yolk - the amnioserosa (Fig. 1.5, “*Drosophila*” schematic) (Rafiqi et al., 2008). In this section I will present a current opinion about the reduction of amnion and the origin of amnioserosa and the changes in *zen* expression that accompanied it, based on the *zen* expression profiles from *M. abdida*, *E. balteatus* and *D. melanogaster* (Rafiqi et al., 2008; Rafiqi et al., 2010; Schmidt-Ott et al., 2010). Since the expression patterns of *Ma-zen* and *Eb-zen* are the same, for simplicity only comparison between *D. melanogaster* and *M. abdida* will be described.

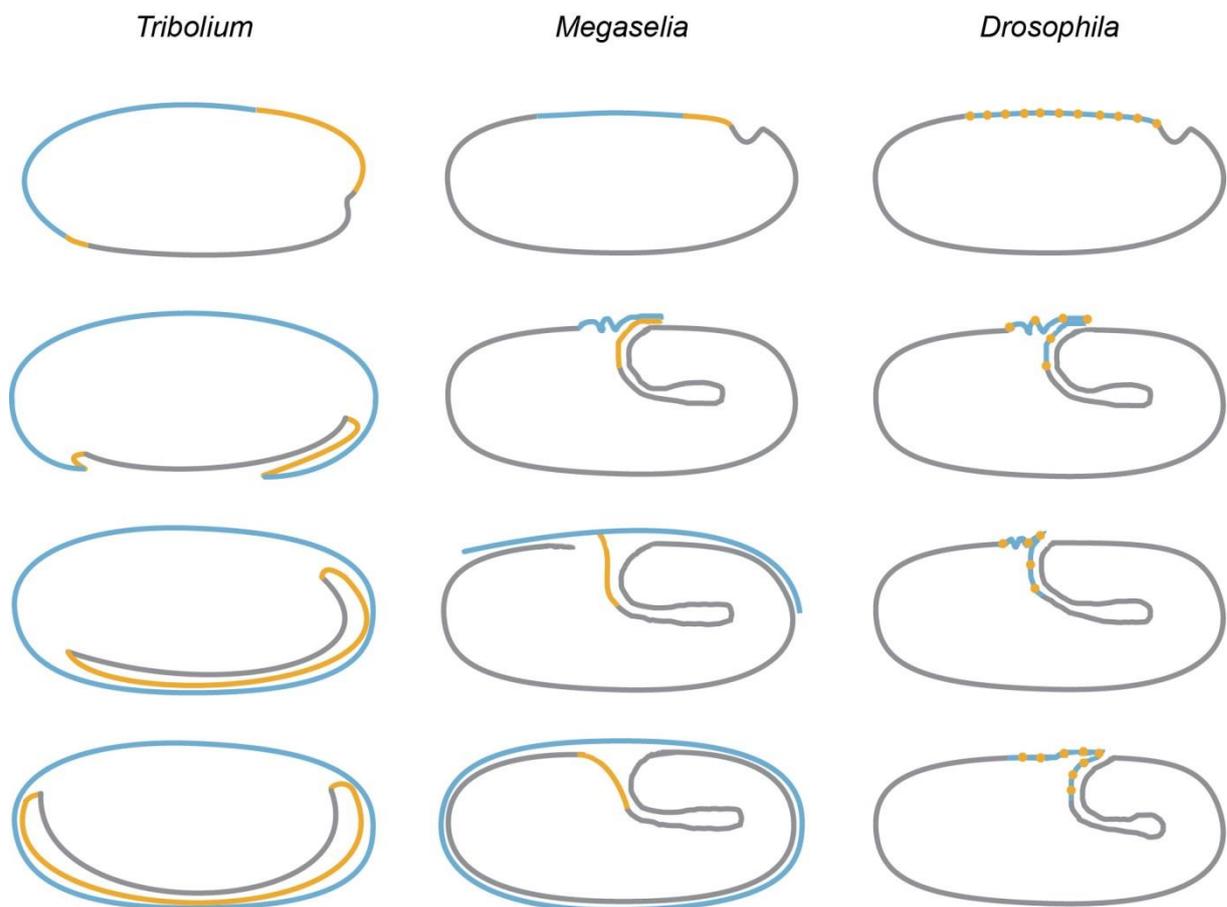


Figure 1.5. Reduction of extraembryonic membranes during dipteran evolution. Schematic overview shows the topology of EEMs in three holometabolous species (beetle *T. castaneum* and flies *M. abdida* and *D. melanogaster*). In *T. castaneum*, the amnion covers the embryo on the ventral side and forms the amniotic cavity, while the serosa covers the embryo, amnion and yolk. In *M. abdida* the amnion is reduced and covers the embryo only dorsally, while the serosa retained its topology and encompasses the egg. In *D. melanogaster*, also the serosa became reduced and moreover fused with the amnion forming a uniform tissue - the amnioserosa. The amnion is depicted in orange, the serosa in blue, the amnioserosa in orange-blue and the embryo in grey (modified from Rafiqi et al., 2008).

Although the EE tissue morphology differs between the compared species, the progression of the embryonic development is similar. *Ma-zen* is not expressed during the blastoderm formation stage before cellularization (Fig. 1.6A). The expression starts during the cellularization when *Ma-zen* is expressed in dorsal blastoderm (Fig. 1.6B). The expression becomes restricted to a dorsal stripe once the serosa forms and eventually, at the beginning of gastrulation, *Ma-zen* is expressed only in the serosa (Fig. 1.6C). On the other hand, the expression of *Dm-zen* starts before the blastoderm cellularization stage (Fig. 1.6F) and the expression in the dorsal part of the embryo is much broader than the expression domain of *Ma-zen*. During the cellularization of the blastoderm, *Dm-zen* expression is confined to the dorsal region, where the amnioserosa anlage, but also the future embryonic region (presumptive head region, eventually *Dm-zen* is expressed in optic lobes) are localized (Fig. 1.6G). While during the gastrulation of the embryo *Dm-zen* is expressed in proctodeum (depression of ectoderm of anal region) (Fig. 1.6I, arrow), *Ma-zen* expression remains only in the serosa (Fig. 1.6D). Finally, the difference between the *Ma-zen* and *Dm-zen* expression is temporal. Unlike in *D. melanogaster*, in *M. abdida*, *zen*'s expression continues also after gastrulation (Fig. 1.6E, J). In *M. abdida*, the amnion and the serosa derive from the amnioserosal fold. The postgastrular expression of *zen* drives the expansion of the serosa and consequently serosa disjoins from the amnion, and covers the whole egg until it closes ventrally. Dorsal amnion stays connected to the embryo.

The differences in *zen* expression between *D. melanogaster* and *M. abdida* suggest that the loss of postgastrular *zen* expression (like described for *D. melanogaster*) led to the reduction of serosa, which ultimately gave rise to the amnioserosa. RNAi experiments in *M. abdida* showed that while silencing *zen* before gastrulation leads to expansion of the amniotic domain and loss of serosal tissue identity, the knockdown of *Ma-zen* during germband retraction leads to the differentiation of serosal tissue, but the disjunction from the amnion does not occur. This result further supported the hypothesis that the postgastrular *zen* expression is necessary for the serosal expansion and that its suppression may

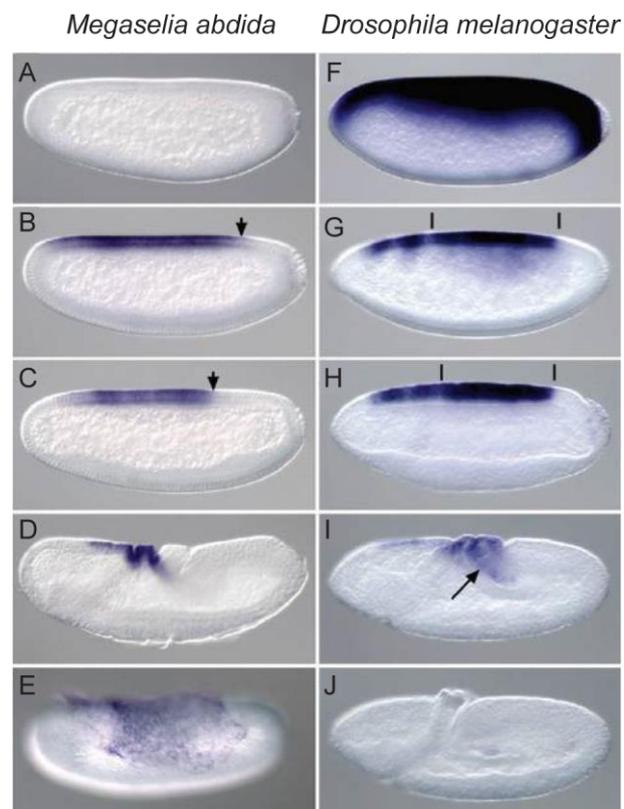


Figure 1.6. Expression of *Ma-zen* and *Dm-zen* in *M. abdida* and *D. melanogaster*. Matching stages of *M. abdida* (left) and *D. melanogaster* (right) are shown: before cellularization (A, F), during cellularization (B, G), the beginning of gastrulation (C, H), during gastrulation (D, I) and after gastrulation (E, J). Horizontal bars (G, H) indicate position of the amnioserosa anlage. Arrowheads (B, C) indicate shortening of the *Ma-zen* domain. Arrow (I) indicates expression of *Dm-zen* in proctodeum. Anterior is left and dorsal up (reproduced from Rafiqi et al., 2008 and Schmidt-Ott et al., 2010).

have triggered the origin of the amnioserosa (Rafiqi et al., 2008; Rafiqi et al., 2010; Schmidt-Ott et al., 2010).

1.6 Insect model organism *Tribolium castaneum*

For the last two decades *T. castaneum* has been emerging as the second insect model organism, leading away from the *D. melanogaster* centric research in insects. During this time, the methodological toolkit was growing and by today there are several sophisticated possibilities how to visualize or alter the embryonic development of *T. castaneum*.

The growth of the methodological toolkit would have not been possible without the sequencing of the genome (Richards et al., 2008). Transgenesis have been successfully established (Berghammer et al., 1999; Berghammer et al., 2009) and recently complemented with the possibility of performing targeted genome editing using CRISPR/Cas9 method (Gilles et al., 2015). Silencing of gene expression by pRNAi (Bucher et al., 2002) became a routinely performed method especially in the field of evo-devo (evolution of development). This method was subsequently exploited for a screen (iBeetle screen) (Schmitt-Engel et al., 2015), which has been performed for high number of *T. castaneum* genes and resulted in the resourceful database (iBeetle-Base) (Donitz et al., 2015). Heat shock-mediated misexpression of genes, as well as GAL4/UAS system, have been demonstrated on a proof-of-principle basis as well (Schinko et al., 2010; Schinko et al., 2012).

A large scale insertional mutagenesis screen fundamentally contributed to the possibilities of embryonic development visualization by generating over 500 enhancer trap lines (Trauner et al., 2009). Several of the enhancer trap lines with the enhanced green fluorescent protein (EGFP) expression in the EEMs have been thoroughly described by our group (Koelzer et al., 2014; Hilbrant et al., 2016). Apart from that, one of the enhancer lines is expressing GFP in all nuclei of developing *T. castaneum* embryo (Sarrazin et al., 2012) and rapidly became an important tool for the live imaging of developmental processes. In addition, visualization of *T. castaneum* embryogenesis is nowadays possible also due to the established transient fluorescent labeling technique (Benton et al., 2013).

Finally, a successful RNA-sequencing (RNA-seq) approach to identify transcriptional regulation in *T. castaneum* has been published recently (Stappert et al., 2016). The described methodological toolkit available for research of *T. castaneum* development along with the ease of laboratory culture handling, serves as evidence that *T. castaneum* is a suitable model organism for investigation of embryonic and EE development.

1.6.1 Extraembryonic development during embryogenesis of *Tribolium castaneum*

T. castaneum was the sole model organism used for the experiments in this project, therefore, in this section, I will introduce its EE development and different developmental stages. In the first hours after egg lay, the undifferentiated blastoderm undergoes twelve synchronized cell proliferation cycles, resulting in the uniform blastoderm (Fig. 1.7A) (Handel et al., 2000). Cell

fates in *T. castaneum* embryo are determined during blastoderm differentiation. The most anterior region acquires serosal tissue identity with the typical morphology of large and widely spaced squamous cells. The cells in the posterior part of the blastoderm form germ rudiment, which gives rise to both amnion and the embryo proper (germband) (Fig. 1.7B, C). During the blastoderm differentiation the most posterior cells flatten and shift into the yolk, forming a primitive pit (Fig. 1.7B, C). As the serosa expands towards the posterior pole, primitive pit is overgrown and shifted ventrally by the posterior amniotic fold (Fig. 1.7D, D') (Handel et al., 2000).

Afterwards, the germ rudiment invaginates dorsal-posteriorly towards anterior and at the same time the amnion extends from posterior to anterior as it starts to cover the embryo on the ventral and lateral sides. In the meantime, the serosa extends from posterior to ventral side and now covers both the amnion and the germ rudiment (Fig. 1.7E, E'). When the serosa border reaches about half of the ventral side of the embryo, it forms a small opening, through which the germband is visible. This stage is called the serosal window stage (Fig. 1.7F, F'). While the serosal window is closing, the circumference of the window is actually formed by amniotic cells. During the final steps of the serosal window closure, the amnion and the serosa separate from each other and for the first time form two discrete tissues (Panfilio, 2008; Hilbrant et al., 2016). After the serosal window closes and serosa encompasses the entire egg, the germband starts to extend posteriorly. Stages from the uniform blastoderm formation until the serosal window closure represent early embryogenesis (Fig. 1.7A-F', red rectangle).

After the extension of germband is fully reached (Fig. 1.7G), germband retraction is initiated. During the retraction process the embryo shortens in anterior-posterior direction and thickens in dorsal-ventral direction (Fig. 1.7H). The complete embryo retraction is reached once the embryo reaches the same length as the anterior-posterior axis. During the whole process of the embryo extension and retraction, the amnion and the serosa retains the topology formed after the serosal window closure (two separate membranes) (Hilbrant et al., 2016).

After the embryo retraction is reached, membrane rupture occurs at the anterior-ventral side, where the membranes are apposed to each other and form a bilayer (Fig. 1.7H, asterisk). While no precise place of the rupture was described for the serosa, the amnion ruptures in the specialized cells of the rupture competence zone (amniotic cap) (Hilbrant et al., 2016). While the amnion initiates the rupture, the serosa drives the withdrawal of both of the membranes by contracting. The amnion and the serosa withdraw simultaneously towards posterior, subsequently they snap over the abdomen and retract towards dorsal side (Fig. 1.7I), where they form dorsal organ. At the same time, the dorsal epidermis of the embryo expands dorsally following the retraction of the membranes (Strobl and Stelzer, 2014).

The dorsal organ undergoes apoptosis and sinks into the yolk while facilitating the dorsal closure of embryo (Panfilio et al., 2013). Once the embryo closes dorsally, the EE development is complete. The embryonic development continues for another approx. 12 h, before the larva hatches. The developmental stages investigated within the presented project did not exceed the post-rupture stage, therefore, pre- and post-rupture stages will be, hereafter, referred to as stages of late embryogenesis (Fig. 1.7H-I, blue rectangle).

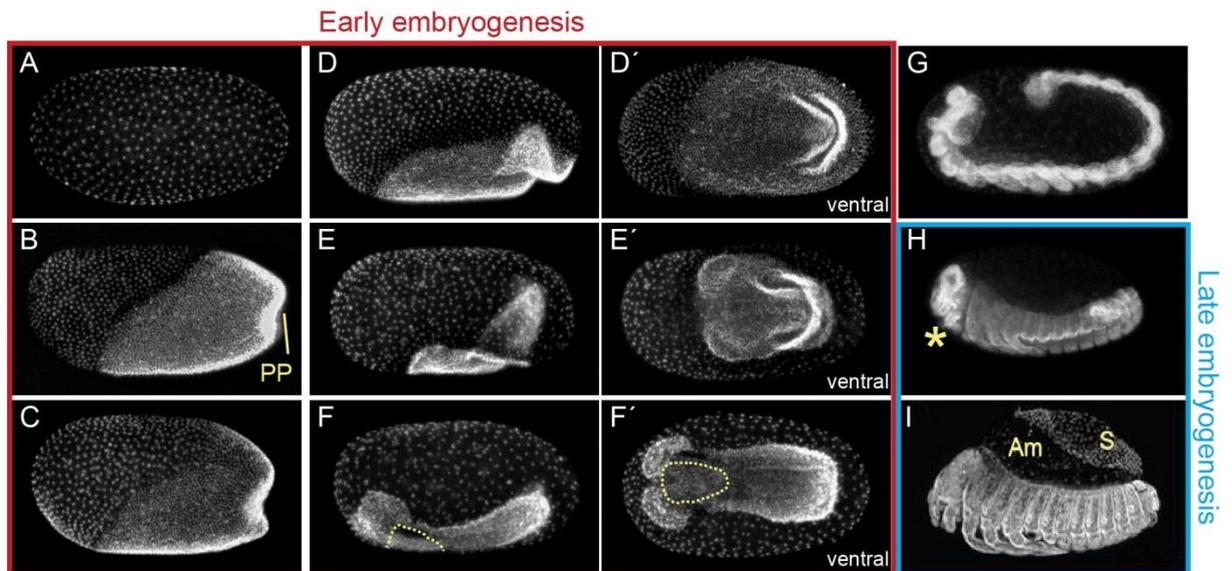


Figure 1.7. Developmental stages of *Tribolium castaneum* during early and late embryogenesis. Nuclear staining of *T. castaneum* embryos of different developmental stages: uniform blastoderm (A), early differentiated blastoderm (B), late differentiated blastoderm (C), posterior amniotic fold (D, D'), early serosal window (E, E'), late serosal window (F, F'), extended germband (G), retracted germband (pre-rupture) (H) and post-rupture stage (I). Note that in the extended and the retracted germband stages (G and H), the germband, amnion and yolk are covered by serosa, however due to the dense nuclear signal from embryo, the serosa and the amnion are not visible. PP-primitive pit, Am-amnion, S-serosa. The rim of the serosal window is highlighted by the dashed line. The position of the membrane rupture is depicted by the asterisk. Unless stated otherwise, the views are lateral with anterior left (with an exception for A, where view cannot be determined). (Micrograph I was reproduced from Koelzer et al., 2014)

1.7 Aims of the study

Insect *Hox3* orthologues, known as *zerknüllt* (*zen*), represent a particular case of Hox genes, which have lost the canonical function in axial patterning of the embryo and rather acquired a novel function in EE domain. Concomitantly with the switch from *Hox3* embryonic to EE function of *zen*, complex EEMs arose during insect evolution. Two distinct roles of *zen* genes in EEMs have been described: early specification function and late morphogenesis function. In the holometabolous beetle *T. castaneum*, *zen* has undergone lineage specific tandem duplication, which generated two fully functional copies (*Tc-zen1* and *Tc-zen2*) each carrying one of the functions: *Tc-zen1* specifies serosal tissue identity during early embryogenesis and *Tc-zen2* is responsible for morphogenesis of matured EEMs during late embryogenesis.

The ultimate aim of this project is to decipher how the two *T. castaneum* paralogues acquired two distinct functions. Our first approach is to pinpoint the differences between transcriptional and translational regulations of *Tc-zen* genes by describing in detail their transcript and protein expression profiles during early and late embryogenesis. With the second approach, we identify downstream transcriptional targets of both *Tc-zen* genes during early embryogenesis by performing RNA-seq after pRNAi followed by differential expression (DE) analysis. The subsequent comparative analysis of identified targets should provide insight into whether the neo- or sub-functionalization hypothesis applies to the case of *T. castaneum* paralogues. Finally, with the third approach, we identify transcriptional target genes of *Tc-zen2* during important developmental events of late embryogenesis: before and after the rupture of the EEMs, when *Tc-zen2* function takes place. The candidate targets are identified by RNA-seq after pRNAi experiment and the subsequent DE analysis. The functional profile of identified *Tc-zen2* targets is retrieved by gene ontology (GO) term analysis. Ultimately the results from all the three approaches should elucidate the specific changes that occurred on the transcriptional and translational level, and which might have triggered the functional divergence of *T. castaneum* paralogues.

2 MATERIAL AND METHODS

2.1 *Tribolium castaneum* stock maintenance

2.1.1 *Tribolium castaneum* husbandry

For all experiments *Tribolium castaneum* San Bernardino (SB) strain (Brown et al., 2009) was used as the wild type (WT) reference. Beetles were kept in the dark at 30 °C with the relative humidity between 40-60%. Cohorts of beetles were maintained in plastic boxes with mesh windows for aeration. Plastic boxes were half-filled with the flour mixture consisting of wheat flour “Extra Type 405” (Diamant) and dark wheat flour “Type 1050” (Diamant) (in 2:1 ratio) supplemented with 0.33 g/kg Fumagilin-B (Medivet Pharmaceuticals Ltd.) for protection against pathogens. The flour mixture was enriched in nutrition by adding 18.75 g of yeast per 1 kg of the flour mixture (collectively hereafter referred to as stock flour). For general stock keeping, approximately 22 g of beetles were kept on 900 g of the stock flour. Since females lay the highest amount of eggs within the first months of their life, a new population of beetles was established every third month. Smaller populations of beetles (eg.: after RNAi experiment) were kept in plastic vials half-filled with the stock flour and closed with a foam lid.

2.1.2 Egg collection

In order to collect the eggs, the beetles were first set on pre-sieved egg lay flour “Instant Type 405” (Diamant). After desired collection interval, the adults were separated from the flour containing eggs with a test sieve with 710 µm mesh size (Retsch). Afterwards, the eggs were separated from the egg lay flour with a test sieve with 300 µm mesh size (Retsch). The eggs were either directly processed further, or incubated for a defined time (incubation period) in order to reach the developmental stage of interest. During the incubation period the eggs were stored in a collection basket (mesh size \leq 180 µm) on the egg lay flour at 30 °C with relative humidity between 40-60%. The reached developmental stage of the eggs was calculated in hours after egg lay (h AEL). The minimal age of the eggs corresponds to the incubation period and the maximum age corresponds to the incubation period plus the collection interval.

2.1.3 Egg dechoriation

To clean the collected eggs from remaining flour and yeast, the eggs were first rinsed with tap water. Next, to remove the chorion, the collection basket with eggs was placed in a petri dish fully filled with bleach (“DanKlorix Hygienereiniger”, Colgate-Palmolive) containing 4-5% sodium carbonate and 1-4% sodium hypochlorite, to digest the chorion. The eggs were swirled in the bleach for 5 min. Finally, to remove the bleach, the eggs were rinsed again with tap water and placed in a clean petri dish fully filled with tap water. Dechorionated eggs were collected from the water surface with a brush and the excess water was removed by placing the brush on Whatman gel blot paper (Whatman International Ltd.).

2.1.4 Fast freeze of eggs

In order to minimize the delay in the developmental stage caused during dechoriation procedure, the eggs were rapidly frozen. Dechorionated dried eggs were transferred on the brush to an Eppendorf tube, which was immediately placed in liquid nitrogen. The amount of eggs in one Eppendorf tube corresponded to the volume of 25-50 μl of water. The eggs were stored until further processing (mRNA and protein extraction) at $-80\text{ }^{\circ}\text{C}$.

2.1.5 Fixation and devitellination

Due to the fact that serosa secretes the serosal cuticle (Panfilio, 2008) devitellinization by methanol shock after fixation is only possible in eggs younger than 16 h AEL. The serosal cuticle sticks to vitelline membrane preventing it from bursting. The eggs possessing the vitelline membrane cannot be used for antibody (AB) or *in situ* hybridization (ISH) staining in whole mount form, because the membrane is not permeable for any AB or ISH probe.

Dechorionated *T. castaneum* eggs were transferred to a glass vial containing fixation solution consisting of 2 ml phosphate-buffered saline (PBS), 2 ml of 10% methanol-free formaldehyde and 4 ml of heptane. The eggs were fixed for 20 min (AB staining) or for 1.5 h (ISH) on the rocker. Next, the lower aqueous phase (PBS + formaldehyde) was removed and replaced with approx. 3 ml of ice cold 100% methanol. Subsequently, the vial was thoroughly shaken for 20 s. The methanol shock caused burst of vitelline membrane and allowed the eggs to sink from the interface between heptane and methanol to the bottom of the vial. The devitellinated eggs were transferred from the vial to a new Eppendorf tube and were subsequently washed 3-5 times with 100% methanol. The eggs were stored in 100% methanol at $-20\text{ }^{\circ}\text{C}$ until further processing.

2.2 Basic molecular methods

2.2.1 RNA extraction

RNA was isolated according to Stappert et al., 2016 and the TRIzol Reagent protocol (Ambion, Life Technologies). Eggs stored at $-80\text{ }^{\circ}\text{C}$ were thawed on ice for approx. 5-10 min. Eggs were first homogenized in 100 μl of the TRIzol. The homogenate was centrifuged at 12000 g and $4\text{ }^{\circ}\text{C}$ for 10 min. The supernatant was transferred to a new Eppendorf tube and 400 μl of TRIzol were added. Samples were incubated for 5 min at room temperature (RT). Next, 100 μl of chloroform were added and the tubes were shaken thoroughly by hand for 15 s. The samples were incubated at RT for 2-3 min, followed by centrifugation at 12000 g and $4\text{ }^{\circ}\text{C}$ for 15 min. The aqueous phase containing RNA was transferred to a new Eppendorf tube. For RNA precipitation, 250 μl of isopropanol were added and the samples were incubated for 10 min at RT. Afterwards, the samples were centrifuged at 12000 g and $4\text{ }^{\circ}\text{C}$ for 10 min. The supernatant was removed and the pellet was washed twice with 500 μl of 70% ethanol. The washing was followed by short centrifugation at 12000 g and $4\text{ }^{\circ}\text{C}$ for 5 min. After the second washing step, the ethanol was carefully removed and the pellet was air dried in the open

Eppendorf tubes at RT for 10 min. The pellet was resuspended in nuclease-free water (Ambion, Life Technologies) (10-20 µl depending on the further use) in heating block at 900 rpm and 60 °C for 10 min. The concentration of RNA was measured on a spectrophotometer (NanoDrop 2000c, Thermo Scientific). The isolated RNA was stored at -80 °C until further use.

2.2.2 cDNA synthesis

Complementary DNA (cDNA) was synthesized by reverse transcription of RNA using the SuperScript VILO cDNA synthesis kit (Invitrogen, Life Technologies) according to the manufacturer's protocol. The input amount of RNA to the reaction was 2 µg. Synthesized cDNA was stored at -80 °C until further use.

2.2.3 Primer design

The online interface Primer3 version 4.0.0 (Untergasser et al., 2012) was used for primer design. The full or partial sequence of the gene of interest was pasted to the interface and, depending on the intended use, the size of the final PCR product was set to 600-800 bp (ISH probe and dsRNA) or to 100-150 bp (RT-qPCR). The primer pair with the best thermodynamic parameters scores with respect to GC content (approx. 50%), self and partner complementarity (low), melting temperature (approx. 60 °C) and primer size (20-22 bp) was chosen. To avoid any unspecific amplification, the primers were blasted against the *T. castaneum* genome. Only primer pairs without any sequence identity to regions other than the gene of interest, and/or with sequence identity only to intergenic regions, were used. To further enable ISH probe and dsRNA synthesis, the primers were equipped with linker sequence at the 5' end: GGCCGCGG for forward primer and CCCGGGGC for reverse primer. Primers were synthesized by Sigma-Aldrich. All primer sequences used in this project are listed in Table 2.1.

Table 2.1. List of all primer sequences used for different purposes of this project.

TC gene identifier and primer orientation	sequence	amplicon size [bp]
ISH probes		
<i>TC000921 (Tc-zen1) / F</i>	ggccgcggTCCCAATTTGAAAACCAAGC	688
<i>TC000921 (Tc-zen1) / R</i>	cccggggcCGTTCCACCCTTCCTGATAA	
<i>TC000922 (Tc-zen2) / F</i>	ggccgcggAACGCCCCAGTTTCAACAA	546
<i>TC000922 (Tc-zen2) / R</i>	cccggggcCTCATCCTTCACCACCACCT	
dsRNA		
<i>TC000921 (Tc-zen1) / F</i>	ggccgcggTTTGAAAACCAAGCCGTTCT	203
<i>TC000921 (Tc-zen1) / R</i>	cccggggcCGTTGGGGTTGAGTTTCTTG	(short fragment)
<i>TC000921 (Tc-zen1) / F</i>	ggccgcggTTTGAAAACCAAGCCGTTCT	682
<i>TC000921 (Tc-zen1) / R</i>	cccggggcCGTTCCACCCTTCCTGATAA	(long fragment)

<i>TC000922 (Tc-zen2) / F</i>	ggccgcggCAATGTCGCCGCAATCGACG	250
<i>TC000922 (Tc-zen2) / R</i>	cccggggcACACAATTCTTCCCTTGTA	
RT-qPCR		
<i>TC000921 (Tc-zen1) / F</i>	TCCACCTTCTGATTGGAAGT	101
<i>TC000921 (Tc-zen1) / R</i>	CGTTGGGGTTGAGTTTCTTG	
<i>TC000922 (Tc-zen2) / F</i>	TCGAAGTGTCCCTCTCAGAAA	101
<i>TC000922 (Tc-zen2) / R</i>	GGAGGAGGTGTACGCAGTTC	
<i>TC008261 (Tc-RpS3) / F</i>	ACCGTCGTATTCGTGAATTGAC	186
<i>TC008261 (Tc-RpS3) / R</i>	ACCTCAAACACCATAGCAAGC	
<i>Tc-zen1</i> candidate target genes (miniscreen#1) both ISH probes and dsRNA		
<i>TC000107 / F</i>	ggccgcggCTTACACCATGGGCGAGATT	555
<i>TC000107 / R</i>	cccggggcCAGCAGCGTCAAACATGACT	
<i>TC007258 / F</i>	ggccgcggGGAAGTCTTTTCGGACAACA	552
<i>TC007258 / R</i>	cccggggcGACCTCAGCAGCGTAACTCC	
<i>TC006727 / F</i>	ggccgcggCAGTTGAAGACGCGAATGAA	548
<i>TC006727 / R</i>	cccggggcAGGTTTAGGTGCCTCGGTTT	
<i>TC015108 / F</i>	ggccgcggCCAAATTGTGTGGCGTAATG	580
<i>TC015108 / R</i>	cccggggcTGTGGAATGCAGGGTAATGA	
<i>TC013480 / F</i>	ggccgcggAGGCTGGCCTTATTCCATTT	521
<i>TC013480 / R</i>	cccggggcCAGGACCACTTCCCTCCGTTA	
<i>TC015555 / F</i>	ggccgcggGCACAAACTGAACGGGTTTT	503
<i>TC015555 / R</i>	cccggggcAAAAATCCTCAATGCGAGGTC	
<i>TC008400 / F</i>	ggccgcggGCAGTTTTGCTCGTTTTGGT	313
<i>TC008400 / R</i>	cccggggcGCAAAGCGTATTGCTCACA	
<i>TC011141 / F</i>	ggccgcggGTTCCAAAGGCGAATACGAA	528
<i>TC011141 / R</i>	cccggggcTCGGATCATCACAGGTGAAA	
<i>TC031198 / F</i>	ggccgcggCGGTTACTTGTGGCCTTGTT	582
<i>TC031198 / R</i>	cccggggcGAGGAACGCTCTTCTTGCAC	
<i>TC011283 / F</i>	ggccgcggCAGGACCGGACTTTATTGGA	761
<i>TC011283 / R</i>	cccggggcAAAAGCACCCGAATTTTGTG	
<i>TC013404 / F</i>	ggccgcggTTTTGCAACGATTCTGTGCT	745
<i>TC013404 / R</i>	cccggggcCAAAGATCAGTCGGCATT	
<i>TC014502 / F</i>	ggccgcggTGTGATACTTGCCGTTGCTC	760
<i>TC014502 / R</i>	cccggggcTCTGTTATTTTCCGGTGCTG	
<i>TC013320 / F</i>	ggccgcggCTGATTAAGCGGGCAATAA	772
<i>TC013320 / R</i>	cccggggcAATCGGAAAACACCATCTCG	
<i>TC016348 / F</i>	ggccgcggGATGATGGAACCACCAAACC	457
<i>TC016348 / R</i>	cccggggcCAGGACACATCTGTGCACT	
<i>TC034701 / F</i>	ggccgcggGAGGAATTACTCCCGGCTTC	584
<i>TC034701 / R</i>	cccggggcTCAGATTCATCTGCACTCG	
<i>TC012744 / F</i>	ggccgcggCGTTTTCCATCGTTTTCGTT	540
<i>TC012744 / R</i>	cccggggcGGCGGAATTATCCCAAACCT	
<i>Tc-zen2</i> candidate target genes (miniscreen#1) both ISH probes and dsRNA		
<i>TC007326 / F</i>	ggccgcggCCTGATGGCAAGTGCTACAA	615
<i>TC007326 / R</i>	cccggggcCGGGTGCAGTTGGTAGTTTT	
<i>TC011068 / F</i>	ggccgcggACCAAACAAGACCCTCAACG	606
<i>TC011068 / R</i>	cccggggcGAGTCTTGGTGGTTCCGGTGT	
<i>TC000511 / F</i>	ggccgcggCTCACCGAAGCAACAGATCA	640
<i>TC000511 / R</i>	cccggggcTGACTTCAGACGTGGACGAG	

<i>TC000853</i> / F	ggccgcggTTTATCTTCGCCACGTCAT	359
<i>TC000853</i> / R	cccggggcGCGGGTCTCGGAATAACC	
<i>TC002837</i> / F	ggccgcggAGTTTTTCACCCACACAG	651
<i>TC002837</i> / R	cccggggcATCATGGGCGGTGTATTCAT	
<i>TC011635</i> / F	ggccgcggTATTTGCAACACCGAACAA	653
<i>TC011635</i> / R	cccggggcGCTGAACGTGAATTGGAGGT	
<i>TC008204-RB</i> / F	ggccgcggCACCGAGAAGTTCGAAAAGC	799
<i>TC008204-RB</i> / R	cccggggcAAGCACAAACACCAAGCAAA	
<i>TC033464-RA</i> / F	ggccgcggCAGAAACCAAAGCAATTTCCA	689
<i>TC033464-RA</i> / R	cccggggcGCAAAAATCTGTCCGAAAGC	
<i>TC011724-RA</i> / F	ggccgcggCCAAGAACCTGGACCAAGAA	802
<i>TC011724-RA</i> / R	cccggggcGAGTTGATTTTCGACGGTGGT	
<i>TC000446-RA</i> / F	ggccgcggACCAACGGGCTAAAAAGCTC	789
<i>TC000446-RA</i> / R	ggccgcggACCAACGGGCTAAAAAGCTC	
<i>TC015615</i> / F	ggccgcggCCGAAACTACCCACTTGAA	629
<i>TC015615</i> / R	cccggggcTTGCCATGAAATCCCAAAT	
<i>TC002092</i> / F	ggccgcggAGTGGTGTTTTTGCCCTTTG	632
<i>TC002092</i> / R	cccggggcACACCAGTAACGCAACCACA	
<i>TC002831</i> / F	ggccgcggGCCTTCATGAAATACATCTTCG	203
<i>TC002831</i> / R	cccggggcTGTGGCTCTTCATGGACGTA	
<i>TC008236</i> / F	ggccgcggCAACAGCCCAAGTATGTCCA	626
<i>TC008236</i> / R	cccggggcCTGATTTGCGTCGGTCTGTA	
<i>TC014361</i> / F	ggccgcggGGCCCTGTATACCTCCCAAT	658
<i>TC014361</i> / R	cccggggcCCGCCAGAAAGAACAGAAAG	
<i>TC001119</i> / F	ggccgcggAATACGGCTACGTGGACGAC	688
<i>TC001119</i> / R	cccggggcCCACCTTGGGGTAGAGCATA	
<i>TC013146</i> / F	ggccgcggCTGCAAGGGCTTCTTCAAAC	630
<i>TC013146</i> / R	cccggggcAAGCTGCTGCAGAACTCCAT	
<i>TC000546</i> / F	ggccgcggATCCGCGAAAGGACACATAG	628
<i>TC000546</i> / R	cccggggcTTGGACGACATTCGGTAACA	
<i>TC002942</i> / F	ggccgcggAAAACAATCACCGCCAAGTC	638
<i>TC002942</i> / R	cccggggcGCCCAAACATGTACCAAAC	
<i>Tc-zen2</i> candidate target genes (miniscreen#2) RT-qPCR		
<i>TC010840-RA</i> / F	TGCGCCTCTCTTCAGTACCT	127
<i>TC010840-RA</i> / R	CGCCAGGTAAAGGCATACAC	
<i>TC011665-RA</i> / F	AAGACGCAGCTTTGACCAAT	142
<i>TC011665-RA</i> / R	ATCATCATACCGCCCATCAT	
<i>TC005982-RA</i> / F	GTATTGTGTACGCGGGGACT	117
<i>TC005982-RA</i> / R	TTGTTGAAAACCCACCCTCT	
<i>TC006575-RA</i> / F	TTACCATTTGTCCCGAGTCC	144
<i>TC006575-RA</i> / R	CGAACTTCGTGTCCGAAATA	
<i>TC014041-RA</i> / F	TATGGCAGCCACAAGAAGC	150
<i>TC014041-RA</i> / R	GTTGGGGTGGTGTCTAGAT	
<i>TC014143-RA</i> / F	CGAAGACGATAAAGAGGGCTA	123
<i>TC014143-RA</i> / R	TTCATGGCACTATACTGGTTCG	
<i>TC014497-RA</i> / F	GCTCTTCGTTTCACTTGTGG	137
<i>TC014497-RA</i> / R	TGCCGTCACTGGTCTCATAAC	
<i>TC007162-RA</i> / F	CGTCGCAACCTGTAAGTCTG	148
<i>TC007162-RA</i> / R	TCGTTTCATCAGCGTGAAGTC	
<i>TC008606-RA</i> / F	CAAGCTGGCCTCGTCACTAT	134
<i>TC008606-RA</i> / R	ATGCAGTCGCACATCATT	

<i>TC011349-RA / F</i>	ACCATCGTTACCCTCATTGC	122
<i>TC011349-RA / R</i>	TCCCTGCATTTCGATATAGCC	
<i>TC031481-RA / F</i>	GCACCCCGATAACGGATTA	114
<i>TC031481-RA / R</i>	GGGATTTACCATTACTGGA	
<i>TC033856-RA / F</i>	GAAGAGGCCGAAACTACGA	150
<i>TC033856-RA / R</i>	GCCCCTTACCACCGACTAT	
Universal primers for T7 polymerase		
5' universal	GAGAATTCTAATACGACTCACTATAGggccgcgg	NA
3' universal	AGGGATCCTAATACGACTCACTATAGGGcccggggc	

2.2.4 Polymerase chain reaction

Polymerase chain reaction (PCR) was used for amplification of specific fragments of the genes of interest. Two consecutive PCR reactions were performed in order to amplify gene fragment for either ISH probe or dsRNA synthesis. For both, the reaction setup and the thermal cycling conditions of the first and the second PCR reaction were the same:

PCR mix (20 µl)

1 µl	cDNA (100 ng)
1 µl	forward primer (10 µM)
1 µl	reverse primer (10 µM)
10 µl	REDTaq ReadyMix (Sigma-Aldrich)
7 µl	nuclease-free H ₂ O

Thermal cycling conditions

1)	94 °C	5 min
2)	94 °C	30 s
3)	57 °C	30 s
4)	72 °C	1 min
5)	34 times to step 2	
6)	72 °C	10 min

In the second PCR reaction mix, 1 µl of the first PCR reaction was always used as a template. Therefore, the amplified fragment always contains at the 5' and 3' ends the linker sequences for the 5' and 3' universal primers, which contain promoter sequence for T7 polymerase. The choice of the universal primers depends on whether the amplified fragment will be further used for ISH probe or dsRNA synthesis. If the antisense ISH probe was synthesized, forward specific primer and the 3' universal primer were used in the second PCR reaction. If the control sense ISH probe was synthesized, the 5' universal primer and specific reverse primer were used in the second PCR reaction. If the dsRNA was synthesized, both 5' and 3' universal primers were used in the second PCR reaction.

To confirm the presence of correct PCR product, 5 µl of the reaction were analyzed on a 1% agarose gel. The gel was run in TRIS-Acetate-EDTA-Buffer at 135 V. The size of the PCR product was determined by comparison with Smart Ladder MW 1700-10 (Eurogentec).

2.2.5 TA cloning

To minimize the background of the *Tc-zen1* and *Tc-zen2* ISH staining, the probes were cloned into a vector using TA cloning system (Invitrogen). The same cloning system was also used in the case when the second PCR reaction did not yield PCR product of a desired size or yielded more than one PCR product. The unpurified fresh PCR product of the first PCR reaction (5 μ l) was mixed together with 2 μ l of pCRII vector, 1 μ l of ExpressLinkT4 DNA ligase (5 units) and 2 μ l of ExpressLinkT4 DNA ligase buffer. The ligation reaction was incubated for 15 min at RT and subsequently used for One Shot chemical transformation: 2 μ l of the ligation reaction were gently mixed with One Shot TOP10 chemically competent *Escherichia coli* cells (Invitrogen) and incubated on ice for 30 min. The heat shock was performed at 42 °C for 30 s and was followed by cold shock by placing the reaction tube back on ice. The cells were recovered by incubation in 200 μ l of Super Optimal Catabolite medium (SOC, Invitrogen) at 37 °C for 1 h on the rotor. The cell suspension was plated on the lysogeny broth (LB) medium plates supplemented with 0.01% ampicillin (Sigma) and 0.004% x-Gal (Sigma), and incubated overnight (O/N) at 37 °C.

To confirm the successful insertion of PCR fragment, 8 colonies were picked from the plate using a pipette tip and dipped into 8 μ l of nuclease-free water. The tip was first left in the water for 10 min to allow for cells detachment and afterwards stored for the later use. The suspension was subsequently used as a template for “colony” PCR. The size of the PCR product was analyzed by gel electrophoresis. If the desired PCR product was obtained, the tip with the remaining cells containing the vector with the insert was used for inoculation of 2 ml LB medium supplemented with 0.01% ampicillin. The cells were incubated O/N at 37 °C on the rotor.

The cell culture was harvested and the vector was isolated using ZR Plasmid Miniprep kit (Zymo research) according to the manufacturer’s protocol. Finally the isolated vector with the cloned PCR fragment of the gene of interest was used as a template for the first PCR described in the section 2.2.4.

2.3 Quantitative reverse transcription PCR

Quantitative reverse transcription PCR (RT-qPCR) was employed to quantify the presence of transcript of the gene of interest in various developmental stages and to evaluate the strength of knockdown (KD) in the RNA-seq after RNAi experiments. Libraries of cDNA (preparation described in the section 2.2.2) were used as template for the reaction. The expression of the gene of interest was normalized to the ribosomal protein S3 (*Tc-RpS3*). Two different master mixes were used in this project: SYBR Green PCR Master Mix (Applied Biosystems, Life Technologies) and GoTaq® qPCR Master Mix (Promega). For all RT-qPCR runs, 7500 Fast Real-time PCR cycler (Applied Biosystems) was used. Note that in order to investigate expression levels of *Tc-zen2* it is necessary to use SYBR Green master mix, which usage, particularly in the case of *Tc-zen2*, has been proven to obtain more consistent results.

RT-qPCR mix (20 µl)

5 µl	cDNA (20 ng)
0.8 µl	forward primer (10 µM)
0.8 µl	reverse primer (10 µM)
10 µl	Master Mix (SYBR Green/GoTaq)
3.4 µl	nuclease-free H ₂ O

Thermal cycling conditions

- 1) 95 °C 2 min
- 2) 95 °C 3 s
- 3) 60 °C 30 s
- 4) 39 times to step 2
- 5) 95 °C 15 s
- 6) 60 °C - 95 °C
(1% temperature increase after each measurement - melting curve reading)
- 7) 60 °C 15 s

The RT-qPCR experiments were performed in three (KD strength evaluation) or four (expression profile of *Tc-zen1* and *Tc-zen2*) biological replicates (BRs), each of three technical replicates (three same reactions within one plate). Note that technical replicate in the result section in the tables 4.1 and 4.3 refers to several different cDNA samples prepared from eggs collected after different number of days after injection (DAI) within one BR. The number of cDNA samples in each of three BRs was different, because the mRNA was not successfully extracted from every egg collection. The KD strength was then measured for each cDNA sample (technical replicate) individually. Therefore, in all three BRs in the above mentioned tables, the KD strength is represented as a range.

2.3.1 RT-qPCR data analysis

Raw RT-qPCR data were exported from 7500 software version 2.3 (Applied Biosystems). The baseline correlation and Cp values determination was performed using LinRegPCR version 12.16 (Ruijter et al., 2009; Tuomi et al., 2010). The final expression ratio (R) was calculated in Excel 2010 (Microsoft) according to the following formula:

$$R = \frac{(E_{target})^{\Delta C_{p_{target}}(C_{p_{ref.s.}} - C_{p_{target}})}}{(E_{ref.g.})^{\Delta C_{p_{ref.g.}}(C_{p_{ref.s.}} - C_{p_{ref.g.}})}}$$

where E_{target} and $E_{ref.g.}$ refer to the mean PCR efficiency of the target gene of interest and of the reference gene (*Tc-RpS3*), and $\Delta C_{p_{target}}$ and $\Delta C_{p_{ref.g.}}$ refer to the difference between mean Cp value of the target gene of interest and of the reference gene. The ΔC_p is calculated as subtraction of the Cp value of the reference gene (*ref.g.*) from the Cp value of reference sample (*ref.s.*). The reference sample is a pool of all cDNA samples used for the entire experiment (in all BRs). The Cp value represents the cycle number, at which the amplicon reaches the threshold of fluorescence (the lower the Cp value, the higher the expression of the gene of interest).

2.4 *In situ* hybridization

2.4.1 Probe synthesis

Antisense digoxigenin (dig) labeled single stranded RNA (ssRNA) ISH probes were synthesized in the transcription reaction containing diluted second PCR reaction outcome (1:1 dilution with nuclease-free water), performed with 5' specific and the 3' universal primers according to the section 2.2.4, as a template. The reaction mix contained 6 µl of template, 2 µl of dig-coupled uridine triphosphate labelling mix (Roche), 2 µl of T7 RNA-polymerase (40 U) (Roche), 2 µl of transcription buffer (Roche), 1 µl of RNase inhibitor (Roche) and 7 µl of nuclease-free water. The reaction was incubated at 37 °C for 4 h. The transcription reaction was terminated by adding 30 µl of nuclease-free water and 50 µl of stop solution containing 0.2 M sodium acetate (pH 6.0). Afterwards, 5 µl of tRNA and 10 µl of lithium chloride were added and the RNA was precipitated in 300 µl of 100% ethanol at -20 °C for 30 min. Next, samples were centrifuged at 20000 g and 4 °C for 15 min. The supernatant was discarded and the RNA pellet was washed twice with 70% ethanol. After the second wash, the samples were shortly centrifuged at 12000 g and 4 °C for 5 min. Finally, the ethanol was removed and the pellet air dried at RT in the open Eppendorf tube for 10 min. The pellet was resuspended in 100 µl of probe solution containing 50% formamide in 2x saline-sodium citrate (SSC). Probes were stored at -20 °C until further use.

2.4.2 Probe hybridization

ISH was performed using both WT and KD embryos. Dechorionated and fixed eggs were first rehydrated by gradual transfer from 100% methanol to 0.1% PBT (0.1% Tween-20 in PBS). Eggs were first brought to 30% PBT in methanol, then 70% PBT in methanol and finally to pure PBT. To ensure no residual methanol was present, the eggs were washed 3 times with PBT for 5 min. After the last wash, PBT was removed and the eggs were incubated with 1:1 solution of PBT: hybridization solution I (Hyb I, 50% formamide, 25% SSC and 0.1% Heparin in Milli-Q water) for 10 min at RT on the rotating wheel. Next, the eggs were incubated in pure Hyb I for 10 min at RT on the rotating wheel. The Hyb I was replaced with Hyb II (Hyb I containing 0.1% salmon sperm) and the eggs were incubated at 55 °C for 1 h in the heating block. After the pre-hybridization, the Hyb II was removed and the eggs were covered with 100 µl of Hyb II containing 2 µl of dig-labeled ssRNA probe. The samples were incubated at 55 °C O/N on rotating wheel in hybridization oven. The following day, the samples were first washed 3 times for 5 min and afterwards 4 times for 30 min with warm (55 °C) Hyb I. Next, the eggs were first washed with 2:1 and afterwards 1:2 solution of Hyb I:PBT at 55 °C for 10 min. Finally, the eggs were washed 4 times with PBT for 10 min at RT. All the washing steps were performed on rocker.

2.4.3 Digoxigenin antibody incubation

The samples were blocked in blocking solution containing 1% bovine serum albumin (BSA) and 0.5% normal goat serum (NGS) twice for 30 min at RT on the rocker. Next, the first AB

was diluted in the fresh blocking solution (anti-dig AB, Roche, 1:5000). The eggs were incubated with the first AB at 4 °C O/N on the rocker. The following day, the eggs were first washed 3 times for 5 min and afterwards 3 times for 15 min with PBT at RT.

2.4.4 Colorimetric detection

In order to allow for colorimetric detection of the anti-dig AB, the samples were incubated in alkaline phosphatase (AP) buffer containing 100 mM Tris (pH 9.5), 50 mM MgCl₂, 100 mM NaCl and 0.2% Tween-20 in Milli-Q water. The eggs were incubated in the AP-buffer twice for 5 min at RT. The colorimetric reaction was performed with staining solution containing 18.75 mg/ml nitro blue tetrazolium chloride (NBT) and 9.4 mg/ml 5-bromo-4-chloro-3-indonyl phosphate (BCIP) in 67% dimethyl sulfoxide (DMSO), which was diluted with the AP-buffer (1:500). To prevent photobleaching, the samples were kept in the dark during the staining procedure. When specific expression pattern was observed, the staining reaction was stopped with PBT: the eggs were washed 6 times with PBT for 5 min at RT on the rotating wheel. In order to reduce background, the eggs were gradually dehydrated and rehydrated with ethanol diluted in PBT. The eggs were incubated in 50% and 75% ethanol in PBT for 5 min, in 100% ethanol for 10 min and subsequently rehydrated with 50% and 25% ethanol in PBT. Finally, the eggs were washed 3 times with PBT for 5 min at RT. After the final wash, PBT was removed and the eggs were covered in mounting medium containing 4',6-diamidin-2-phenylindol (DAPI) (Vectashield, Vector laboratories).

In general, in order to investigate expression profile of the unknown gene, staining experiment was performed at least twice with parallel positive (probe of gene with well-known expression pattern) and negative (sense probe) control staining experiments. Note that in order to quantify gene expression strength difference between the KD and WT embryos, both WT and KD samples were incubated in the staining solution for the exactly same time period. The staining reaction was stopped as soon as the well-known expression pattern of the gene was observed in the WT samples. In order to confirm the result, the experiment was repeated 3 times.

2.5 Gene expression silencing

2.5.1 Double stranded RNA synthesis

Double stranded RNA (dsRNA) was synthesized using MEGAscript T7 kit (Ambion, Life Technologies). Diluted second PCR reaction outcome (1:1 dilution with nuclease-free water) performed with the 5' and the 3' universal primers according to the section 2.2.4 was used as a template. The reaction contained 8 µl of the template, 2 µl of all four dNTPs, 2 µl of T7 polymerase, 2 µl of reaction buffer and 6 µl of nuclease-free water. The reaction was incubated at 37 °C O/N in the heating block. The following day, the transcription reaction was stopped by adding 115 µl of nuclease-free water and 15 µl of 5 M ammonium acetate solution in 100 mM EDTA. For purification, 150 µl of phenol-chloroform solution (Roth) were added and the suspension was vortexed for 30 s. The samples were centrifuged at 5000 g and RT for

5 min. The upper aqueous phase was transferred to a new Eppendorf tube. The dsRNA was precipitated in 150 µl of isopropanol at -20 °C for 1 h. The samples were centrifuged at 20000 g and 4 °C for 15 min. The supernatant was discarded and the pellet was washed twice with 300 µl of 70% ethanol. After the last wash, the samples were centrifuged at 12000 g and 4 °C for 5 min and ethanol was completely removed. The pellet was air dried for 5 min at RT in the open Eppendorf tube. The pellet was resuspended in 16 µl of nuclease-free water. The dsRNA concentration was measured with a spectrophotometer. The samples were diluted to the concentration of 1 µg/µl and distributed to 10 µl aliquots. Isolated dsRNA was stored at -20 °C until further use.

2.5.2 Parental RNA interference

Parental RNA interference (pRNAi) was used to transiently silence the expression of the genes of interest. Virgin *T. castaneum* females were injected with dsRNA of the gene of interest in the pupal stage. The injection was performed with fine needles produced from glass capillaries (Hilgenberg). The glass capillaries were pulled apart with a laser needle puller P-2000 (Sutter Instrument Co.) with the following settings: heat 400, filament 4, velocity 60, delay 225, pull 150. To create a sharp tip, the needles were opened by cutting the tip in 45° angle with a razor blade.

Female pupae were stuck with the dorsal side of their most posterior segments to the double sided tape (Doppelband Fotostrip, Transparent, TESA), which was stuck with the other side to the microscope slide. Pupae were injected with the needle containing dsRNA from the lateral side between the 3rd and the 4th abdominal segment until the pupae stretched from the injected volume. For the small scale RNAi experiments (miniscreen) approx. 50 pupae were injected with dsRNA of one gene. However, for the large scale RNAi experiments (RNA-seq after RNAi) about 350 pupae for one collection interval and per one BR were injected with dsRNA of one gene (either *Tc-zen1* or *Tc-zen2*). On average, 10 µl of dsRNA was necessary to inject approx. 35-40 female pupae. After the injection, pupae were gently removed from the double sided tape and placed on the stock flour in the plastic vials. Always 1 male per 4 females was added to the cohort. Approx. 4-5 DAI the pupae eclosed. After 7 DAI the first egg lays were collected.

2.6 Phenotypic scoring after RNA interference

2.6.1 Nuclear staining

Phenotype after pRNAi was investigated on different *T. castaneum* developmental stages. Possible embryonic defects were investigated in the stages from the uniform blastoderm to the retracting germband stage using nuclear staining. The fixed embryos were washed 3 times with 0.5% PBT (0.5% Tween-20 in PBS). After the last wash, the eggs were covered with mounting medium containing DAPI (Vectashield, Vector Laboratories).

2.6.2 Cuticle preparation

Larval cuticle preparations were used to determine defects during the embryonic development. First, the larvae or eggs (in case of developmental defects due to the KD) were dechorionated as described in the section 2.1.3. Next, the larvae and eggs were transferred from water surface onto a microscope slide with brush. The larvae and eggs were separated one by one and equally distributed on the slide. Afterwards, the specimens were covered with 100 µl of Hoyer's solution containing 0.6% gum Arabic, 1.2 M chloral hydrate and 217 mM glycerol (diluted 1:1 with lactic acid). The larvae and eggs in the Hoyer's solution were covered by a cover slide. To allow for complete tissue digestion, the samples were incubated at 60 °C O/N.

2.6.3 Serosal cuticle integrity determination

As previously mentioned, serosa secretes chitin cuticle (Panfilio, 2008), therefore the phenotype after RNAi of the gene responsible for the serosa tissue specification (*Tc-zen1*) or genes, which might play role in the serosa maintenance (*Tc-zen1* targets), can be scored by observing the presence, absence or integrity of the serosal cuticle. Fixed embryos were transferred directly from the heptane-PFA interface onto a small piece (1x0.5 cm) of Whatman gel blot paper. The eggs were subsequently transferred onto a double sided tape, which was stuck from the other side to a petri dish. The eggs were covered by PBS and the cuticle was observed under the dissecting microscope (Zeiss). The integrity of the serosal cuticle was investigated with a sharp needle by evaluation of the cuticle resistance. The KD samples were compared with WT samples in parallel.

2.7 Protein expression

2.7.1 Protein extraction

Dechorionated eggs stored at -80 °C were first thawed on ice for approx. 5-10 min. Eggs were homogenized in 200 µl of RIPA buffer (Sigma-Aldrich) and the homogenate was incubated on ice for 30 min. Samples were centrifuged at 20000 g and 4 °C for 20 min. After the centrifugation the debris was collected at the bottom of an Eppendorf tube in a form of pellet, but the residual debris was also found on the top of the supernatant in a form of a white foam layer. Roughly 170 µl of supernatant were transferred to a new Eppendorf tube. Since it is not possible to completely avoid the white foam layer, the supernatant was additionally centrifuged at 20000 g and 4 °C for 5 min. Precisely 153 µl of supernatant were transferred to a new Eppendorf tube. In order to measure the protein concentration of the sample, 3 µl of the supernatant were set aside for the Bradford assay (described below). To the remaining 150 µl of supernatant, 150 µl of 2x sample buffer containing 120 mM Tris (pH 6.8), 12.8% glycerol, 4% SDS, 0.1% bromphenol blue and 0.2 M DTT were added. Samples were boiled at 100 °C for 3 min in heating block and afterwards stored at -20 °C until further use.

2.7.2 Protein concentration measurement

Protein concentration of the samples was measured using Bradford reagent (Sigma-Aldrich), which contains Brilliant Blue G dye. After the protein in the sample binds to the dye, the dye-protein complex causes a shift in the absorption maximum of the dye. The amount of absorption is proportional to the protein present in the sample. In order to calculate the unknown protein concentration in the samples, first the concentration of protein standard samples (predefined protein concentration of BSA) has to be measured. Protein standard solution (2 mg/ μ l, Sigma-Aldrich) was diluted with nuclease-free water to the following BSA concentrations: 250, 500, 750 and 1000 μ g/ μ l.

Always 300 μ l of Bradford reagent were mixed either with 10 μ l of the standard samples or 1 μ l of the samples of unknown protein concentration. Bradford reagent was supplemented with 1 μ l of RIPA buffer when mixed with the standard samples and with 10 μ l of nuclease-free water when mixed with the samples of unknown protein concentration. The blank sample was supplemented with both 1 μ l of RIPA buffer and 10 μ l of nuclease-free water. Samples were measured in three technical replicates.

The standard curve was created from the absorbance values of the standard samples and the predefined BSA concentration values in the standard samples. The unknown protein concentration of the samples was afterwards calculated based on the linear regression data distribution model of the BSA absorbance values of the standard samples.

2.7.3 SDS-PAGE

Protein extracts from *T. castaneum* eggs in different developmental stages were run on sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Always 50 μ g of total protein were loaded into each well. Both collection (3%) and separation (12%) gel were run at 200 V. The gel was run for 2.5 h allowing for broad distribution of proteins of size between 40 and 55 kDa. Pre-stained protein ladder (PageRuler, Thermo Fisher Scientific) was used as size standard.

3% collection gel

1.2 ml	90% Mix
0.2 ml	30% Acrylamide
5 μ l	40% APS
5 μ l	TEMED

12% separation gel

3 ml	60% Mix
2 ml	30% Acrylamide
10 μ l	40% APS
10 μ l	TEMED

90% Mix

31 ml Upper Tris
80 ml Milli-Q water

Upper Tris (pH 6.7)

6.1 g Tris
3 ml 10% SDS
fill up to 100 ml with Milli-Q water

5x running buffer

30 g Tris
10 g SDS
144 g Glycine
fill up to 2 l with Milli-Q water

60% Mix

200 ml Lower Tris
160 ml Glycerol
120 ml Milli-Q water

Lower Tris (pH 8.8)

18.2 g Tris
4 ml 10% SDS
fill up to 100 ml with Milli-Q water

Transfer buffer

29 g Tris
14.6 g Glycine
1 l Methanol
1.875 g SDS
fill up to 5 l with Milli-Q water

2.7.4 Western blotting

Separated proteins were transferred from the gel onto a nitrocellulose membrane (Thermo Fisher Scientific) at 100 V for 1 h. The success of transfer was tested by Ponceau staining (Ponceau S solution, Sigma-Aldrich). The membrane was submerged in the Ponceau solution for 5 min. After the protein visualization the lanes and the correct orientation of the membrane were marked. The Ponceau stain was removed by incubation of the membrane in 0.1 M NaOH for 1 min. The NaOH was removed and the membrane was washed with Milli-Q water 3 times for 5 min.

Membrane was blocked for 1 h with TBST blocking solution containing 0.1 M Tris (pH 7.5), 150 nM NaCl, 0.1% Tween-20 and 3% milk powder (Bebivita). After the blocking, the membrane was incubated with the first AB diluted in the fresh blocking solution (Tc-Zen1 and Tc-Zen2 AB, 1:1000, Tub AB, 1:10000) O/N at 4 °C. Following day, the first AB was removed and the membrane was washed 3 times with the TBST blocking solution for 10 min. Next, the membrane was incubated with the secondary AB diluted in the fresh TBST blocking solution (anti-rabbit and anti-mouse HRP ABs, Novex, 1:10000) for 1 h at RT. The secondary AB was removed and the membrane was washed with the TBST blocking solution 3 times for 10 min.

After the last wash, the membrane was dried gently with the paper towel. Next, the membrane was incubated in ECL substrate (WesternSure® ECL Substrate, LI-COR) for 5 min. Digital detection was performed on western blot developing machine (C-DIGIT, LI-COR). Signal was measured for 12 min with high sensitivity settings. The software (Image Studio, LI-COR) collects six pictures with six different signal intensities and exposure

settings in order to allow for a sensitive data collection (cumulative signal integration of multiple time points of the blot).

2.7.5 Cryo-sectioning

Eggs older than 16 h AEL cannot be devitellinated by methanol shock (see section 2.1.5). Therefore, in order to perform AB staining of the embryos older than 16 h AEL, the embryos had to be sectioned.

Dechorionated and fixed embryos were brought from methanol to 0.1% PBT solution as described in section 2.4.2. Eggs were washed twice with PBT for 5 min. Afterwards, the eggs were embedded in warm liquid sucrose-agarose embedding medium containing 15% sucrose and 2% agarose in PBS. After the medium became solid, small blocks containing one embryo each, were cut out with scalpel blade. The blocks were stored in the 30% sucrose solution O/N at 4 °C. Next day the blocks were embedded in Tissue freezing medium (Leica Biosystems) on metal block holder and fast frozen in ice cold isopenten (2-methylbutan, Roth). The frozen embryos were sectioned with a cryostat (CM1850, Leica Biosystems) at -20 °C to 20 µm (longitudinal) or 30 µm (cross) sections. The cryo-sectioned material was collected on the specialized microscope slides (SuperFrost® Ultra Plus, Menzel Gläser, VWR) and dried O/N at RT.

2.7.6 Antibody staining

The following protocol was used for both AB staining of whole mounts and cryo-sectioned material. Unless stated otherwise, the steps were performed at RT. In the case of whole mounts, embryos were first brought from methanol to 0.1% PBT as described in section 2.4.2. The embryos or cryo-sectioned material were washed twice with PBT for 5 min. Afterwards, the samples were blocked with blocking solution containing 2% BSA and 1% NGS in PBT. The first AB was diluted in the fresh blocking solution (Tc-Zen1 and Tc-Zen2 AB, 1:1000). Samples were incubated with the first AB in light-proof boxes O/N at 4 °C on rocker. Following day, the first AB was removed and the samples were washed with PBT 6 times for 10 min. Next, the samples were blocked with the blocking solution for 1 h. The secondary AB was diluted in the fresh blocking solution (Alexa488, Invitrogen, 1:400). The samples were incubated with the secondary AB for 3 h. Afterwards, the secondary AB was removed and the samples were washed with PBT 6 times for 10 min. Finally, PBT was removed and the samples were covered with mounting medium containing DAPI (Vectashield, Vector Laboratories) and mounted on microscope slides either with two (whole mounts) or with one (cryo-sectioned material) spacer.

2.8 Visualization of specimens

2.8.1 Sample mounting for microscopy

The eggs were transferred from Eppendorf tube onto microscope slide, which was equipped with two spacers (2 halves of cover slip on top of each other glued with commercially available nail polish) at both edges. Long cover slip (24x50 mm) was placed on top of the spacers, to which it was connected with 2 drops of glycerol. The cover slip was gently touching the surface of the eggs. This mounting method enables gentle rotation of the eggs during microscopy and therefore, orientation of the embryos to various views is possible.

2.8.2 Microscopy and picture processing

Specimens prepared according to different protocols (ISH or AB staining) were visualized with different microscopes. Pictures of ISH staining with corresponding DAPI staining were taken with AxioPlan 2 microscope (Zeiss). The same microscope was used for obtaining the pictures of cuticle preparations, but dark field was applied. For each specimen, series of pictures in the consecutive focal planes were taken with AxioVision software (Zeiss). Later, the series of pictures was combined to a projection with Helicon Focus 6 software.

Fluorescent AB staining of both whole mount and sectioned specimens were visualized with Axio Imager 2 equipped with ApoTome 2 (Zeiss). First, the Z stacks for each color were taken. Afterwards, the Z stacks were combined to maximum intensity projection (MPI) with ZEN blue software (Zeiss).

The high magnification pictures of AB staining of cryo-sectioned specimens were obtained with confocal microscope LSM 700 with M2 Imager (Zeiss). First, the Z stacks for each color were taken. Afterwards, the Z stacks were combined to MPI with ZEN 2 black software (Zeiss).

All the pictures were later processed in Photoshop (CS5.1, Adobe). Any changes (contrast, brightness, color) were applied to whole picture. Micrographs were combined to panels with Illustrator (CS5, Adobe).

2.9 *In silico* analysis

2.9.1 Sequence alignment

Pairwise alignment of *Tc-zen1* and *Tc-zen2* coding sequences (CDS) was performed in ClustalX version 2.1 (Larkin et al., 2007). The sequence alignment was edited in BioEdit sequence alignment editor (Hall, 1999).

2.9.2 Identification of conserved non-coding regions

Hox3 loci of four closely related *Tribolium* species were compared: *T. castaneum* (Brown et al., 2009), *T. freemani* (Hinton, 1948), *T. madens* (von Charpentier, 1825) and *T. confusum* (Du Val, 1863). Genomic sequences in the format of assembled scaffolds of *T. freemani*, *T. madens*, and *T. confusum* were obtained in FASTA-formatted files (version 26 March 2013 for each species' file) from the BeetleBase.org FTP site at Kansas State University (<ftp://ftp.bioinformatics.ksu.edu/pub/BeetleBase/>). The Hox3 locus sequence of *T. freemani*, *T. madens*, and *T. confusum* was obtained by blastn search of *zen1* and *zen2* genomic DNA sequence of these three species in their respective complete genome sequences using a standalone installation BLAST+ (version 2.2.30 (Altschul et al., 1997; Camacho et al., 2009)). Subsequently, 5 kb long region upstream of *zen1* and downstream of *zen2* sequence was extracted with “Fetch Alignments/Sequences” tool with “Extract genomic DNA” function on the web-based platform Galaxy (<https://usegalaxy.org/>).

In order to identify conserved non-coding regions among Hox3 loci of four *Tribolium* species, two different programs were used: standalone program Multiple Species Sequence Analysis version 1.1.0 (MUSSA, California Institute of Technology) (<http://mussa.caltech.edu/mussa>) and online based mVista tool (Mayor et al., 2000; Frazer et al., 2004). Both programs required submission of Hox3 locus genomic sequences in FASTA format and the corresponding annotation text file, in which the position of *zen* genes was annotated.

First, MUSSA performed all possible pairwise alignment comparisons between the four Hox3 loci sequences. In the second step, multiple sequence alignment comparison of all four queries was performed and regions with the same sequence in all four Hox3 loci sequences were visualized with connecting red or blue lines regardless of the region's position. The comparison was performed with default setting of 30 nt sliding window and the threshold was manually changed from 90% to 100% of sequence identity.

The second program, mVista, investigates conserved domains by recursively finding strong anchors from the collection of maximal matches in the sequences. The default parameters, with which the analysis was run, were 70% sequence identity within 100 nt long sliding window. Both programs were also used for pairwise comparison of *Tc-zen1* and *Tc-zen2* genomic sequences.

2.10 RNA-sequencing after RNA interference

2.10.1 RNA-sequencing

In order to identify genes that are differentially expressed after RNAi of *Tc-zen1* and *Tc-zen2*, the transcriptomes of KD and corresponding WT samples were sequenced. For this purpose RNA-sequencing (RNA-seq) after RNAi was employed. The most common technique used for RNA-seq is sequencing by synthesis (Illumina). The samples are sequenced on a glass slide with 8 lanes called a flow cell.

In the first step, the quality of RNA samples (prepared as described in the section 2.2.1) was examined by capillary electrophoresis. Since the RNA samples quality was high, the cDNA libraries were created according to TrueSeq protocol (Illumina). In this procedure, the cDNA is fragmented and each fragment is tagged at both ends with adaptor sequences. The adaptor sequences consist of three different regions: sequencing primer binding sites, indices and regions complementary to the oligo present on the flow cell. Fragments of cDNA in one sample are tagged with the same index region in order to allow for multiplexing of several samples on one lane.

Next, the fragmented cDNA is bound to the flow cell through complementary sequences of adaptors and is subsequently clonally amplified through bridge amplification process, which generates clusters. This process is repeated multiple times and occurs simultaneously for millions of clusters across the flow cell resulting in clonal amplification of all the fragments. For the sequencing by synthesis, DNA polymerase and 4 fluorescently tagged dNTPs are added. During each sequencing cycle, all four single labeled dNTPs are competing for addition to the growing chain, but only one is successfully incorporated based on the sequence of the template. The fluorescent label serves as reversible terminator for polymerization, because the next dNTP cannot be incorporated unless the label is enzymatically cleaved.

After the addition of each nucleotide (after every single cycle), the clusters are excited by a light source and a characteristic fluorescent signal is emitted. The emission wavelength along with the signal intensity determine the base call. The cDNA fragments were sequenced from both ends in a pair-end sequencing process. Hundreds of millions of clusters are sequenced in a massively parallel process (Wang et al., 2009; Wilhelm and Landry, 2009; Illumina ©Inc, 2010).

2.10.2 Sample preparation for RNA-sequencing experiments

Two separate RNA-seq experiments have been performed in this project. First, the *Tc-zen1^{RNAi}* and *Tc-zen2^{RNAi}* samples from early developmental stages were prepared. The eggs of the injected females were staged to 6-10 h AEL (*Tc-zen1^{RNAi}*) and 10-14 h AEL (*Tc-zen2^{RNAi}*). WT samples in the respective time points were collected in parallel. KD efficiency was evaluated by RT-qPCR (both *Tc-zen1^{RNAi}* and *Tc-zen2^{RNAi}*) and phenotypic penetrance was scored by serosa integrity determination (*Tc-zen1^{RNAi}*) and cuticle preparations (*Tc-zen2^{RNAi}*). The eggs were collected in three biological replicates (BRs).

For the second RNA-seq experiment, *Tc-zen2^{RNAi}* samples from two different late developmental stages were prepared. The eggs of the injected females were staged to 48-52 h AEL (pre-rupture stage) and to 52-56 h AEL (post-rupture stage). Respective WT samples were collected in parallel. KD strength was confirmed by RT-qPCR and phenotypic penetrance was scored by cuticle preparations. The eggs were collected in three BRs.

The RNA was isolated according to the protocol described in the section 2.2.1. Samples of isolated RNA were sent to Cologne Centre for Genomics (CCG), where RNA-seq was performed. Samples of each RNA-seq experiment were multiplexed on two lanes (Table 2.2).

Table 2.2A. *Tc-zen1* KD (blue) and corresponding WT1 samples (orange), *Tc-zen2* KD (purple) and corresponding WT2 samples (green) from early developmental stages were multiplexed on 2 lanes according to the following scheme:

lane	sample					
1	BR1_zen1	BR2_zen1	BR1_WT1	BR2_WT1	BR3_zen2	BR3_WT2
2	BR1_zen2	BR2_zen2	BR1_WT2	BR2_WT2	BR3_zen1	BR3_WT1

Table 2.2B. *Tc-zen2* KD samples from the pre-rupture stage (48-52 h AEL, blue) and the corresponding WT samples (48-52 h AEL, red), *Tc-zen2* KD samples from the post-rupture stage (52-56 h AEL, purple) and the corresponding WT samples (52-56 h AEL, green) were multiplexed on 2 lanes according to the following scheme:

lane	sample							
1	BR1_WT 48-52	BR2_WT 48-52	BR1_KD 48-52	BR2_KD 48-52	BR3_WT 52-56	BR3_KD 52-56	BR4_WT 48-52	BR4_KD 48-52
2	BR3_WT 48-52	BR3_KD 48-52	BR1_WT 52-56	BR2_WT 52-56	BR1_KD 52-56	BR2_KD 52-56	BR4_WT 52-56	BR4_KD 52-56

Note that, BR for RNA-seq experiment was defined as the RNA sample isolated from the eggs, which were laid by females from only one cohort. Three different cohorts were used for both RNA-seq experiments in order to obtain three BRs. However, in order to make the results from the second RNA-seq experiment (from late developmental stages) statistically more robust, we sequenced two samples (two technical replicates) from the second BR. Because it is highly unlikely that the eggs collected in different days from one cohort are laid by the same females, we decided to consider the second sample from the BR2 as the fourth BR.

2.11 Generating pipeline for RNA-sequencing data analysis

2.11.1 Joining the files

To compare the list of genes in the output files of the differential expression (DE) analysis performed with two different programs, the files were joined (merged) according to the column, in which the content shared by both files (name of genes) is present. The merging was performed with “Join, Subtract and Group” tool with “Join two Datasets” function on the web-based platform Galaxy (<https://usegalaxy.org/>). The data from merged files were transferred to Excel, where subsequently the calculation of overlapping genes was performed.

2.11.2 Ranking test

To compare fold change (FC) values assigned to the same differentially expressed genes identified by two different programs, the Bland Altman Leh ranking test (Bland and Altman, 1986) was employed. The test was performed with “BlandAltmanLeh” package (Lehnert, 2015) in R Studio (R Core Team, 2016). The test assigns ranks to FC values of the same gene in both datasets and evaluates agreement between two methods, by calculating the mean and the difference between the ranks. Based on the data distribution model the test calculates the critical value, until which the results from the two different methods can be considered the same.

2.12 RNA-sequencing data analysis

2.12.1 Quality control

Prior to any further data analysis, the quality of the raw sequencing reads obtained after RNA-seq has to be examined. Quality control allows for detection of sequencing errors, PCR artifacts and contaminations (e.g.: overrepresented sequences of ribosomal or mitochondrial RNA, bacterial contamination). The quality control of the Illumina reads obtained in this project was examined with FastQC (Andrews, 2014), which provides an overview of sequence quality, GC content, the presence of adaptors, overrepresented *k*-mers and duplicated reads. Acceptable levels of *k*-mers, duplications and the GC-content are organism- and experiment-specific, but these values have to be homogenous for samples within one experiment.

2.12.2 Trimming

The first step of the RNA-seq data analysis is trimming. In general, quality of the read sequence decreases towards the 3' end of the read. However, because of adaptor sequences the quality might also be reduced at the 5' end. If the quality of the 5' end is not sufficient, it is better to remove the low-quality 5' part of the read to increase the quality of mapping. Reads obtained in this project were trimmed with the Trimmomatic (version 0.36) (Bolger et al., 2014) with the following command:

```
java -jar trimmomatic-0.36.jar PE -phred33 -threads 16 -trimlog <logfile name> <input file names> -baseout <prefix for output file names> ILLUMINACLIP:adapters/Adapter_indices_TcasRNA-seq-PE.fa:2:30:10:6:TRUE LEADING:40 SLIDINGWINDOW:4:20 MINLEN:36
```

The low quality bases at the 5' end were removed using “LEADING:40” function, which trimmed all the bases from the 5' end with the quality score lower than 40. Adaptor sequences were discarded with “ILLUMINACLIP” function, with which all the predefined TrueSeq adaptor sequences were trimmed. To remove poor quality bases in the read, “SLIDINGWINDOW:4:20” function was used to remove any base with quality score lower than 20 in the region of 4 bases. Finally, any read shorter than 36 bases was automatically

discarded with the “MINLEN:36” function. Quality control of the reads was again examined after the trimming procedure with FastQC.

2.12.3 Filtering overrepresented sequences of ribosomal and mitochondrial RNA

The quality control of the raw sequencing reads reported overrepresented sequences of *T. castaneum* ribosomal and mitochondrial RNA. In general contamination with these RNA sequences can lower the mapping efficiency. Therefore RNA reads containing these sequences were filtered out by mapping with Bowtie2 (Langmead and Salzberg, 2012). First, a database of all the *T. castaneum* ribosomal and mitochondrial RNA was created by searching the NCBI gene database with the query “tribolium [organism] AND (ribosomal OR mitochondrial OR mitochondrion) NOT (whole genome shotgun) NOT (Karoochloa purpurea)”. FASTA-formatted sequences of all these genes were downloaded and a database compatible with Bowtie2 was created using “bowtie2-built” function. Subsequently, RNA reads were mapped to this database and only those reads that did not map were kept and extracted to a separate file. Quality control of the reads was again examined with FastQC after the filtering procedure and no overrepresented sequences of ribosomal and mitochondrial RNAs were detected.

2.12.4 Mapping

Mapping efficiency represents the percentage of mapped sequencing reads and it highly depends on the quality of the reference genome used. Since the annotation of the *T. castaneum* genome is in continuous process, several automated gene prediction sets, three final assembly versions and two official gene sets (OGS) versions are currently available (summarized in Table 2.3).

Table 2.3. Overview of currently available versions of different *Tribolium castaneum* gene predictions, corresponding assemblies and corresponding official gene sets (OGS).

Gene prediction	Corresponding assembly version	Corresponding OGS
augustus 2	----	----
augustus 3	<i>Tcas</i> 3.0	OGS2
augustus 4	<i>Tcas</i> 4.0	----
augustus 5	<i>Tcas</i> 5.2	OGS3

2.12.4.1 Mapping to official gene set

Trimmed and filtered sequencing reads were mapped to *T. castaneum* OGS3. A major complication in quantifying the transcript abundancies is the fact that reads do not always

map uniquely, but rather map to several locations (multi-mapping reads). For this reason RSEM (RNA-seq by Expectation Maximization) (Li and Dewey, 2011) was employed for mapping. RSEM implements an expectation maximization algorithm, which assigns multi-mapping reads partially to different locations and afterwards it computes maximum likelihood transcript abundance estimates. With this application the raw read count values are presented as non-integers. RSEM uses Bowtie2 as a default aligner for the mapping. For mapping to the OGS3 “Tcas5.2_GenBank.corrected_v5.renamed.mrna.fa” file was used (available in iBeetle Genome Browser - http://bioinf.uni-greifswald.de/tcas/genes/tcas5_annotation/). The transcript abundance estimation was calculated with “rsem-calculate-expression” and “calc-ci” functions. A diagnostic file containing statistics on mapping efficiency was obtained with “rsem-plot-model” function.

2.12.4.2 Mapping to the genome

Trimmed and filtered sequencing reads were mapped to the *T. castaneum* genome (*Tcas* 5.2 assembly version) with STAR aligner (Dobin et al., 2013). For mapping to the genome “GCA_000002335.3_Tcas5.2_genomic.fna” file containing complete genome sequence and “GCA_000002335.3_Tcas5.2_genomic.gtf” file containing annotations of all genes were used. These files were obtained from NCBI genome database (<ftp://ftp.ncbi.nih.gov/genomes/all>). The mapping with STAR was performed with both strict and relaxed parameters. With strict parameters (“outFilterMultimapNmax 1”; “outFilterMismatchNmax 1”) one read could be mapped to maximum 10 locations and only 1 mismatch was allowed. With the relaxed parameters (“outFilterMultimapNmax 100”; “outFilterMismatchNmax 10”) the multimapping allowance was set to 100 locations and the number of mismatches was set to 10. Statistics on mapping efficiency were obtained in the output file using the “Within” function.

2.12.5 Feature counting

Before any biological interpretation can take place, read mapping results have to be summarized in terms of read coverage for genomic features of interest. For RNA-seq data, the strategy taken is to count the number of reads (read summarization) that fall into annotated genes (by mapping to the reference genome). In principle, counting reads that map to a catalogue of features is straightforward. However, one has to consider the problem of counting reads that fall into an intronic region or beyond an annotated region. Therefore establishment of a catalogue of features in the counting software is necessary. In this project the features (exons) were summarized in the annotation file used for mapping to the genome (GCA_000002335.3_Tcas5.2_genomic.gtf). With featureCounts (available in SourceForge Subread, version 1.5.1) (Liao et al., 2013, 2014) raw read counts from the STAR output mapping files were extracted based on the annotation file.

2.12.6 Filtering out genes with low read count

Raw read counts data were transferred from both RSEM and featureCounts output files to Excel file. The count data were organized in a table (count table) which reports, for each sample of all conditions, the number of reads that have been mapped to each gene. Measuring a large fraction of the genes with low read counts can produce a dataset that is biased towards identifying differentially expressed genes with low read counts, since these genes are measured with higher noise (Busby et al., 2013). Therefore, all the genes, which had less than 10 reads assigned in both KD and corresponding WT samples, were filtered out using “count if” function in Excel.

2.12.7 Differential expression analysis

A basic task in the analysis of count data derived from RNA-seq is the detection of differentially expressed genes. In this project, two programs for DE analysis were used: DESeq2 (Love et al., 2014) and EBSeq (Leng et al., 2013). Both of the programs implement negative binomial data distribution model. DE analysis was performed with default settings of both programs. The programs reported the \log_2 FC values of gene expression between the conditions. Genes with \log_2 FC values higher than ± 1 (at least two times higher/lower expression compared to WT) were considered differentially expressed.

2.12.8 Principal component analysis

Difference between the sequenced samples was evaluated by principal component analysis (PCA). PCA is a statistical procedure that converts possibly correlated variables into a set of values of linearly uncorrelated variables (principal components). According to the transformation definition, the first PC has the largest possible variance, which means that it accounts for as much variability in the dataset as possible. Raw read count datasets were used as input files for PCA, performed with the DESeq2 package in R Studio.

2.12.9 Analysis of shared targets

To identify the number of target genes that *Tc-zen1* and *Tc-zen2* share during early embryogenesis, comparative analysis of differentially expressed genes after *Tc-zen1*^{RNAi} and *Tc-zen2*^{RNAi} was performed. Due to the fact that WT and KD samples of two different time points (described in the section 2.10.2) (WT1 and *Tc-zen1* KD samples: 6-10 h AEL and WT2 and *Tc-zen2* KD samples: 10-14 h AEL, hereafter referred to as WT shift) were sequenced, the expression change of the genes during the WT shift was taken into account. Therefore, three groups of isoforms were created according to the following procedure: first, raw read count data from WT1 and WT2 samples were filtered in order to discard all the isoforms that were assigned less than 10 reads (as described in the section 2.12.6). Next, DE analysis of filtered datasets was performed with the DESeq2. Resulting group of isoforms with P_{adj} value

≤ 0.1 was divided into three groups based on the FC values: non-differentially expressed isoforms ($-2 > FC < 2$), strongly upregulated isoforms ($FC \geq 2$) and strongly downregulated isoforms ($FC \leq -2$). These three groups of isoforms were used as the input datasets for the further analyses.

The group of non-differentially expressed isoforms within the WT shift was first compared to all strongly differentially expressed isoforms after *Tc-zen1*^{RNAi} ($-2 \geq FC \geq 2$, $P_{adj} \leq 0.1$). The group of isoforms overlapping between these two datasets was afterwards compared with all strongly differentially expressed isoforms after *Tc-zen2*^{RNAi} ($-2 \geq FC \geq 2$, $P_{adj} \leq 0.1$). The genes overlapping between these two datasets were marked as shared by *Tc-zen1* and *Tc-zen2*. In the second step of the comparative analysis, the group of isoforms not differentially expressed within WT shift and strongly differentially expressed after *Tc-zen1*^{RNAi} was compared with all weakly differentially expressed isoforms after *Tc-zen2*^{RNAi} ($-1 < FC < 1$, $P_{adj} \leq 0.1$). The genes identified after the second comparison were marked as shared by *Tc-zen1* and *Tc-zen2*, but with the lower threshold by *Tc-zen2*.

Similarly to the previous comparison, the isoforms that were strongly up- and downregulated within the WT shift were first compared with all strongly differentially expressed isoforms after *Tc-zen1*^{RNAi} ($-2 \geq FC \geq 2$, $P_{adj} \leq 0.1$). The group of overlapping isoforms was subsequently compared first with all strongly differentially expressed isoforms after *Tc-zen2*^{RNAi} ($-2 \geq FC \geq 2$, $P_{adj} \leq 0.1$) and afterwards with all weakly differentially expressed isoforms after *Tc-zen2*^{RNAi} ($-1 < FC < 1$, $P_{adj} \leq 0.1$). Based on these comparisons, genes that are activated and repressed by *Tc-zen1* and at the same time shared with *Tc-zen2* (on both high and lower thresholds) were identified.

To identify genes that are either activated, or repressed by *Tc-zen2* (both with high and lower thresholds) and at the same time shared with *Tc-zen1*, the isoforms that were strongly up- and downregulated within the WT shift were first compared with all strongly differentially expressed isoforms after *Tc-zen2*^{RNAi} ($-2 \geq FC \geq 2$, $P_{adj} \leq 0.1$). The group of overlapping isoforms was subsequently compared with all strongly differentially expressed isoforms after *Tc-zen1*^{RNAi} ($-2 \geq FC \geq 2$, $P_{adj} \leq 0.1$). In the second step, the isoforms that were strongly up- and downregulated within the WT shift were compared with all weakly differentially expressed isoforms after *Tc-zen2*^{RNAi} ($-1 < FC < 1$, $P_{adj} \leq 0.1$). The group of overlapping isoforms was subsequently compared to all strongly differentially expressed isoforms after *Tc-zen1*^{RNAi} ($-2 \geq FC \geq 2$, $P_{adj} \leq 0.1$). All the datasets were compared by joining (merging) as described in the section 2.11.1.

2.12.10 Gene ontology term analysis

To retrieve a functional profile of differentially expressed genes identified after the RNA-seq during late embryogenesis, gene ontology (GO) term analysis was performed with Blast2GO (Conesa et al., 2005). Four datasets of differentially expressed genes were used as input files for the GO term analysis: differentially expressed genes after *Tc-zen2*^{RNAi} in the pre-rupture stage (48-52 h AEL), differentially expressed genes after *Tc-zen2*^{RNAi} in the post-rupture stage (52-56 h AEL), differentially expressed genes after the WT shift from the pre- to post-rupture

stage (WT 48-52 vs 52-56 h AEL) and differentially expressed genes after *Tc-zen2^{RNAi}* after the shift from the pre- to post-rupture stage (*Tc-z2KD* 48-52 vs 52-56 h AEL). The datasets were blasted against two different databases available in NCBI: “nr” database (non-redundant, GenBank CDS translations + PDB + SwissProt + PIR + PRF excluding environmental samples from WGS projects, version 9 July 2017) and “*Drosophila*” database (*Drosophila melanogaster* gene database, version 9 June 2017). Blasting was performed using insect taxonomy filter.

GO term analysis consists of three steps:

- 1) BLAST: gene description is assigned to the gene of interest based on the sequence homology
- 2) MAPPING: GO terms are assigned to the blast results
- 3) ANNOTATION: the information obtained from blast and mapping steps is evaluated

Only genes that passed all the three steps were considered in the further analysis: 51-54% of genes (when blasted against the *Drosophila* database) and 61-63% of genes (when blasted against the nr database). The functional profile of datasets was retrieved with the “Make Combined Graph” function. *T. castaneum* gene sequences were assigned to each GO term, which were divided into the groups according to the level of GO term. Only GO terms of the level 5 were used for the further analysis. In the next step, GO terms falling into categories of interest were grouped according to similarity in function (summarized in Table 2.4). Afterwards a unique count of *T. castaneum* gene sequences was calculated per each category of interest and the percentage was compared to the rest of the GO terms in the level 5 for each GO type.

Table 2.4. List of GO terms assigned to each category of interest in each GO type.

Category of interest	GO type: Biological process
Stress	response to endoplasmic reticulum stress
	cellular response to oxidative stress
	regulation of translation in response to stress
	regulation of response to osmotic stress
	stress-activated protein kinase signaling cascade
	positive regulation of stress fiber assembly
	age-dependent response to oxidative stress
	regulation of stress fiber assembly
Cuticle	chitin-based cuticle development
	molting cycle, chitin-based cuticle
	cuticle pigmentation
	chitin-based cuticle sclerotization
	ecdysis, chitin-based cuticle
	regulation of adult chitin-containing cuticle pigmentation
	regulation of chitin-based cuticle tanning
	cuticle pattern formation

Cytoskeleton	cytoskeleton organization
	actin cytoskeleton organization
	actin filament organization
	actin filament bundle assembly
	cytoskeleton-dependent cytokinesis
	actin filament-based movement
	regulation of actin cytoskeleton organization
	establishment or maintenance of cytoskeleton polarity
	cytoskeleton-dependent intracellular transport
	microtubule cytoskeleton organization involved in mitosis
	cytoskeletal anchoring at plasma membrane
	oocyte microtubule cytoskeleton organization
	establishment or maintenance of microtubule cytoskeleton polarity
	regulation of actin filament length
Epithelium and morphogenesis	epithelium development
	morphogenesis of an epithelium
	epithelial tube morphogenesis
	cell morphogenesis
	cell part morphogenesis
	cell projection morphogenesis
	imaginal disc morphogenesis
	post-embryonic animal organ morphogenesis
	imaginal disc-derived appendage morphogenesis
	post-embryonic appendage morphogenesis
	cell morphogenesis involved in differentiation
	sensory organ morphogenesis
	epithelial cell development
	gland morphogenesis
	digestive tract morphogenesis
	epithelium migration
	regulation of organ morphogenesis
	Malpighian tubule morphogenesis
	dendrite morphogenesis
	morphogenesis of embryonic epithelium
	embryonic hindgut morphogenesis
	regulation of cell morphogenesis
	regulation of morphogenesis of an epithelium
	antennal morphogenesis
	trachea morphogenesis
	epithelial cell proliferation involved in renal tubule morphogenesis
	epithelial tube formation
	mesoderm morphogenesis
	transepithelial transport
	spermathecum morphogenesis
	heart morphogenesis
	branching morphogenesis of an epithelial tube

Epithelium and morphogenesis	negative regulation of cell morphogenesis involved in differentiation
	morphogenesis of a branching epithelium
	chaeta morphogenesis
	regulation of cell proliferation involved in imaginal disc-derived wing morphogenesis
	male genitalia morphogenesis
	imaginal disc-derived leg joint morphogenesis
	intestinal epithelial cell differentiation
	positive regulation of cell morphogenesis involved in differentiation
	post-embryonic genitalia morphogenesis
	cell elongation involved in imaginal disc-derived wing morphogenesis
Wing disc and pupal and metamorphosis	instar larval or pupal development
	wing disc development
	instar larval or pupal morphogenesis
	metamorphosis
ECM and adhesion	cell-matrix adhesion
	extracellular matrix organization
	negative regulation of cell-cell adhesion
Regulation of gene expression	gene expression
	regulation of gene expression
	negative regulation of gene expression
	positive regulation of gene expression
	regulation of translational initiation
	chromatin silencing
	negative regulation of chromatin silencing
regulation of chromatin silencing	
Transmembrane transport	transmembrane transport
	regulation of transmembrane transport
	regulation of transmembrane transporter activity
	positive regulation of ion transmembrane transporter activity
	positive regulation of ion transmembrane transport
Category of interest	GO type: Molecular function
Regulation of gene expression	DNA binding
	transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding
	transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding
	transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding
	transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific binding
	flavin adenine dinucleotide binding
	purine ribonucleotide binding
	purine nucleoside binding
	purine nucleotide binding

Regulation of gene expression	ribonucleoside binding
	purine ribonucleoside triphosphate binding
	RNA binding
	repressing transcription factor binding
	regulatory region nucleic acid binding
	chromatin insulator sequence binding
Cuticle	structural constituent of chitin-based larval cuticle
	chitin deacetylase activity
Cytoskeleton	myosin binding
	actin binding
	actinin binding
	actin filament binding
Transmembrane transport	ion transmembrane transporter activity
	secondary active transmembrane transporter activity
	primary active transmembrane transporter activity
	organic acid transmembrane transporter activity
	sulfate transmembrane transporter activity
	active ion transmembrane transporter activity
	monoamine transmembrane transporter activity
	taurine transmembrane transporter activity
	phosphate ion transmembrane transporter activity
serotonin transmembrane transporter activity	
Category of interest	GO type: Cellular component
Regulation of gene expression	nucleus
	chromatin
	nuclear chromosome part
	nuclear chromosome
	chromosomal region
	nuclear transcription factor complex
Cytoskeleton	cytoskeleton
	contractile fiber
	unconventional myosin complex
	polymeric cytoskeletal fiber
	cell cortex region
	cell cortex part
	myosin II complex
ECM and adhesion	focal adhesion

3 METHOD DEVELOPMENT

3.1 Generating a pipeline for RNA-sequencing and differential expression data analysis

In this project I performed two RNA-sequencing (RNA-seq) experiments: in the first one, *Tc-zen1*^{RNAi} and *Tc-zen2*^{RNAi} samples, and corresponding wild type (WT) samples in the stages of the expression peaks of *Tc-zen* genes (WT1: 6-10 h AEL and WT2: 10-14 h AEL, respectively; hereafter referred to as early stages) were sequenced. In the second RNA-seq experiment, *Tc-zen2*^{RNAi} and corresponding WT samples of the pre-rupture and the post-rupture stages (48-52 h AEL and 52-56 h AEL, respectively; hereafter referred to as late stages) were sequenced.

Data obtained after the RNA-seq of the early stages were first analyzed by quickNGS pipeline (Wagle et al., 2015). Unfortunately, this pipeline did not include the crucial step of filtering out the genes with low read counts (≤ 10). Not including this step in the pipeline caused generation of high number of genes that were falsely designated as differentially expressed (false positive). In addition, the RNA-seq of the late stages samples was planned to be run on new Illumina machines producing reads with the length of 75 bp, while the “early stages” RNA-seq raw datasets had 100 bp long reads. Thus, we decided to reanalyze the “early stages” RNA-seq datasets and simultaneously generate a pipeline that would be used for the “late stages” RNA-seq data analysis as well.

I compared the output results of several bioinformatic tools in order to choose the ones that most accurately reflect the underlying biological events and processes. To describe the consequences of the different read lengths, we produced a 75 bp dataset by trimming the original 100 bp long reads from their 3' region. Afterwards, I ran the entire comparative analysis for both 100 bp and 75 bp datasets in parallel (Fig. 3.1).

3.1.1 Quality control, trimming and filtering

I first performed quality control assessment (per base sequence quality, per sequence GC content, sequence length distribution, overrepresented sequences, adaptor content, etc.) of the raw sequencing data using FastQC (Andrews, 2014) (Fig. 3.1A). Next, the adaptor sequences and the base pairs with quality score lower than 40 were trimmed from the 5' region of the reads with Trimmomatic (Bolger et al., 2014) (Fig. 3.1B). Afterwards, quality control was performed again and overrepresented sequences of mitochondrial and ribosomal RNA were identified. The overrepresented RNA sequences were filtered out by mapping to the database of these sequences (see Methods section 2.12.3) using Bowtie2 (Langmead and Salzberg, 2012) (Fig. 3.1C). Approx. 9 million reads mapping to the database were filtered out from each file. Finally, the trimmed and filtered datasets were run through the last quality control and passed all the check points.

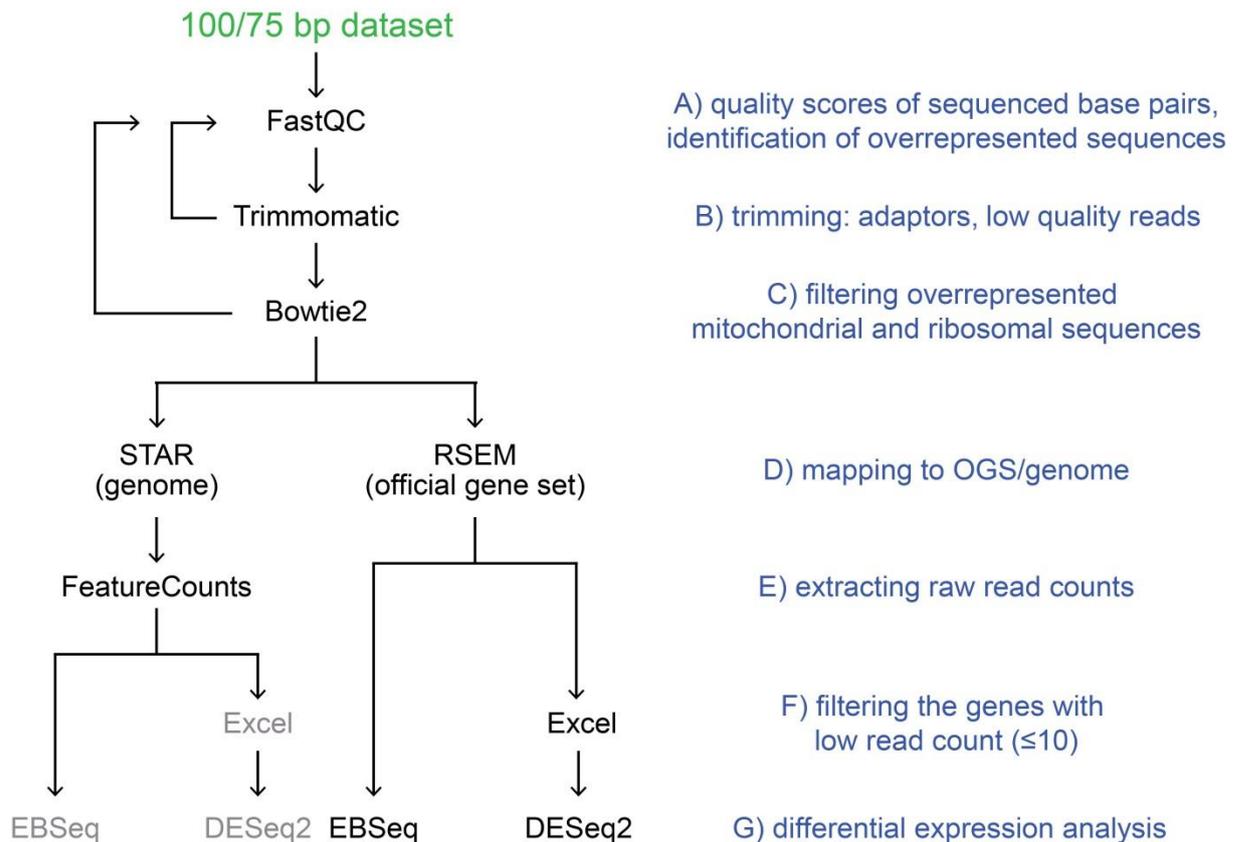


Figure 3.1. Pipeline for RNA-sequencing and differential expression data analysis. Several bioinformatic tools were used for particular steps (A-G) of RNA-seq data analysis. The results from these programs were subsequently compared in order to establish the final pipeline. Tools highlighted by grey color were not used, due to the fact that FeatureCounts outputs did not report sufficient information for further analysis.

3.1.2 Mapping RNA-sequencing data to the genome and to the official gene set

The trimmed and filtered reads were mapped to both *Tribolium castaneum* genome (assembly version *Tcas* 5.2) using STAR (Dobin et al., 2013) and the official gene set 3 (OGS3, 18536 isoforms) using RSEM (Li and Dewey, 2011) (Fig. 3.1D). Next, depending on whether the reads were mapped with STAR or RSEM, the raw read counts were extracted either with FeatureCounts (mapping with STAR) or directly obtained from the RSEM output files (Fig. 3.1E). Subsequently, the genes with less than or equal to 10 reads in both WT and knockdown (KD) samples were filtered out in Excel, using the “count if” function (Fig. 3.1F). Approx. 7300-7500 genes were filtered out from the datasets obtained from RNA-seq of early developmental stages.

Unfortunately, the FeatureCounts output did not contain standard *T. castaneum* gene identifiers (TC IDs), but instead it reported RNA IDs, chromosome coordinates for each exon and DNA strand direction. In general, gene identifiers should be present in the annotation file (gff), which is required for mapping to the genome. However, *T. castaneum* gff file does not contain TC IDs and therefore, they could not have been reported in the raw read count output of FeatureCounts. Thus, we tried to extract TC IDs based on the RNA IDs from another gff file available for *T. castaneum* genome. However, the RNA IDs between the two gff files did

not correspond. Therefore, we decided not to pursue differential expression (DE) analysis with the genome mapping results obtained from STAR (Fig. 3.1, grey text). DE analysis was performed only with the raw read count datasets, which were obtained from the mapping to the OGS3 by RSEM.

3.1.3 Differential expression analysis

I performed DE analysis using two different programs based on the same data distribution models (negative binomial), but different DE tests (P_{adj} -value calculations): DESeq2 (Love et al., 2014) and EBSeq (Leng et al., 2013) (Fig. 3.1G). Although the programs use the same underlying data distribution model, they did not generate lists with the same genes. Thus, the DE analysis results from both programs were further compared in detail. Ultimately, I investigated the influence of both the read length (75 bp and 100 bp) and the programs used (DESeq2 and EBSeq) on the DE analysis results.

3.1.3.1 Influence of the read length

To begin with, I evaluated the impact of the read length on the number of genes identified by both DESeq2 and EBSeq. Both programs identified more genes by analysis of the 75 bp dataset than the 100 bp dataset (100 more genes with DESeq2 and 40 more genes with EBSeq). These 140 genes comprised both differentially expressed and non-differentially expressed genes. Although the DE analysis of the 100 bp dataset identified a lower number of genes, it also identified genes that were omitted by the analysis of the 75 bp dataset. Therefore, in order to investigate whether by choosing the 75 bp dataset over the 100 bp dataset for DE analysis I lose potential candidate target genes, I evaluated fold change (FC) values of these genes. The genes that were exclusively identified by the DE analysis of the 100 bp dataset had FC values lower than two (considered not strongly differentially expressed). Therefore, by considering the 75 bp dataset as input dataset to the pipeline, I would not lose genes that could potentially be designated as relevant *Tc-zen1* or *Tc-zen2* candidate targets in early embryogenesis.

Next, I compared the mapping efficiency statistics (percentage of mapped reads). Besides the raw read count information, it is also possible to retrieve mapping efficiency statistics from the so called “Diagnostic file”. In this file the alignment statistics describe the proportion of uniquely mapping reads and multi-mapping reads (reads mapping to several genes) to those that were not mapped at all (Fig. 3.2). I compared the mapping efficiency of the shorter 75 bp reads to the mapping efficiency of the longer 100 bp reads. On average, 8-15% more reads mapped uniquely, when their length was 75 bp. Further, only 3-4% more reads mapped to multiple genes when their length was shorter (75 bp). These results suggest that by shortening the read length we gained higher mapping efficiency, while the number of multi-mapping reads did not increase drastically (percentages summarized in Table 3.1). Moreover, this result is consistent with the identification of higher number of genes after DE analysis of the shorter reads (75 bp dataset). Since by analyzing the 75 bp datasets we were

able to achieve higher mapping efficiency and, therefore, identify more genes, I decided not to consider the 100 bp dataset for the further comparative analysis.

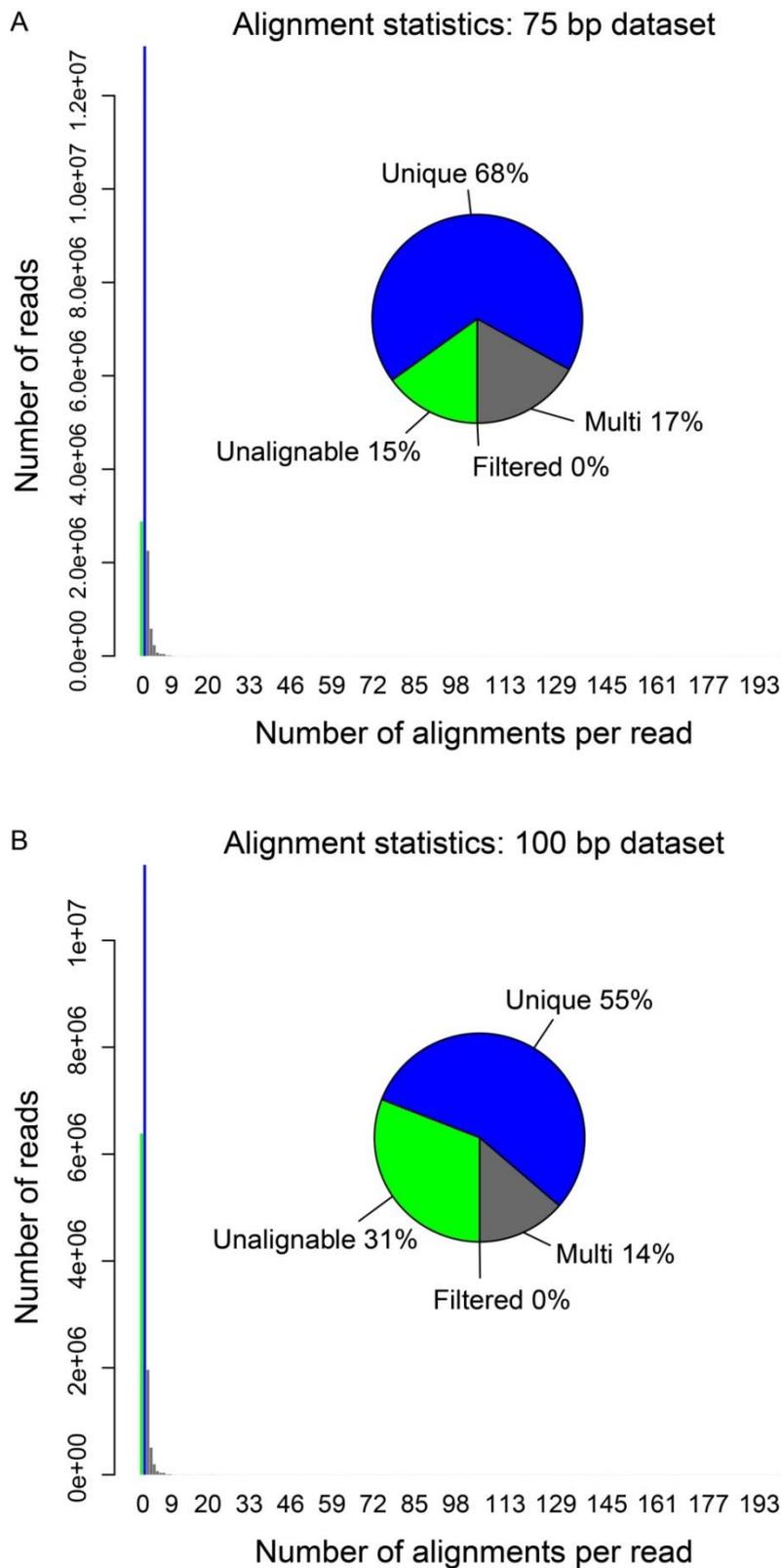


Figure 3.2. Output diagnostic files of *Tc-zen1KD* (biological replicate 1) sample. Comparison of the alignment statistics of one of the “early stages” RNA-seq samples. When the reads of the same sample were trimmed to 75 bp, 68% of the reads aligned uniquely, 17% aligned to multiple genes and 15% did not align at all (A). When the reads of the same sample were kept to the original length of 100 bp, the proportion of the uniquely aligning reads was 55%, multi-aligning reads 14% and 31% of the reads did not align at all (B).

Table 3.1. Comparison of the alignment statistics of all the sequenced samples from early stages. Information about mapping efficiency was retrieved from the RSEM output diagnostic files. The comparison of the mapping efficiency of the same samples mapped as 75 bp long reads and 100 bp long reads is shown. The samples mapped as 75 bp long reads have higher mapping efficiency, gaining 8-15% of uniquely mapping reads and only 3-4% of multi-mapping reads. Overall alignment rate represents the sum of uniquely mapping and multi-mapping reads. BR-biological replicate.

Sample name	Unique-mapping [%]		Multi-mapping [%]		Overall alignment rate [%]	
	100 bp	75 bp	100 bp	75 bp	100 bp	75 bp
<i>Tc-zen1KD_BR1</i>	55	68	14	17	69	85
<i>Tc-zen1KD_BR2</i>	55	69	13	17	68	86
<i>Tc-zen1KD_BR3</i>	54	64	15	19	69	83
WT1_BR1	51	66	14	18	65	84
WT1_BR2	54	68	14	17	68	85
WT1_BR3	56	67	14	17	70	84
<i>Tc-zen2KD_BR1</i>	57	68	14	17	71	85
<i>Tc-zen2KD_BR2</i>	58	68	14	17	72	85
<i>Tc-zen2KD_BR3</i>	57	68	14	17	71	85
WT2_BR1	56	64	15	18	71	82
WT2_BR2	58	67	14	17	72	84
WT2_BR3	55	66	14	17	69	83
range	51-58	64-69	13-15	17-19	65-72	82-86
difference	8-15		3-4		11-19	

3.1.3.2 Influence of the program used for the differential expression analysis

Finally, I compared the datasets of differentially expressed genes identified by EBSeq to those identified by DESeq2. The two programs did not identify the same number of genes: with the same considerations and criteria (both differentially expressed and non-differentially expressed genes, $P_{adj} \leq 0.1$), DESeq2 identified 1904 genes and EBSeq 1496 genes. Out of

these, only 1243 genes were identified by both programs. This means that 661 genes were exclusively identified by DESeq2 and 253 genes exclusively by EBSeq.

In order to evaluate, whether the genes identified by both DESeq2 and EBSeq (1243 genes) are designated as either differentially expressed or non-differentially expressed with similar pattern, I compared the FC values of all the 1243 genes between the two datasets. For this comparison I chose the Bland-Altman-Leh ranking test (Bland and Altman, 1986) that assigns a rank to the FC value of each gene in both datasets and afterwards plots the mean of these ranks against their difference. Next, based on the data distribution, the algorithm calculates the critical value. This analysis showed that *Tc-zen1* targets identified by both programs were assigned comparable \log_2 FC values within the range of ± 1.2719 , but beyond this critical range the *Tc-zen1* target genes were assigned significantly different FC values by the two programs (Fig. 3.3A). In conclusion, the datasets generated by DESeq2 and EBSeq are showing different results for the *Tc-zen1* targets.

Similar pattern was observed for the \log_2 FC values of *Tc-zen2* targets: the two programs assigned comparable results to the genes with the FC values within the range of ± 1.7853 , but beyond this critical range the two programs generated significantly different results (Fig. 3.3B). Overall, the Bland-Altman-Leh test showed that the datasets generated by DESeq2 and EBSeq do not provide the same DE analysis results.

Nonetheless, it has been previously reported that EBSeq does perform more accurately when working with larger number of biological replicates (BRs) (≥ 10), while DESeq2 performs very well even with smaller number of BRs (≤ 3). Moreover, DESeq2 is in general more strict than EBSeq, which calculates P_{adj} values more liberally (e.g.: where DESeq2 would assign higher P_{adj} value, EBSeq would assign a lower one and this gene would pass the chosen cut-off criteria) (Seyednasrollah et al., 2015). Therefore, I finalized the pipeline in the following order:

- A) The quality control of the 75 bp long reads is assessed by FastQC.
- B) The reads are trimmed with Trimmomatic.
- C) The overrepresented sequences of mitochondrial and ribosomal RNA are filtered out with Bowtie2.
- D) The trimmed and filtered reads are mapped to the OGS3 with RSEM.
- E) The raw read count information is directly obtained from the RSEM output files.
- F) The genes with low expression levels (≤ 10 reads in each WT and KD sample) are filtered out in Excel.
- G) The final DE analysis is performed with DESeq2.

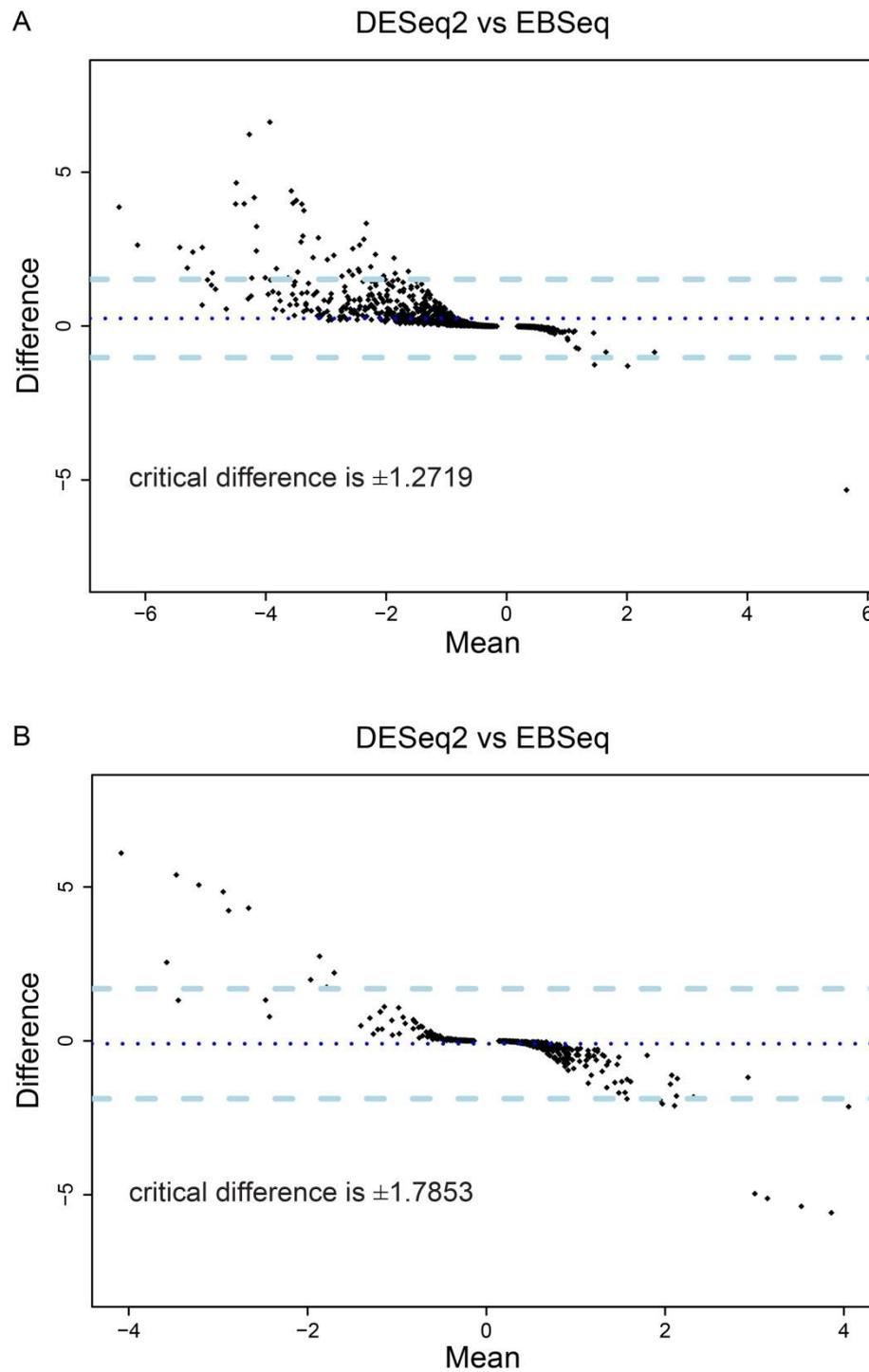


Figure 3.3. Bland-Altman-Leh ranking test of \log_2 fold change values. The \log_2 FC values of *Tc-zen1* (A) and *Tc-zen2* (B) targets were ranked in both DESeq2 and EBSeq datasets. The mean of the ranks was subsequently plotted against the difference in the ranks. The critical difference for the *Tc-zen1* targets is ± 1.2719 and for the *Tc-zen2* targets is ± 1.7853 (light blue dashed lines). Dark blue dotted line represents mean of all the differences between particular ranks.

4 RESULTS

4.1 Hox3 locus sequence conservation

4.1.1 Investigation of conserved non-coding regions between *Tc-zen1* and *Tc-zen2*

It has been previously reported that, although *Tc-zen* genes derived from recent tandem duplication, they acquired distinct functions (van der Zee et al., 2005). *Tc-zen* genes are located in Hox3 locus and are separated by only 172 bp. Therefore, we assumed that in order for them to perform distinct functions in two different developmental stages, a fine-tuned transcriptional regulation, comprising enhancers and insulators, is most likely involved.

In order to investigate whether the transcriptional regulation operates through the same regulatory regions on both *Tc-zen* genes, I compared the complete genomic DNA sequences of *Tc-zen1* and *Tc-zen2*, including 5' and 3' UTRs, and their promoter regions: 200 bp upstream of *Tc-zen1* 5' UTR and 172 bp (the intergenic region) upstream of *Tc-zen2* 5' UTR. To make the *in silico* analysis results more robust, I used two different programs: online based mVista tool (Mayor et al., 2000; Frazer et al., 2004) and standalone Mussa analyzer (<http://mussa.caltech.edu/mussa>).

With the mVista tool a conserved non-coding region was not identified (Fig. 4.1). The only conserved region was identified between the coding sequence of the third exons and it describes conservation levels of the homeobox region (Fig. 4.1, grey rectangle). On the other hand, with the Mussa analyzer, two conserved non-coding regions were identified (Fig. 4.2). One big advantage of the Mussa analyzer is that it considers conserved regions regardless of their position or direction on DNA strand. In fact, by setting the threshold of conservation to 21/30 identities/nt (sliding window), the first conserved region was identified in the reverse-complement direction between the promoter sequence of *Tc-zen1* and the second intron of *Tc-zen2* (Fig. 4.2A, blue line). With the same threshold of conservation, the second conserved non-coding region was identified between the second intron of *Tc-zen1* and the first intron of *Tc-zen2* (Fig. 4.2A, thin red line). Identification of the two conserved non-coding regions was rather surprising and potential transcriptional factor binding sites (TFBS) within these regions should be investigated in the future. However, by increasing the threshold of conservation (from 73% to 100%) only conserved coding region identified is localized within the third exon of both *Tc-zen* genes and spans the homeobox sequence (Fig. 4.2B-F). Absence of highly conserved non-coding regions between the promoter sequences of *Tc-zen* genes implies selective transcriptional regulation of *Tc-zen1* and *Tc-zen2* genes, which could explain, to a certain degree, acquirement of two different functions.

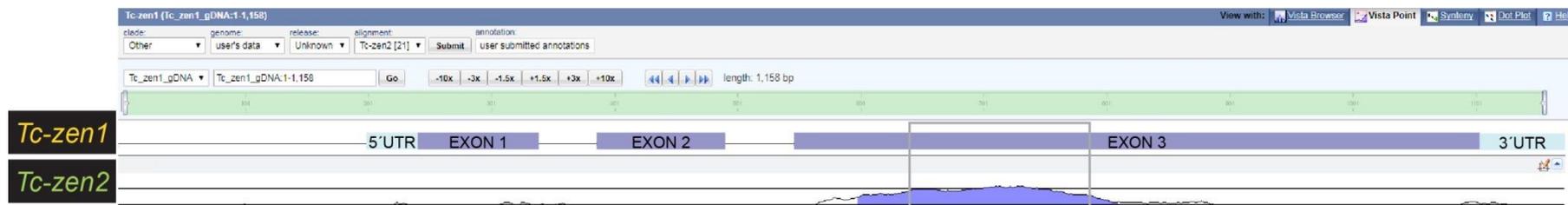


Figure 4.1. *Tc-zen1* and *Tc-zen2* mVista alignment. Using mVista tool, conserved non-coding regions between *Tc-zen1* and *Tc-zen2* genomic DNA were not identified. The only identified conserved region falls into the coding region of the third exon (purple) and shows conservation levels of the homeobox sequences (grey rectangle). The conservation level reaches 75%.

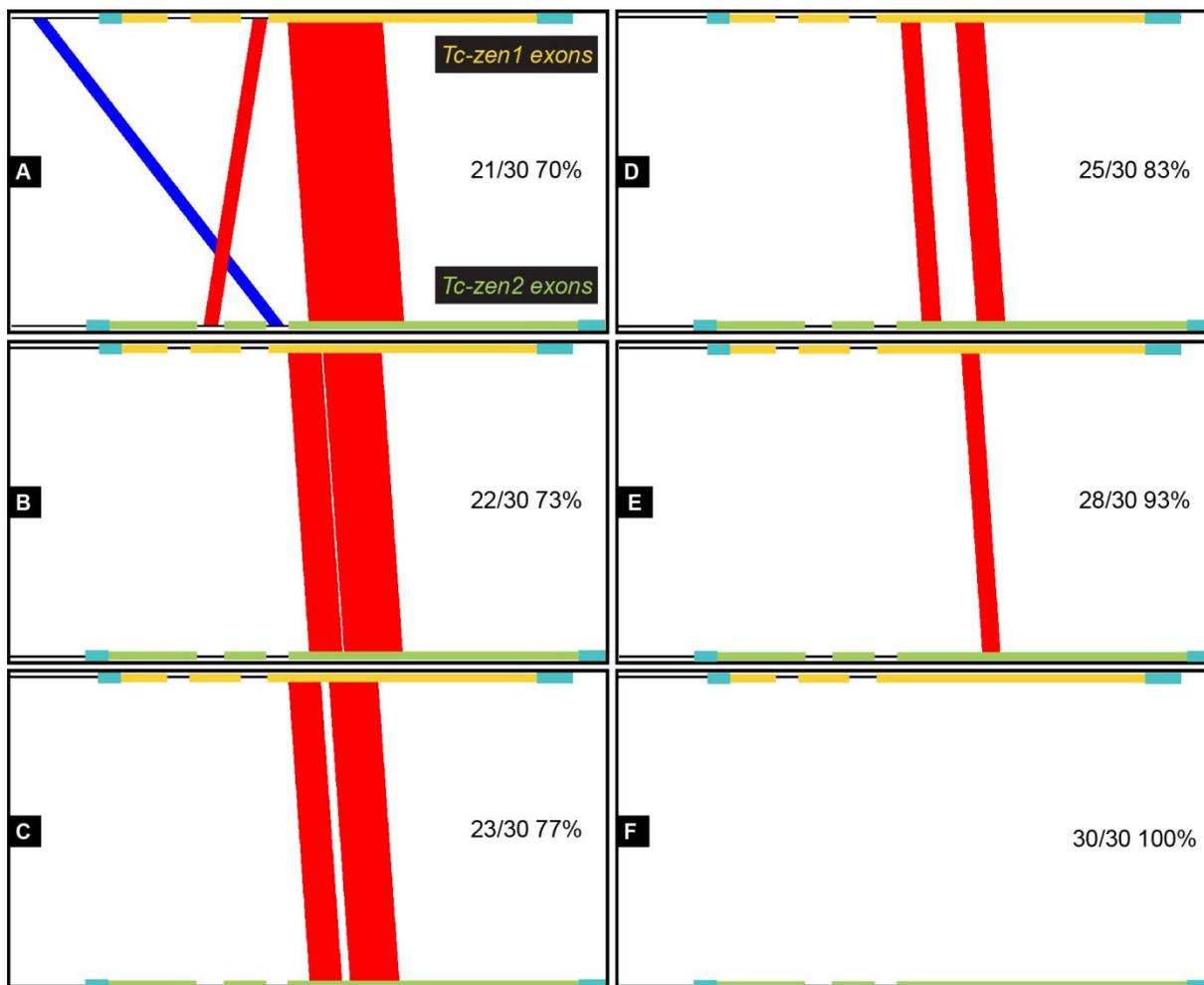


Figure 4.2. *Tc-zen1* and *Tc-zen2* Mussa alignment. Conserved sequences in the same and opposite strand directions between *Tc-zen1* and *Tc-zen2* are displayed with red and blue connecting lines, respectively. Sliding threshold window is defined within each analysis window. Exons of *Tc-zen1* and *Tc-zen2* genes are represented by yellow and green rectangles, respectively. 5' and 3' UTRs are represented by light blue rectangles. The promoter region is represented with a black line located before the 5' UTR of both *Tc-zen* genes. With a 21/30 similarity threshold, two conserved non-coding regions were found: one between the promoter region of *Tc-zen1* and the second intron of *Tc-zen2* and the other between the second *Tc-zen1* intron and the first *Tc-zen2* intron (blue and thin red line, respectively) (A). However, by increasing the similarity threshold, no highly conserved non-coding regions were identified (B-F). High conservation levels with lower thresholds were only identified in the coding region between the third exons of *Tc-zen* genes in the homeobox sequence (A-E).

4.1.2 Investigation of conserved non-coding regions in Hox3 locus of four closely related *Tribolium* species

To identify potential regulatory regions of *Tc-zen1* and *Tc-zen2* located further away from *zen* genes, I performed a large scale comparative *in silico* analysis. I took advantage of the fact that the genomes of four closely related *Tribolium* species (*T. castaneum*, *T. freemani*, *T. madens* and *T. confusum*) are sequenced and compared the regions 5 kb upstream and 5 kb downstream of *zen* genes in the Hox3 loci of all the four species. The *in silico* analysis was again performed with both mVista tool and Mussa analyzer.

Using the mVista tool, I performed comparison of approx. 10-12 kb long regions in Hox3 loci (Fig. 4.3). I identified five conserved non-coding regions: two short regions immediately upstream of the 5' UTR of both *zen1* and *zen2* genes (Fig. 4.3, black rectangles), one upstream in the close proximity of the 5' UTRs of *zen1* gene (Fig. 4.3, green rectangle), one downstream and adjacent to the 3' UTRs of *zen2* genes (Fig. 4.3, yellow rectangle) and the last one located approx. 4 kb downstream of *zen* genes (Fig. 4.3, blue rectangle).

Using the Mussa analyzer, I compared the exact same regions in Hox3 loci of *Tribolium* species as described above. By setting four different thresholds of conservation from 27/30 to 30/30 identities/nt (sliding window), I identified three highly conserved non-coding regions (100% conserved): one upstream of the *zen1* 5' UTR, one downstream of the *zen2* 3' UTR and one approx. 4 kb downstream of *zen* genes (Fig. 4.4D). Additionally, with the lower threshold (90-97% conserved), two further non-coding regions were identified: one immediately upstream of the 5' UTR of *zen1* genes and one in the intergenic region between *zen1* and *zen2* genes (Fig. 4.4A-C, arrows). All the regions identified by Mussa (both highly and less conserved) overlap with the regions identified by mVista tool. Assuming that what is conserved is of functional relevance, regions responsible for transcriptional regulation of *Tc-zen1* and *Tc-zen2* might be located within these identified non-coding sequences. As previously mentioned, a further analysis of TFBS predictions, within the conserved non-coding regions, should be pursued in the future.



Figure 4.3. Conserved non-coding regions in Hox3 loci of four *Tribolium* species identified by mVista tool. The tree is representing the phylogenetic relationship among the four closely related *Tribolium* species used in the analysis (*Tcas* - *T. castaneum*, *Tfre* - *T. freemani*, *Tmad* - *T. madens* and *Tcon* - *T. confusum*). *Tcas* sequence represents the reference sequence, to which the three other species are compared. The conserved non-coding regions are highlighted in pink, the conserved coding regions in purple and conserved UTRs in light blue peaks. Five conserved non-coding regions were identified: two immediately upstream of the 5' UTRs of both *zen* genes (black rectangles), one slightly more upstream of the 5' UTR of *zen1* genes (green rectangle), one slightly more downstream of the 3' UTRs of *zen2* genes (yellow rectangle) and one approx. 4 kb downstream of *zen* genes (blue rectangle). Homeobox sequences of both *zen* genes are highlighted in grey rectangles.

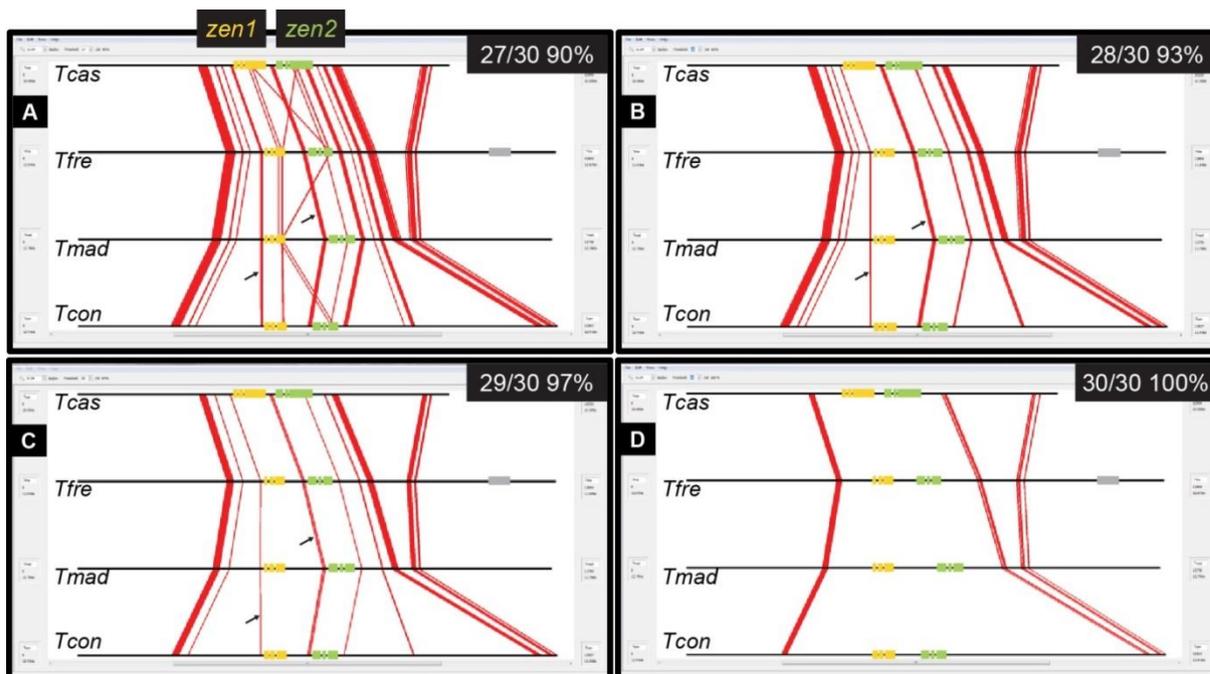


Figure 4.4. Conserved non-coding regions in Hox3 loci of four *Tribolium* species identified by Mussa analyzer. Conserved sequences across the four *Tribolium* species are displayed with red connecting lines. Sliding threshold windows are represented by the black rectangles with the threshold defined inside in white. Exons of *zen1* and *zen2* genes are represented by yellow and green rectangles, respectively. The grey rectangle in the *T. fremani* (*Tfre*) sequence represents stretch of Ns. With the lower threshold of conservation (90-97%), two conserved non-coding regions were identified immediately upstream of the 5' UTR of *zen1* genes and in the intergenic region between *zen1* and *zen2* (A-C, arrows). By increasing the similarity threshold, three different highly conserved non-coding regions were identified: 1 in close proximity to the 5' UTRs of *zen1* genes, 1 in close proximity to the 3' UTRs of *zen2* genes and 1 approx. 4 kb downstream of *zen* genes (D).

4.2 *Tc-zen1* and *Tc-zen2* wild type expression dynamics

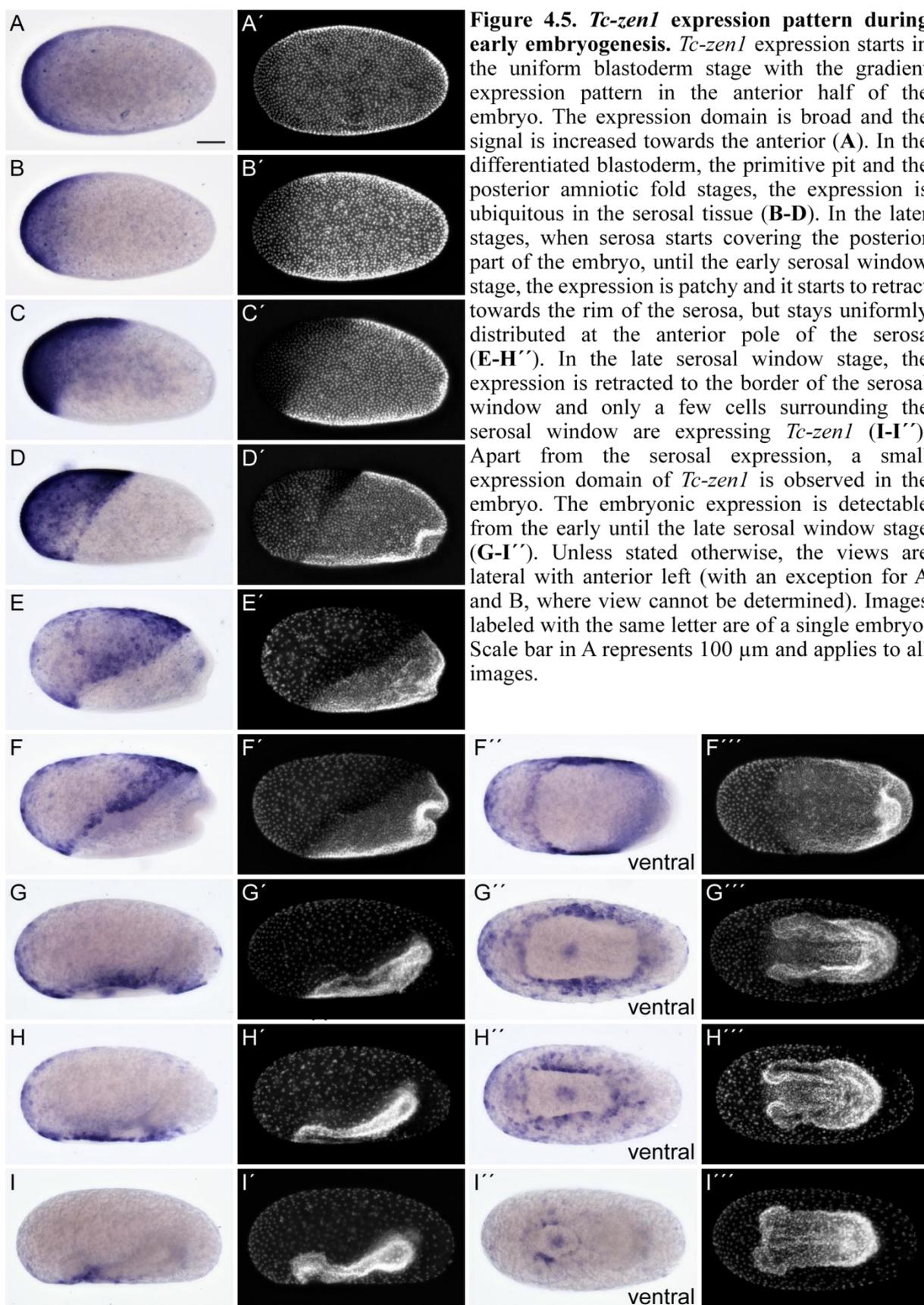
4.2.1 Expression domains of *Tc-zen1* and *Tc-zen2* during early embryogenesis

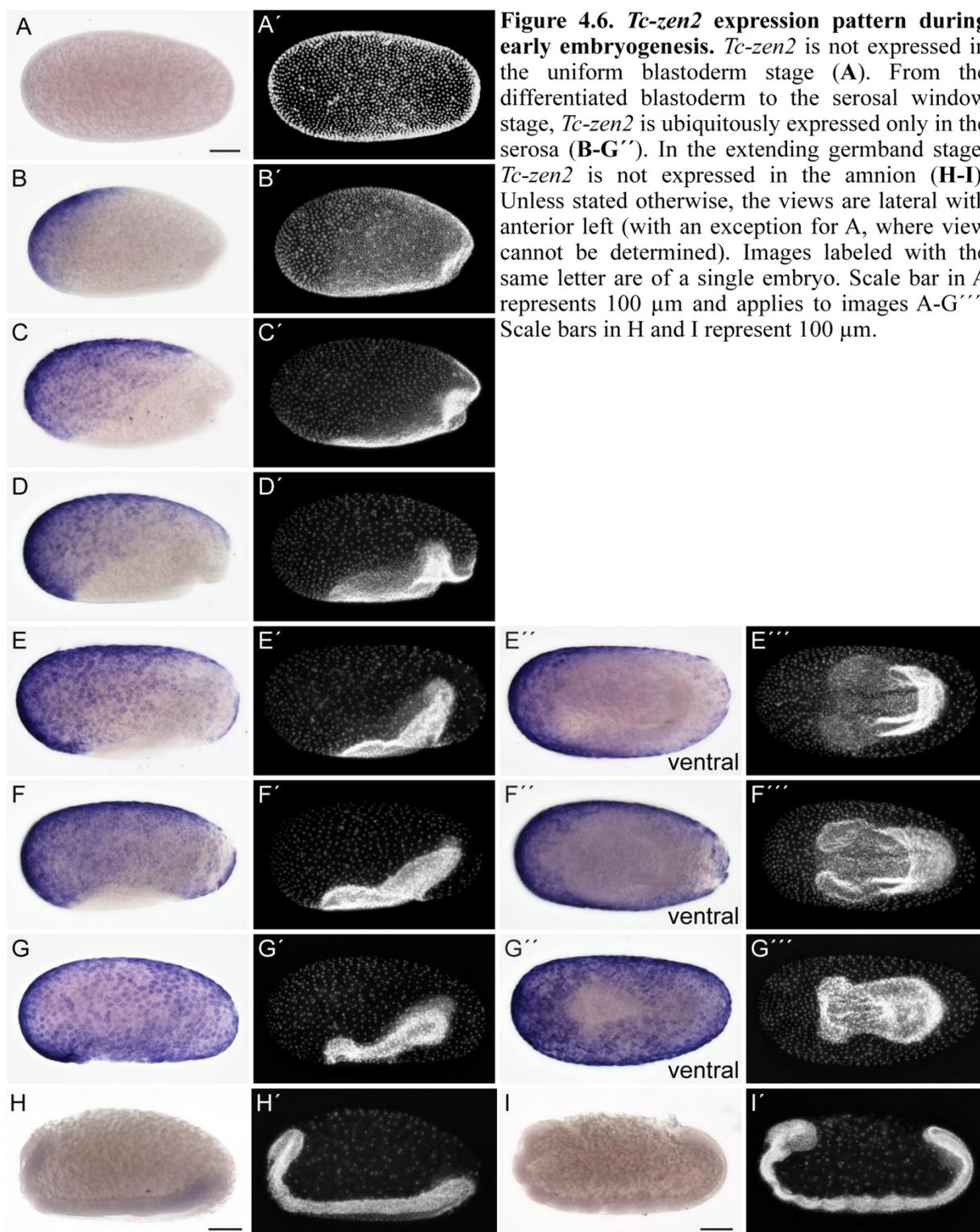
I characterized *Tc-zen1* and *Tc-zen2* expression patterns in wild type (WT) embryos by performing *in situ* hybridization. Carefully staged WT eggs were collected in order that all the important developmental stages were represented. Consistent with their roles in extraembryonic development, both *Tc-zen1* and *Tc-zen2* were exclusively expressed in the serosa, except for a small transient embryonic expression domain of *Tc-zen1* with unknown function (Fig. 4.5 and 4.6).

In the uniform blastoderm stage, *Tc-zen1* is broadly expressed in the anterior half of the embryo with the gradient expression pattern increasing towards the anterior, where serosal tissue identity will be established (Fig. 4.5A). From the differentiated blastoderm stage to the stage of the early posterior amniotic fold, *Tc-zen1* expression pattern is ubiquitous in the serosa (Fig. 4.5B-D). During the developmental process when serosa is expanding towards the posterior, *Tc-zen1* expression is no longer ubiquitous, but rather becomes patchy across the serosa (Fig. 4.5E-F''). In the early serosal window stage, the patchy expression starts to retract to the rim of the serosa, but remains uniform at the anterior pole (Fig. 4.5G-H''). In the late serosal window stage, the expression is retracted to the border of the serosal window and only a handful of cells surrounding the serosal window in the ventral serosa are still expressing *Tc-zen1* (Fig. 4.5I-I''). Additionally, I repeatedly observed a dot-shaped expression domain in embryo visible through the serosal window (Fig. 4.5G-I'').

Likewise, *Tc-zen2* is only expressed in the serosa, however with a slightly different pattern compared to *Tc-zen1*. *Tc-zen2* expression is not observed in the uniform blastoderm stage (Fig. 4.6A) and does not start before the blastoderm becomes differentiated (Fig. 4.6B). Moreover, while *Tc-zen1* is ubiquitously expressed in the serosa only until the stage of the early posterior amniotic fold, *Tc-zen2* retains its ubiquitous expression in the whole serosa through the late serosal window stage (Fig. 4.6B-G'').

It has been previously reported, that, apart from the serosal expression, *Tc-zen2* is expressed also in the anterior-ventral part of the amnion during the extending germband stage (van der Zee et al., 2005). However, no amniotic *Tc-zen2* expression was observed in the extending germband stages in this project (Fig. 4.6H-I).





4.2.2 Expression profile of *Tc-zen1* and *Tc-zen2* transcript during early embryogenesis

To identify developmental stage of *Tc-zen1* and *Tc-zen2* peak expression, I investigated their transcript expression levels throughout early embryogenesis by performing RT-qPCR. To begin with, I investigated the expression in developmental stages from egg lay until the serosal window closure (0-16 h AEL) with 4 h collection intervals. *Tc-zen1* and *Tc-zen2* share overlapping time windows with the expression peak between the end of the uniform and the differentiated blastoderm stages (8-12 h AEL). No maternal expression was observed for either *zen* genes (0-4 h AEL) (Fig. 4.7A).

To describe *Tc-zen1* and *Tc-zen2* expression profile in more detail, I investigated their expression using carefully staged embryos with 2 h collection intervals spanning developmental events from the blastoderm formation to the germband extension (4-20 h AEL). *Tc-zen1* expression starts during the blastoderm formation stage (4-6 h AEL) and peaks in the primitive pit stage (6-10 h AEL). Later on, during developmental processes like blastoderm differentiation, germ anlage condensation, the onset of gastrulation and the serosal window closure (10-16 h AEL), the expression of *Tc-zen1* slowly decreases and abruptly switches off after the serosal window closure (after 16 h AEL) (Fig. 4.7B).

Consistent with its expression pattern in the embryo described above, *Tc-zen2* expression does not start before the primitive pit stage. The peak expression of *Tc-zen2* is during the time window, when numerous important developmental events take place: the germ band condenses, the gastrulation starts and later on, during the germband extension, the serosal window closes (10-14 h AEL). After the serosal window closure, *Tc-zen2* expression gradually wanes (Fig. 4.7B).

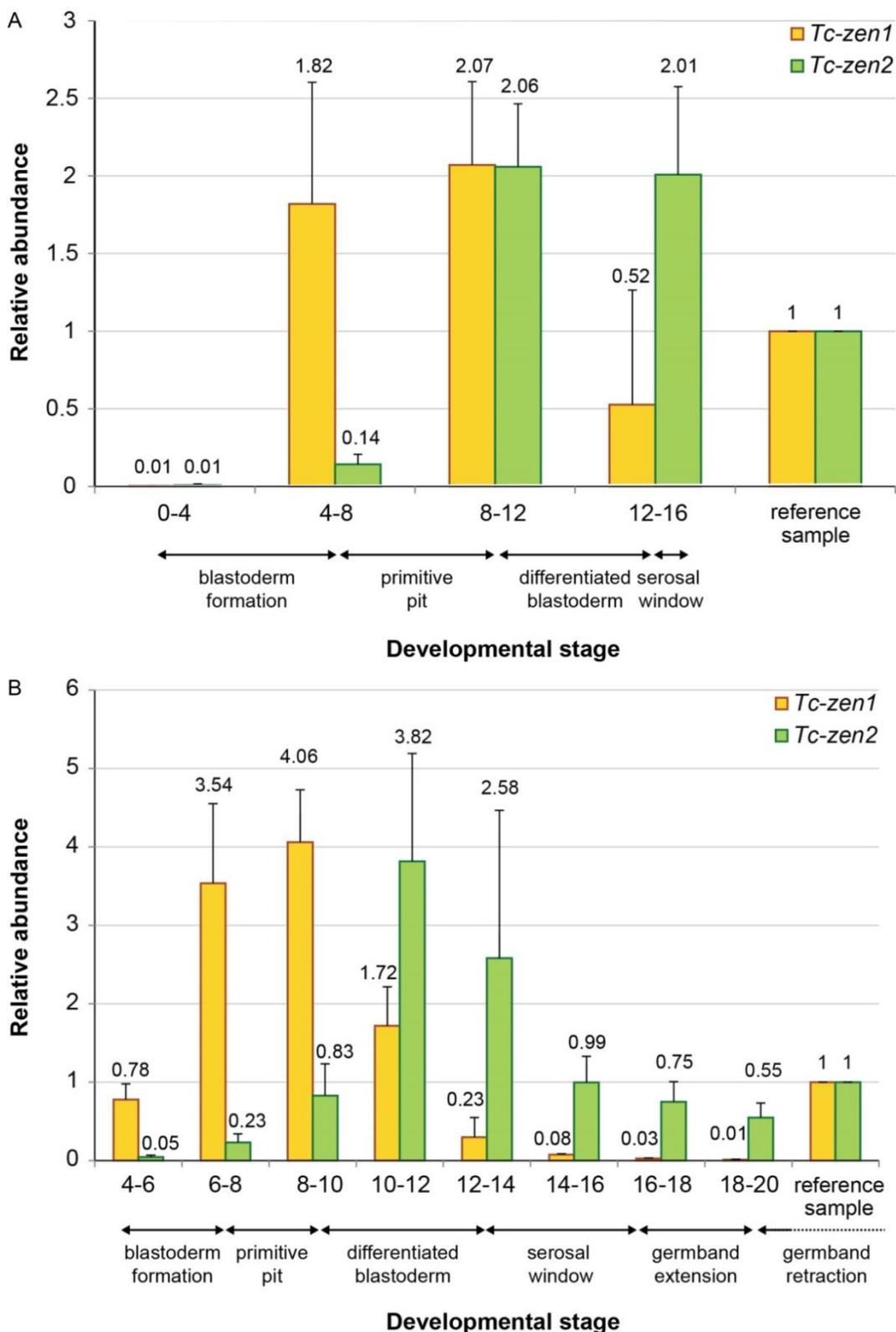


Figure 4.7. Expression profile of *Tc-zen1* and *Tc-zen2* throughout early embryogenesis. Transcript abundance across early embryonic development normalized to the reference gene *Tc-RpS3*. Neither *Tc-zen1* nor *Tc-zen2* are expressed maternally (0-4 h AEL) and they share the expression peak during the differentiated blastoderm stage (8-12 h AEL) (A). Exclusive *Tc-zen1* expression peak occurs in the primitive pit stage (6-10 h AEL) and the expression switches off abruptly after the serosal window stage (16 h AEL). *Tc-zen2* expression peaks between the differentiated blastoderm and the serosal window stage (10-14 h AEL) and wanes gradually (B). Bars indicate mean values; error bars (standard deviation) represent variance across four biological replicates.

4.2.3 Expression profile of *Tc-zen1* and *Tc-zen2* transcript during late embryogenesis

The morphogenesis function of *Tc-zen2* does not manifest before the membrane rupture stage (described later), which takes place 42 h after its expression reaches peak during early embryogenesis (52-56 h AEL). Thus, I investigated *Tc-zen2* transcript expression during late embryogenesis until the membrane rupture stage. For comparison and more reliable interpretation of results, I performed RT-qPCR using the samples from early and late developmental stages.

On one hand, *Tc-zen1* expression is not observed during the late stages. On the other hand, *Tc-zen2* is expressed, but with very low expression levels during the pre-rupture stages (Fig. 4.8, pre-R). This result suggests that after the expression peak during early embryogenesis, *Tc-zen2* retains minimal transcript expression, which persists at least until the membrane rupture stage.

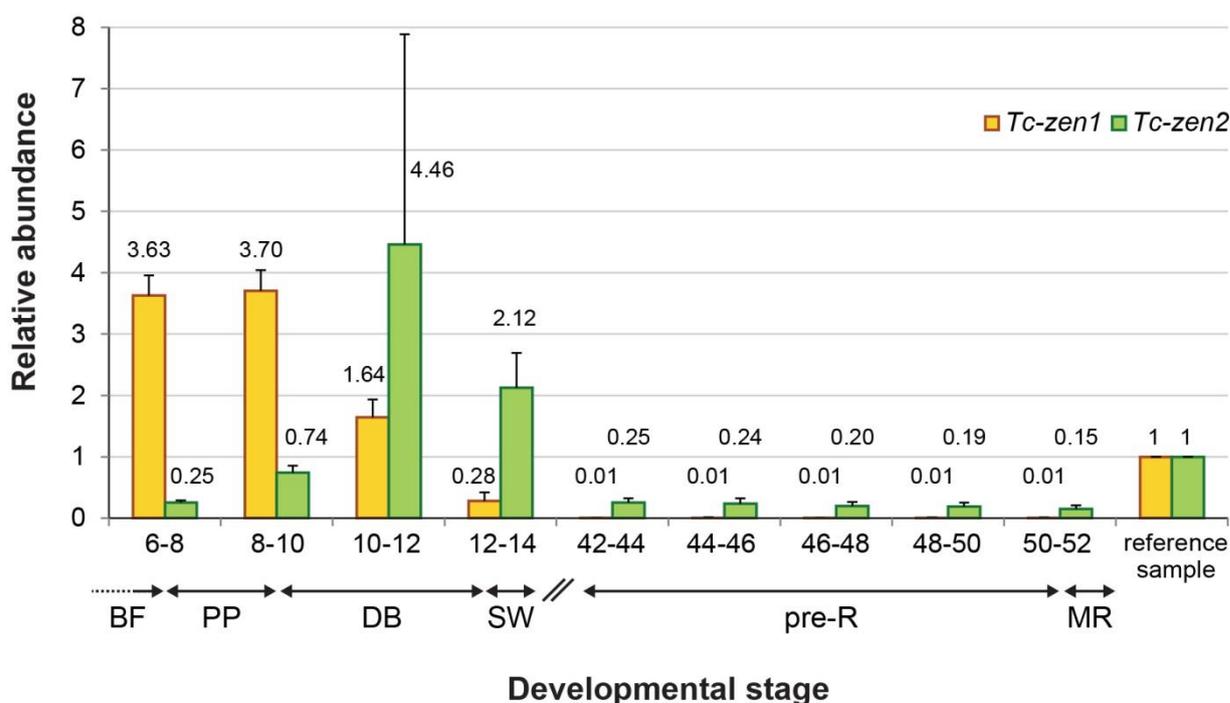


Figure 4.8. Comparison of *Tc-zen1* and *Tc-zen2* expression profile during early and late embryogenesis. Transcript abundance across early and late embryonic development normalized to the reference gene *Tc-RpS3*. During early embryogenesis, the highest *Tc-zen1* expression is in the blastoderm formation and primitive pit stages (6-10 h AEL). Afterwards, the expression decreases and switches off after the serosal window stage. *Tc-zen1* is not expressed during late embryogenesis in the pre-rupture and the membrane rupture stages (42-52 h AEL). *Tc-zen2* expression starts in the blastoderm formation stage and peaks in the differentiated blastoderm and the serosal window stages (10-14 h AEL). *Tc-zen2* expression levels are very low before the membrane rupture stage. Bars indicate mean values; error bars (standard deviation) represent variance across four biological replicates. BF-blastoderm formation; PP-primitive pit; DB-differentiated blastoderm; SW-serosal window; pre-R-pre-rupture; MR-membrane rupture.

4.2.4 Spatial and temporal protein expression profiles of Tc-Zen1 and Tc-Zen2

Apart from the transcript expression profiles, I also investigated Tc-Zen1 and Tc-Zen2 protein expression. Spatial and temporal protein expression was explored by two methods: the expression during early embryogenesis was examined by whole-mount immunostaining of embryos, while the temporal expression was examined by western blots on embryo lysates.

Tc-Zen1 is expressed in the serosa with the same pattern as its mRNA: the expression starts in the undifferentiated blastoderm stage (Fig. 4.9A, E) and the ubiquitous expression pattern in the serosa is retained until the differentiated blastoderm stage (Fig. 4.9B-C, F-G). Later, in the late serosal window stage, Tc-Zen1 expression retracts to the border of the serosal window. Embryonic expression of Tc-Zen1 is not observed, despite the ventral-medial embryonic domain of transcript expression (Fig. 4.9D, H). Neither *Tc-zen2* mRNA nor protein is expressed in the undifferentiated blastoderm stage (Fig. 4.9I, M). Their expression starts from the differentiated blastoderm stage (Fig. 4.9J-K, N-O) and the ubiquitous expression pattern is retained through the serosal window closure (Fig. 4.9L, P).

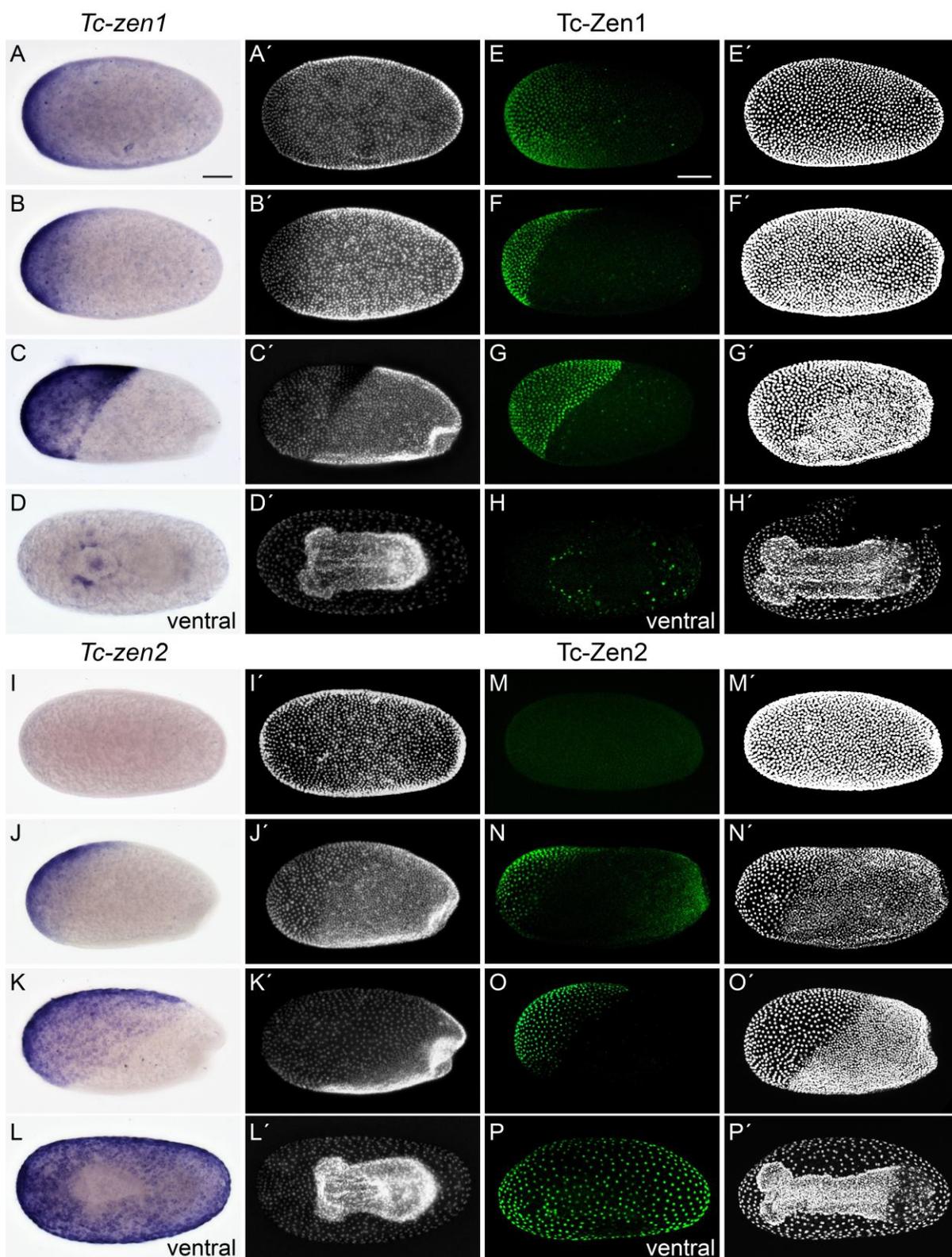


Figure 4.9. Comparison of *Tc-zen1* and *Tc-zen2* mRNA and protein spatial expression patterns. *Tc-Zen1* copies its mRNA expression patterns in the serosa, however no protein expression is observed in the embryo (A-H). On the other hand, *Tc-Zen2* is entirely copying its ubiquitous mRNA expression pattern in the serosa (I-P). The slight offset between *Tc-zen1* and *Tc-zen2* expression with *Tc-zen2* coming on later is observed also in the protein expression patterns (compare A, E to I, M). Unless stated otherwise, the views are lateral with anterior left (with an exception for A, E, B, F, I and M, where view cannot be determined). Images with the same letter are of a single embryo. Scale bar in A and E represents 100 μ m and applies to the ISH (A) and IHC (E) images and their respective DAPI images.

In later embryonic stages (after the serosal window stage) the vitelline membrane is stuck to the serosal cuticle and prevents the antibody penetration. Hence, I investigated temporal expression of Tc-Zen1 and Tc-Zen2 by western blotting, whereas spatial expression during late development was investigated using cryo-sectioned material (see sections 4.2.5 and 4.2.6). Embryonic lysates of the stages spanning from the undifferentiated blastoderm to the membrane rupture stages (4-54 h AEL) were used.

The time course of Tc-Zen1 expression corresponds to the timing of its mRNA expression with a slight temporal shift of two hours (compare Fig. 4.10A with RT-qPCR data in the Fig. 4.7B). Tc-Zen1 protein expression is observed only until the serosal window closure stage (12-14 h AEL). On the other hand, Tc-Zen2 protein expression persists until the membrane rupture stage (Fig. 4.10B-D). Moreover, fluctuations in the Tc-Zen2 expression levels are detected throughout the embryogenesis (from 32-34 to 34-36 h AEL and from 40-42 to 48-50 h AEL), despite the minimal and decreasing *Tc-zen2* mRNA expression after 42 h AEL (compare Fig. 4.10B-D with RT-qPCR data in the Fig. 4.8).

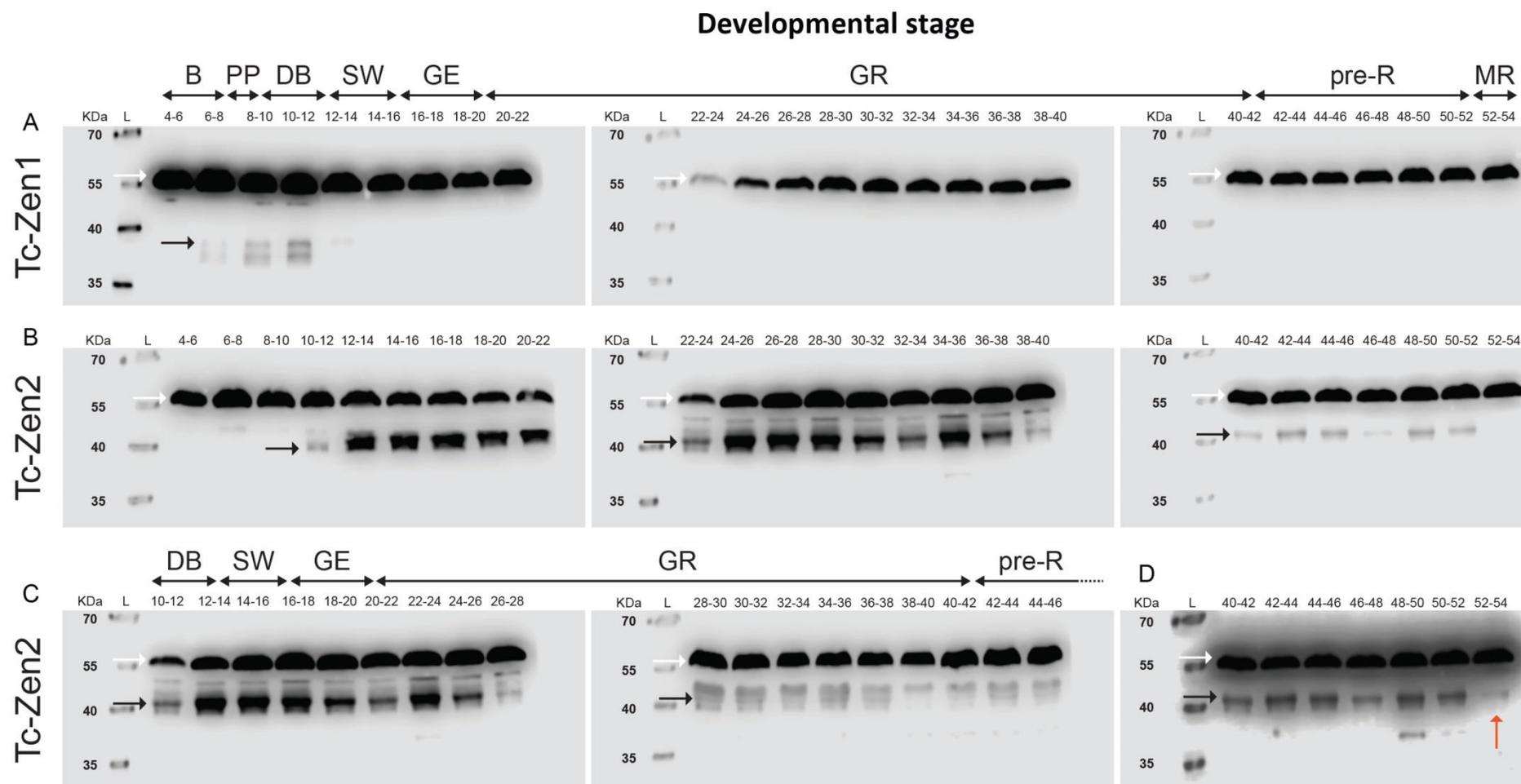
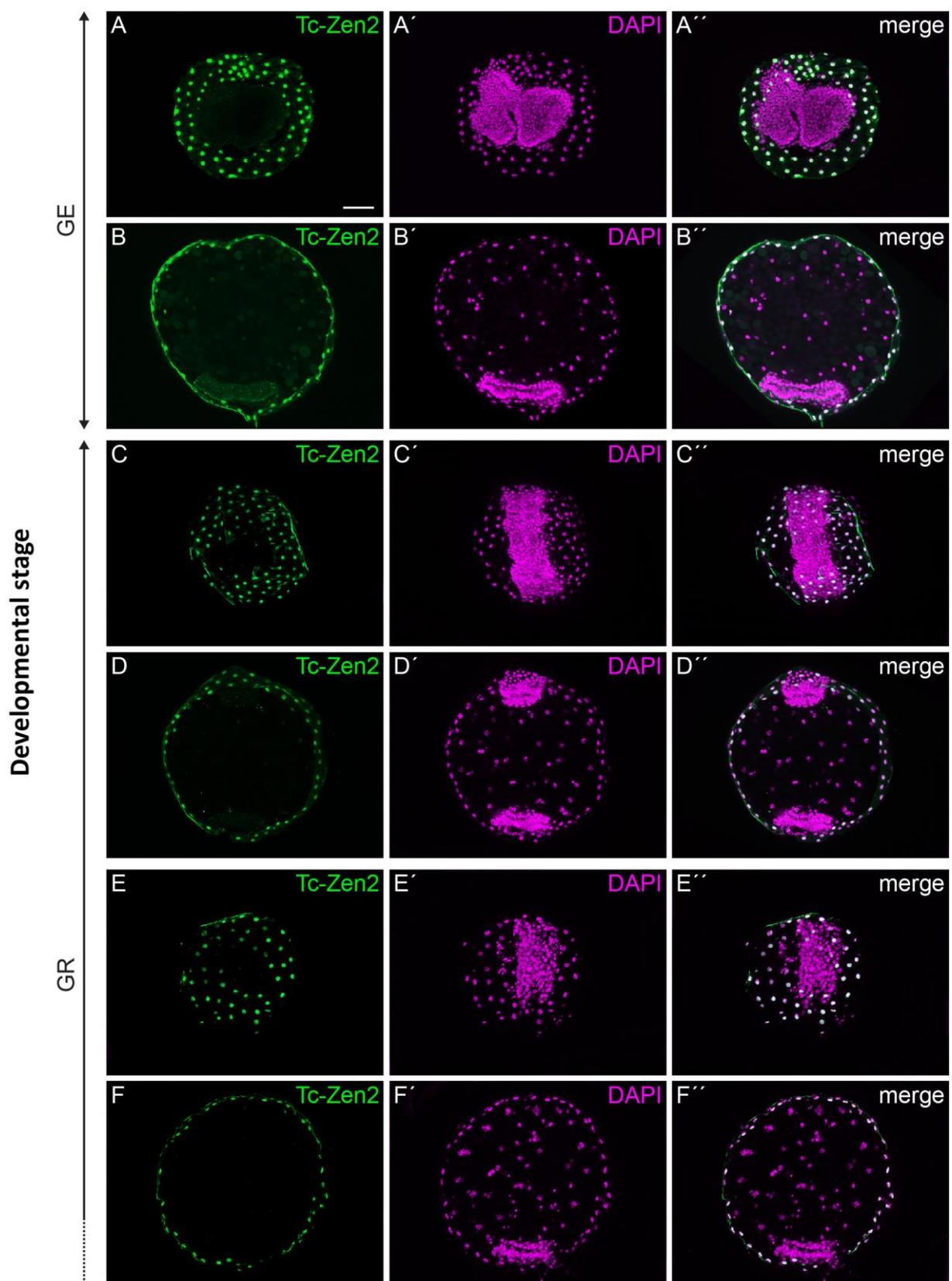


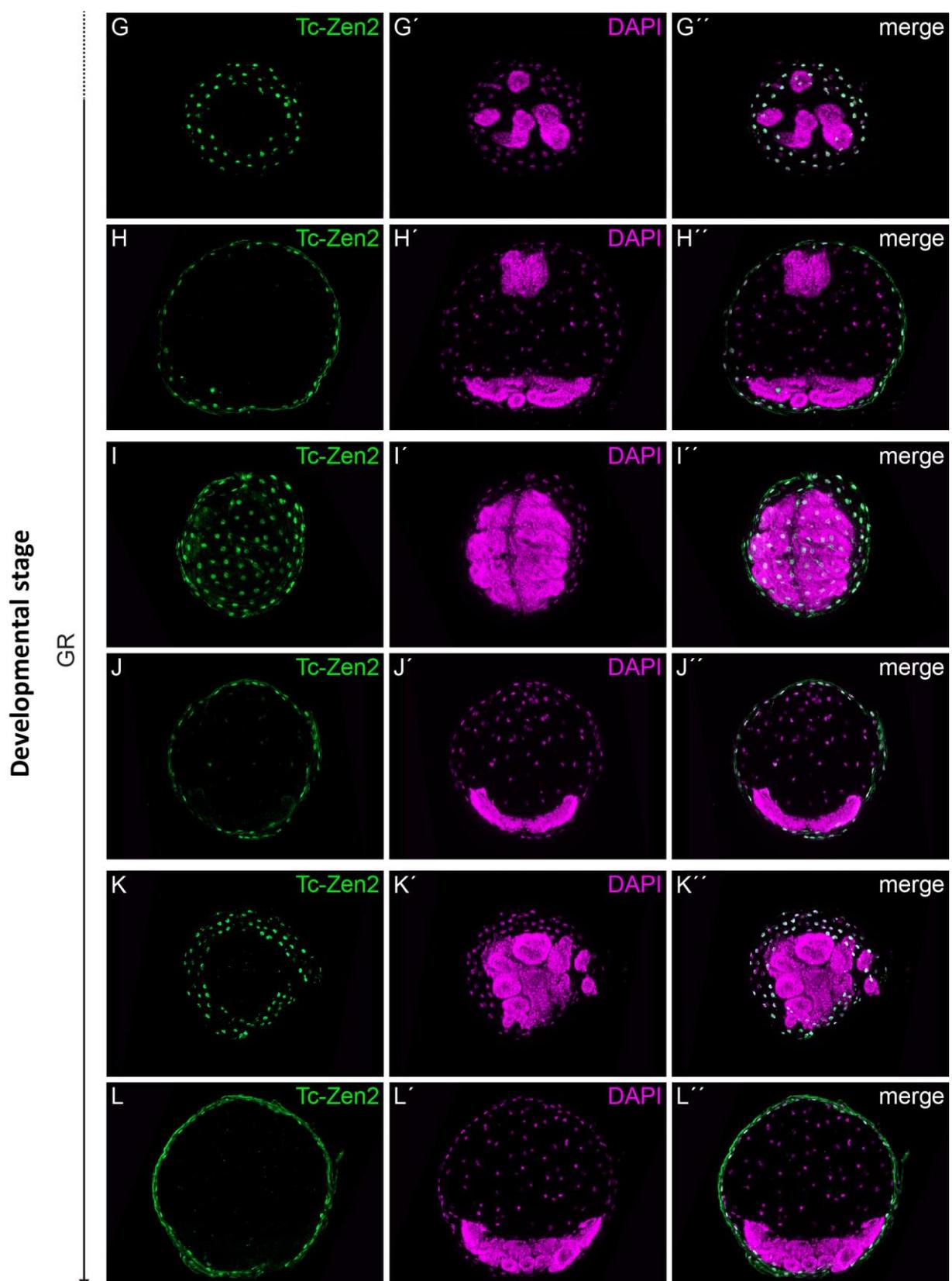
Figure 4.10. Comparison of Tc-Zen1 and Tc-Zen2 temporal expression profiles. Western blots show the temporal expression profiles of Tc-Zen1 (A) and Tc-Zen2 (B-D) from early to late embryogenesis (4-54 h AEL). Tc-Zen1 is expressed only from 6-8 h AEL until 12-14 h AEL (A, black arrow). Tc-Zen2 is expressed from 10-12 h AEL until 54 h AEL, although its expression drops after 52 h AEL (B, black arrow), but it is detectable on higher intensity image (D, red arrow). From 32-34 h AEL to 34-36 h AEL and from 40-42 h AEL to 48-50 h AEL Tc-Zen2 is expressed with fluctuating expression pattern (B, C). White arrow indicates internal loading control (Tc-Tub) present in each well of all western blots. B-blastoderm; PP-primitive pit; DB-differentiated blastoderm; SW-serosal window; GE-germband extension; GR-germband retraction; pre-R-pre-rupture; MR-membrane rupture stage. L-ladder; temporal stages are represented in h AEL.

4.2.5 Cellular localization of Tc-Zen2 transcription factor

As mentioned before (and described later), silencing *Tc-zen2* does not have phenotypic consequences before membrane rupture takes place. However, I have observed Tc-Zen2 protein expression throughout the entire early and late development until this stage (Fig. 4.10B-D). Furthermore, nuclear localization of orthologues of this transcription factor has been shown to be dynamic during embryogenesis in other species, with stages when Zen is cytoplasmic and excluded from the nucleus (Dearden et al., 2000). Therefore, to explore whether Tc-Zen2 changes the localization from cytoplasm to nucleus before its function takes place, I examined the cellular localization of Tc-Zen2 throughout the development until the membrane rupture stage.

From the differentiated blastoderm stage to the serosal window stage, the expression of Tc-Zen2 is specifically localized to the nucleus (Fig. 4.9N-P). Since the vitelline membrane is not permeable to the antibody, localization of Tc-Zen2 in stages after the serosal window stage was investigated on cryo-sectioned material. Because the germband retracts for a rather long period of the developmental time (almost one whole day, approx. 33% of development), 5 different embryos were staged in the consecutive collection intervals (every 4 h) and sectioned in order to cover this long developmental event. Nonetheless, the expression of Tc-Zen2 is only localized to the nucleus also after the serosal window stage and the nuclear localization persists until the pre-rupture stage (Fig. 4.11A-P'). Persistent nuclear localization of Tc-Zen2 might indicate other, so far undescribed functions, taking place from the serosal window closure until the pre-rupture stage.





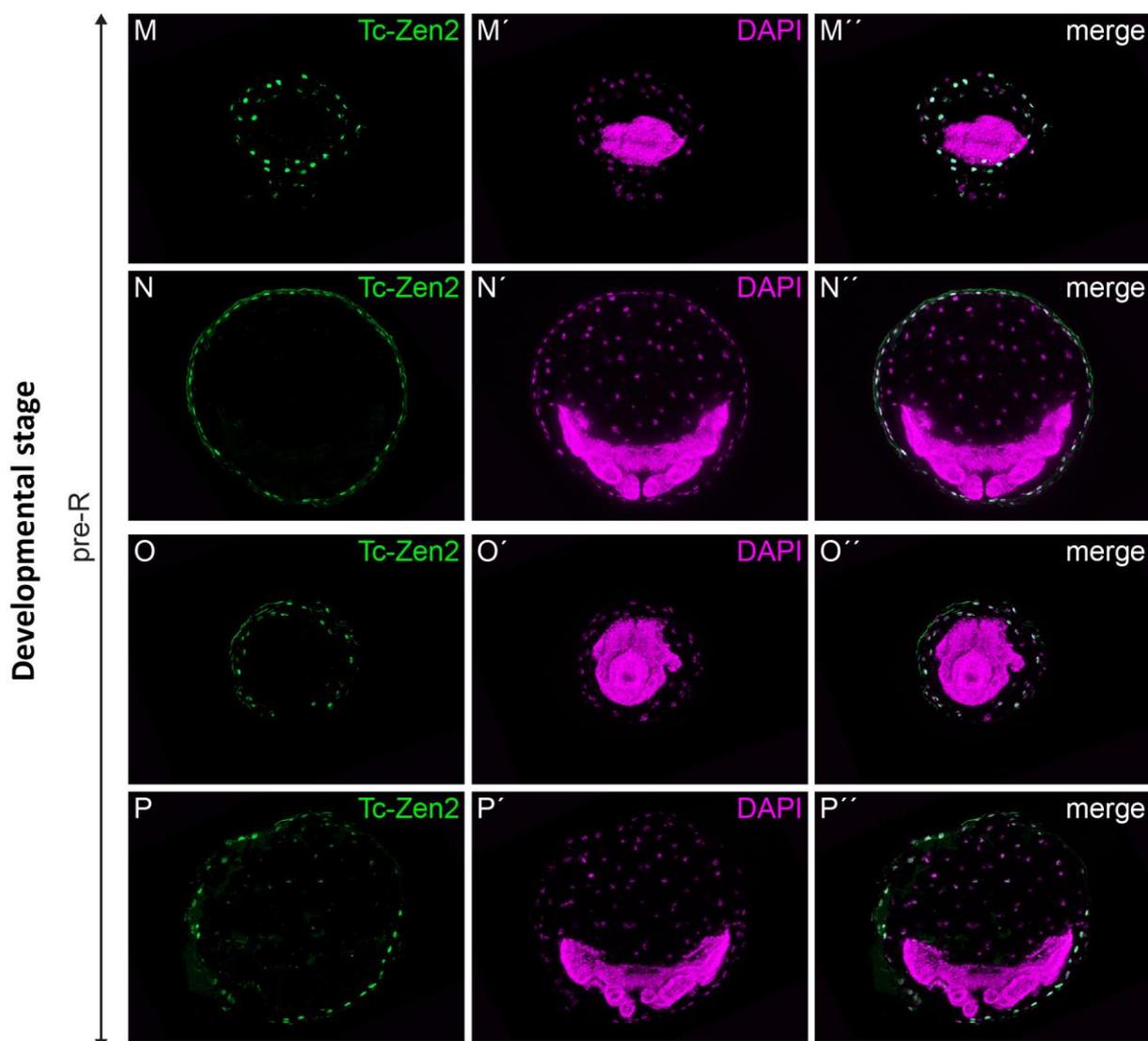


Figure 4.11. Tc-Zen2 expression in cross-sections throughout the embryonic development. Tc-Zen2 antibody staining in the most anterior/posterior and medial cross-sections. Two rows of micrographs represent one embryo. Tc-Zen2 is expressed only in the serosal cells from the germband extension until the pre-rupture stage (A-P''). The expression is localized to the nucleus only. Images with the same letter are of a single embryo. GE-germband extension; GR-germband retraction; pre-R-pre-rupture. Scale bar in A represents 50 μm and applies to all images.

4.2.6 Extraembryonic Tc-Zen2 protein expression

It has been previously suggested that Tc-Zen2 has a function in the amnion during the membrane rupture stage (van der Zee et al., 2005). Therefore, I investigated possible Tc-Zen2 protein expression in the amnion by performing immunostaining on cryo-sectioned material. Since I have shown that Tc-Zen2 expression persists throughout the embryogenesis until the membrane rupture, but it is not clear when exactly its morphogenetic function starts, I explored Tc-Zen2 expression pattern with focus on amnion during several developmental events.

To begin with, I investigated the extended germband stage, where the transcript expression in the anterior-ventral part of the amnion had been previously reported (van der

Zee et al., 2005). On the contrary, Tc-Zen2 is not expressed in any amniotic cell during this stage (Fig. 4.12A-C'). This result is consistent with the one, where no *Tc-zen2* transcript expression has been observed in the amnion during extended germband stages (Fig. 4.6H-I). Next, I performed antibody staining in the germband retraction stage and focused on both the anterior and the posterior part of the amnion. Tc-Zen2 is not expressed in the amnion in this stage either (Fig. 4.12D-F'). Finally, during the pre-rupture stage, I focused on the anterior-ventral part of the amnion, where the membrane rupture takes place. No expression of Tc-Zen2 is observed in this region of the amnion as well as in the posterior one (Fig. 4.12G-I'). This result suggests that Tc-Zen2 expression is exclusive to the serosa throughout the entire lifespan of extraembryonic membranes (EEMs).

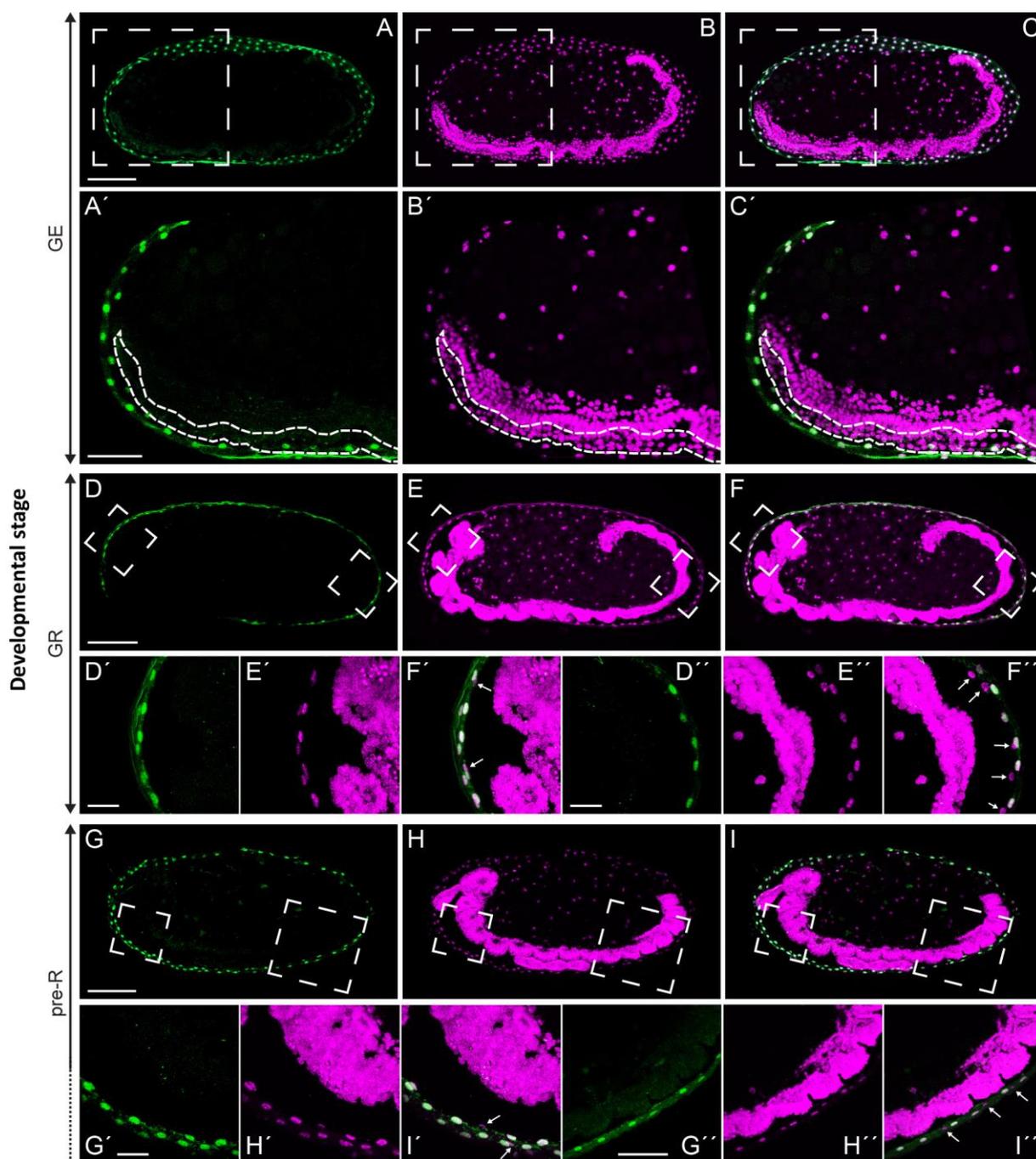


Figure 4.12. Tc-Zen2 is not expressed in the amnion during embryonic development. Tc-Zen2 is expressed only in the serosal cells. Tc-Zen2 is not expressed in the amnion during the extended germband stage (A-C'; A'-C', dashed line). During the germband retraction stage, Tc-Zen2 is not expressed in the anterior (D'-F'; F', arrows), or in the posterior (D''-F''; F'', arrows) part of the amnion. In the pre-rupture stage, Tc-Zen2 is not expressed in the anterior (G'-I'; I', arrows), or in the posterior (G''-I''; I'', arrows) part of the amnion. Green channel represents Tc-Zen2 antibody signal and magenta represents DAPI nuclear signal. GE-germband extension; GR-germband retraction; pre-R-pre-rupture. Scale bar in A, D and G represents 100 μm and applies to A-C, D-F and G-I, respectively. Scale bar in A', G' and G'' represents 50 μm and applies to A'-C', G'-I' and G''-I'', respectively. Scale bar in D' and D'' represents 20 μm and applies to D'-F' and D''-F'', respectively.

4.3 Characterization of *Tc-zen1* and *Tc-zen2* phenotypes after parental RNA interference

4.3.1 Detailed characterization of *Tc-zen1* knockdown

After parental RNA interference (pRNAi), embryos lacking Tc-Zen1 lose completely serosal tissue identity and only small number of cells at the most anterior pole remains widely spaced. However, this remaining tissue is part of the terminal region and despite its morphology (widely spread cells) it is not specified as mature serosal tissue. The rest of the anterior blastoderm cells acquire embryonic fate and contribute to the head region of the embryo proper (van der Zee et al., 2005). Consequently, embryos have considerably bigger head regions compared to the WT embryos (Fig. 4.13A-B). On the other hand, the posterior blastoderm cells contribute to the amniotic tissue, which results in an increased amniotic domain. Due to the fact that the serosa tissue identity is not established, the embryo does not undergo the serosal window closure. The morphogenetic movements are rather taken over by the enlarged amniotic domain. At the posterior pole, amnion performs folding movements similar to those performed by serosa in the WT condition. Nonetheless, the ventral window closure is not reached, since the posterior amniotic fold moves back dorsally as the germband extension continues (Panfilio et al., 2013). Despite the enlarged head region and altered morphogenetic movements, the embryonic development is successfully completed. Under laboratory conditions, larvae hatch and their cuticles do not show any altered phenotype (Fig. 4.13C-D). Since serosa is secreting the serosal cuticle, dechorionated knockdown (KD) embryos are only covered by the vitelline membrane (which along with the chorion in the WT condition, forms a barrier between the embryo and the surrounding environment). Therefore, KD eggs are shiny with a very smooth surface (Fig. 4.13E-F).

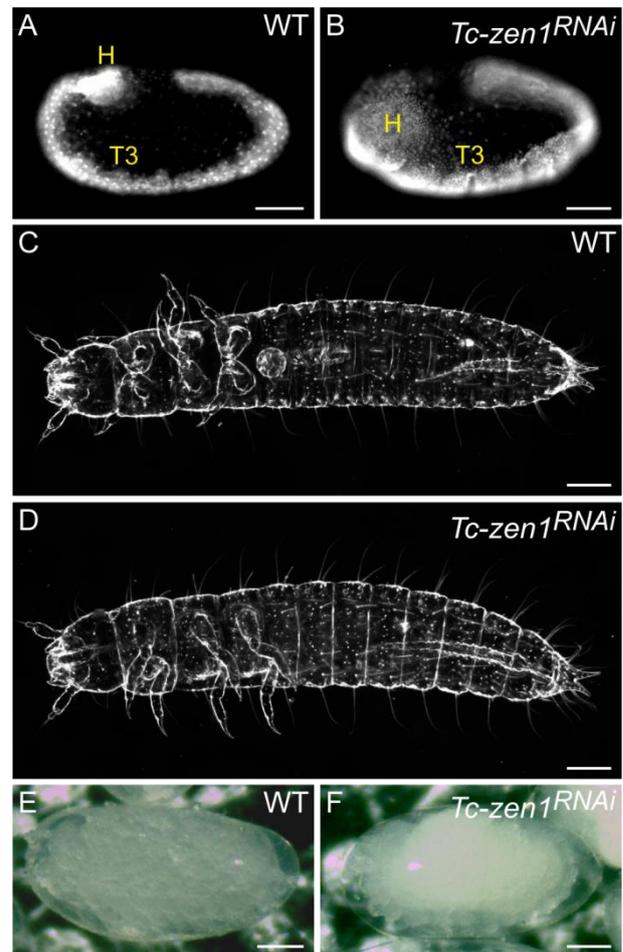


Figure 4.13. *Tc-zen1* knockdown phenotype. Comparison of WT and *Tc-zen1*^{RNAi} phenotype during different developmental stages. During embryogenesis, *Tc-zen1*^{RNAi} embryos display big-headed phenotype, when their head region is considerably larger compared to WT (A, B). *Tc-zen1*^{RNAi} larvae hatch properly and do not show any cuticle phenotype (C, D). WT embryos are covered by the serosal cuticle, which is not see-through, therefore the eggs appear murky (E), while *Tc-zen1*^{RNAi} embryos are only covered by the transparent vitelline membrane and they appear shiny with smooth surface. H-head, T3-third thoracic segment. Scale bars represent 100 μ m.

4.3.2 Knockdown strength and phenotypic penetrance after *Tc-zen1*^{RNAi}

In order to describe the KD strength after *Tc-zen1*^{RNAi}, I investigated *Tc-zen1* KD efficiency by RT-qPCR and by scoring the phenotypic penetrance. In order to make the analysis results more robust, the samples were evaluated in three biological replicates.

The KD efficiency investigation by RT-qPCR was performed with *Tc-zen1*^{RNAi} and corresponding WT samples in the stages of the expression peak of *Tc-zen1* (6-10 h AEL, Fig. 4.7B). The KD of *Tc-zen1* is very strong as well as persistent. Across the three biological replicates, each consisting of several technical replicates, the expression of *Tc-zen1* was reduced to 7-20% of its WT expression (Table 4.1).

As previously mentioned *Tc-zen1* is responsible for the serosa tissue identity determination (van der Zee et al., 2005). Since after *Tc-zen1* KD serosal tissue is not specified, serosal cuticle is not secreted. Therefore, the phenotypic penetrance of *Tc-zen1* was scored on the basis of the embryo being covered by the serosal cuticle or lacking the serosal cuticle (Fig. 4.13E-F). I scored the phenotypic penetrance in three biological replicates. In each replicate, several scoring rounds were performed after different number of days after injection (DAI). No serosal cuticle was observed in *Tc-zen1* KD embryos, therefore the phenotypic penetrance was almost 100% in all the three biological replicates (Table 4.2, Fig. 4.14).

Table 4.1. Knockdown strength after *Tc-zen1*^{RNAi}. The strength of *Tc-zen1*^{RNAi} was verified by RT-qPCR in three biological replicates. The KD strength is represented in the percentages of the WT expression, which is considered to be 100%. n represents number of technical replicates within each biological replicate (BR).

	BR1 [%]	BR2 [%]	BR3 [%]
<i>Tc-zen1</i> ^{RNAi} strength	8-14	7-9	11-20
n [WT]	5	5	4
n [KD]	4	4	5

Table 4.2. Phenotypic penetrance after *Tc-zen1* knockdown. Phenotypic penetrance after *Tc-zen1*^{RNAi} was scored on the basis of embryos being covered by the serosal cuticle (WT) or only by the vitelline membrane (*Tc-zen1*^{RNAi}). Table summarizes the percentages of these phenotypes in three biological replicates (BRs). Ranges represent differences between scoring on different DAI (see also Fig. 4.14)

# BR	WT [%]	<i>Tc-zen1</i> KD [%]
1	100	0.6
2	99-100	1-3
3	100	0.5-6

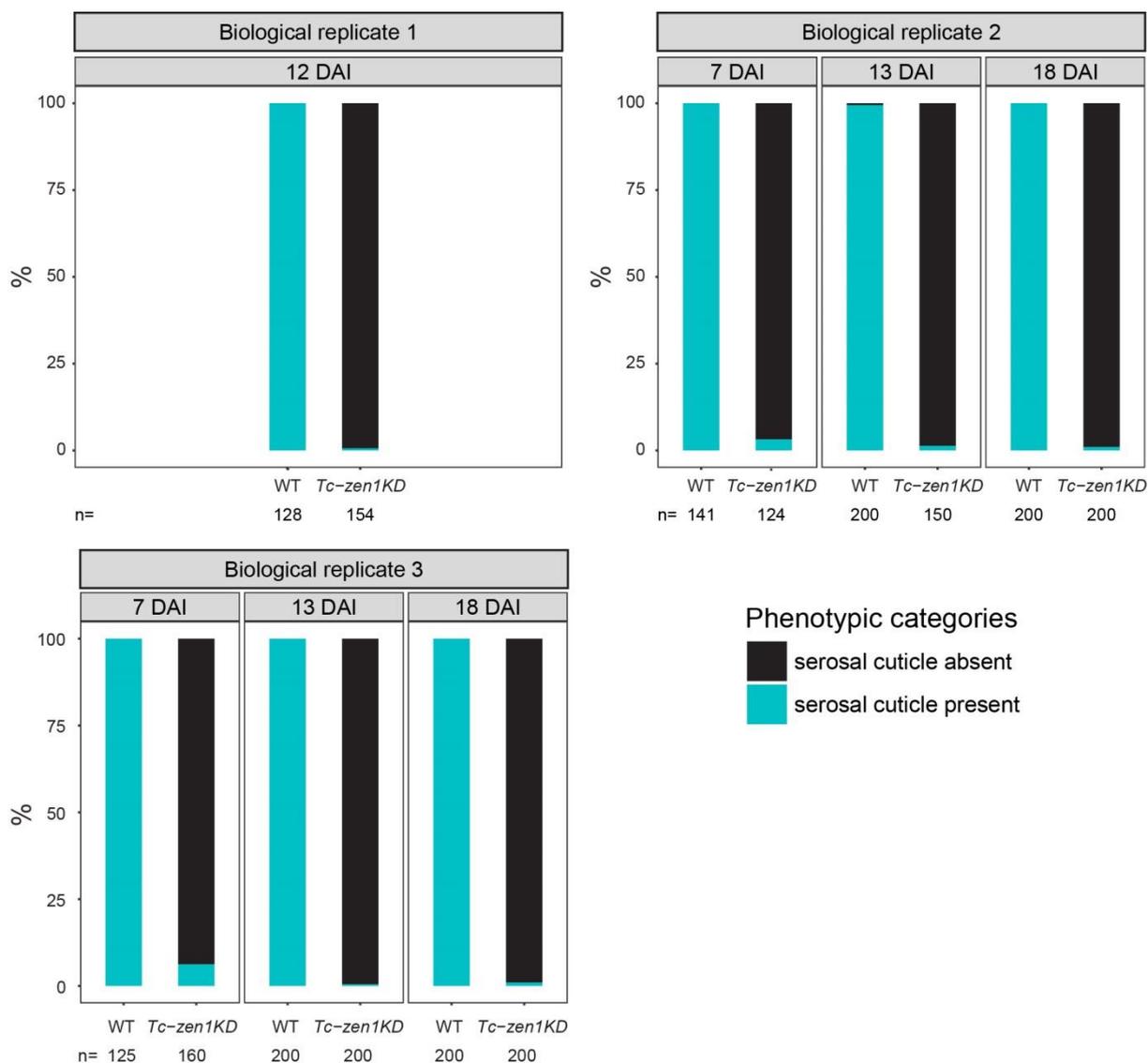


Figure 4.14. Phenotypic penetrance after *Tc-zen1* knockdown. Phenotypic penetrance after *Tc-zen1*^{RNAi} was scored on the basis of the presence or the absence of the serosal cuticle. The scoring was performed after different number of DAI. In all three biological replicates almost 100% of *Tc-zen1*^{RNAi} embryos are lacking serosal cuticle (see also Table 4.2). n represents sample size of each egg collection.

4.3.3 Detailed characterization of *Tc-zen2* knockdown

Tc-zen2 KD does not have any overt phenotypic consequences before the membrane rupture stage. After *Tc-zen2*^{RNAi}, the rupture of EEMs is either completely blocked, or ectopic (van der Zee et al., 2005; Hilbrant et al., 2016), resulting in diverse cuticle defects. Depending on the position of the ectopic rupture, the EEMs withdrawal morphogenetic movements are altered and embryo, or part of the embryo, ends up being inside-out. If the rupture is completely blocked, the whole embryo undergoes ventral closure and folds itself inside-out. Consequently, body structures like bristles, legs and urogomphi are all pointing inwards (Fig. 4.15B).

If the ectopic rupture occurs at the posterior side, the anterior part of the embryo ends up being inside-out, whereas the posterior part remains in the correct orientation (Fig. 4.15C). On the other hand, after the ectopic rupture occurs at the anterior side, the posterior of the embryo folds itself inside-out, whereas the anterior of the embryo remains unfolded (Fig. 4.15D). If the ectopic rupture occurs simultaneously at both the anterior and the posterior side, the embryo does not fold inwards, but rather continues the development in the correct orientation. However, the ectopically ruptured EEMs form a “belt”, which surrounds the embryo (Hilbrant et al., 2016). Therefore, the embryo cannot undergo proper dorsal closure and its dorsal part remains open (Fig. 4.15E). Sometimes the ectopic rupture might occur in different locations. Although the embryo struggles and tries to continue proper development, due to the lack of required morphogenetic movements, it is constrained. Ultimately, after the cuticle preparations, the body structures are difficult to distinguish and identify, and therefore this phenotype is marked as cuticle crumbs (Fig. 4.15F, G).

As the embryonic development after the KD of most of the genes is slowed down, some of the embryos hatch, but the process itself is delayed (Fig. 4.15H). Since the strength of KD is variable across embryos, several embryos are not ostensibly affected by RNAi and the larvae look like WT (Fig. 4.15I). Some of the embryos die during the very early stages, which results in empty egg cuticle phenotype (Fig. 4.15J). The last three categories occur with low frequency after the KD of many genes as well as in WT, and therefore are not regarded as gene-specific phenotypic categories.

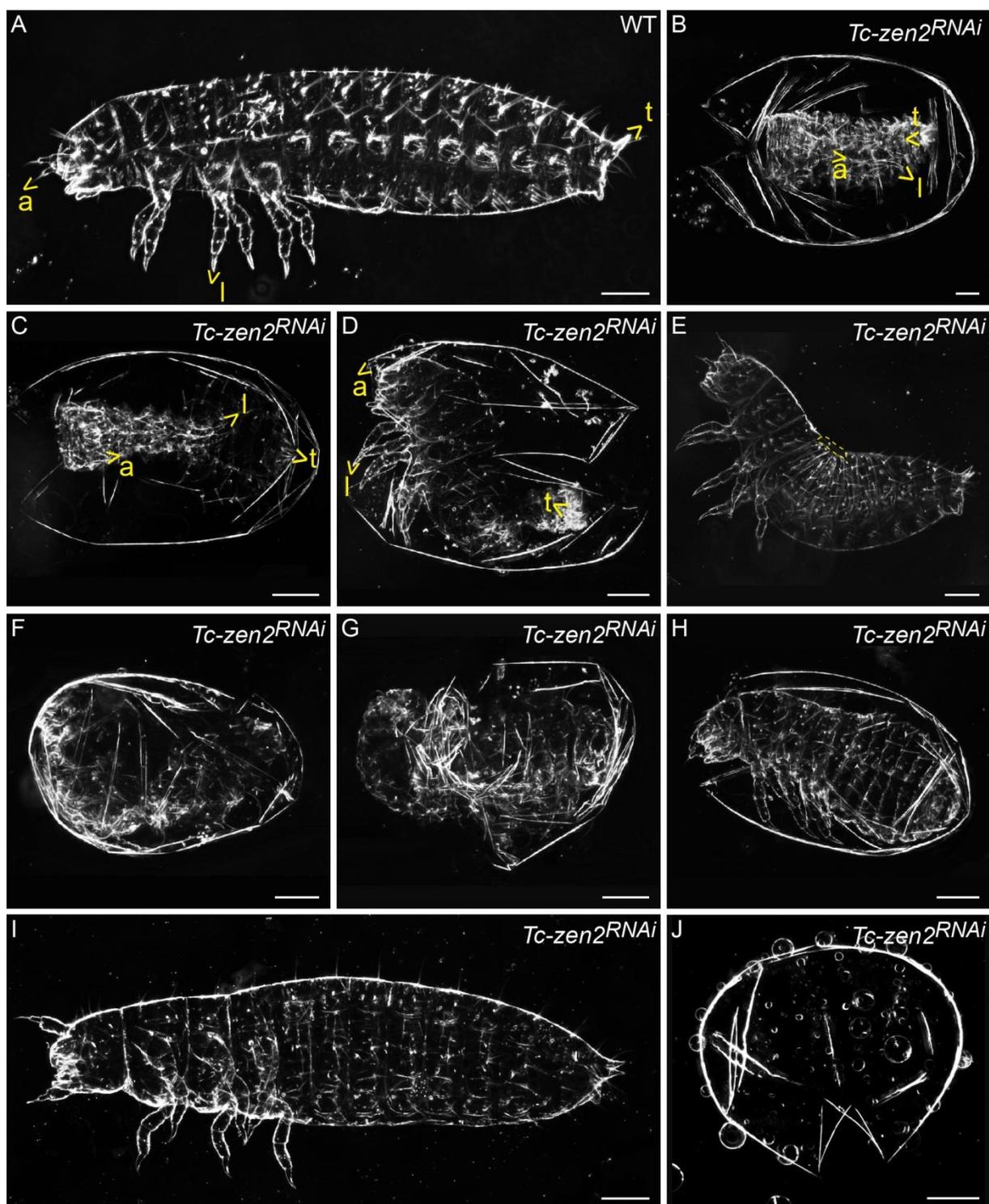


Figure 4.15. *Tc-zen2* knockdown cuticle phenotypic categories. Phenotypic categories based on the cuticle preparations after *Tc-zen2*^{RNAi}: cuticle of the WT hatched larva (**A**); inside-out phenotype (**B**); anterior inside-out (**C**); posterior inside-out (**D**); in the dorsal open phenotype larvae show dorsal hole (**E**, dashed line); cuticle crumbs (**F**, **G**); unhatched (**H**); WT-like (**I**); empty egg (**J**). a-antenna, l-leg, t-tail. Scale bars represent 100 μ m.

4.3.4 Knockdown strength and phenotypic penetrance after *Tc-zen2*^{RNAi}

I investigated the KD efficiency of *Tc-zen2*^{RNAi} by RT-qPCR and phenotypic penetrance scoring. In order to make results of the analysis more robust, the samples were evaluated in three biological replicates. In addition, the reduction of Tc-Zen2 protein expression after *Tc-zen2*^{RNAi} was evaluated by western blots.

The KD efficiency investigation by RT-qPCR was performed with *Tc-zen2*^{RNAi} and corresponding WT samples in the early embryonic stages, in which the expression of *Tc-zen2* was detected (8-24 h AEL, Fig. 4.7B). Compared to the *Tc-zen1* KD strength, *Tc-zen2* KD is less efficient. Across the three biological replicates, each consisting of several technical replicates, *Tc-zen2* expression was reduced to 12-57% of its WT expression (Table 4.3).

Phenotypic penetrance of *Tc-zen2*^{RNAi} was scored on the basis of the cuticle phenotypes (Fig. 4.15). Similarly to the *Tc-zen1* phenotypic penetrance scoring procedure, I scored the cuticle phenotypes after different number of DAI. For simplicity, only the data from two to three DAI per biological replicate are shown. Out of all the cuticle phenotypes observed, 74-92% falls into the gene-specific cuticle phenotypic categories (Fig. 4.16). Data from the 24th DAI in the first biological replicate were excluded from the overall calculation, since after three weeks after pRNAi, the phenotypic penetrance significantly decreases. The data from this day are only shown as an example of this phenomenon. The most abundant gene-specific KD phenotypic categories represented are the anterior inside-out and the complete inside-out (Fig. 4.16, Table 4.4).

Table 4.3. Knockdown strength after *Tc-zen2*^{RNAi}.

The strength of *Tc-zen2*^{RNAi} was verified by RT-qPCR in three biological replicates. The KD strength is represented in the percentages of the WT expression, which is considered to be 100%. n represents number of technical replicates within each biological replicate (BR).

	BR1 [%]	BR2 [%]	BR3 [%]
<i>Tc-zen2</i> ^{RNAi} strength	12-38	14-30	18-57
n [WT]	12	7	6
n [KD]	12	7	6

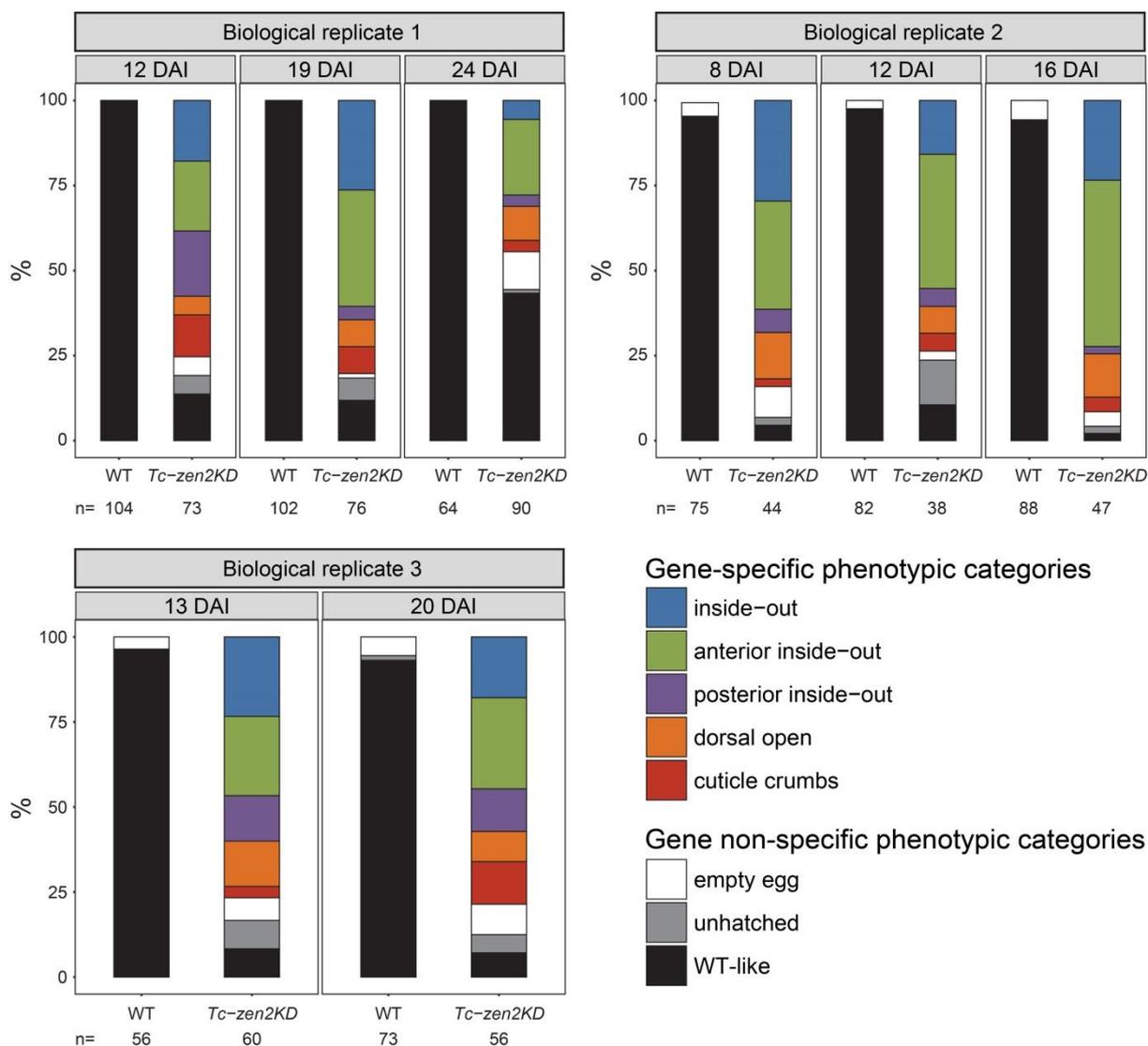


Figure 4.16. Phenotypic penetrance after *Tc-zen2* knockdown. Phenotypic penetrance was scored on the basis of the cuticle phenotypes described in Figure 4.15. The scoring was performed after different number of DAI. The most abundant gene specific phenotypic category is the anterior inside-out, which represents at least 20-48% of all the phenotypes. The most abundant gene non-specific phenotypic category is unhatched, which reached up to 13% (see also Table 4.4). n represents sample size of each collection.

Table 4.4. Phenotypic penetrance after *Tc-zen2* knockdown. Phenotypic penetrance after *Tc-zen2*^{RNAi} was scored on the basis of the cuticle preparations of the KD embryos described in the Figure 4.15. Table summarizes the occurrence and the percentages of different phenotypic categories in three biological replicates (BRs). The categories inside-out, anterior inside-out, posterior inside-out, dorsal open and cuticle crumbs are gene-specific phenotypes and collectively represent KD penetrance. The categories, empty egg, unhatched and WT-like are gene non-specific categories and occur with small frequency also in WT samples. Ranges represent differences between scoring on different DAI (see also Fig. 4.16).

# BR	WT [%]	<i>Tc-zen2KD</i> [%]							
		inside-out	anterior inside-out	posterior inside-out	dorsal open	cuticle crumbs	empty egg	unhatched	WT-like
1	100	5-26	20-34	3-19	5-10	3-12	1-11	1-6	11-43
2	94-97	15-29	31-48	2-6	7-13	2-5	2-9	2-13	2-10
3	93-96	17-23	23-26	12-13	9-13	3-12	6-9	5-8	7-8

Tc-zen2 morphogenesis function manifests before the membrane rupture stage, which takes place two days after *Tc-zen2* reaches its expression peak during early embryogenesis. Due to the fact that *Tc-zen2* transcript levels during the late developmental stage, where the morphogenesis function takes place (48-56 h AEL), are very low (Fig. 4.8), it is not possible to evaluate KD efficiency during the late development by RT-qPCR. Therefore, I decided to investigate the presence of Tc-Zen2 protein after *Tc-zen2*^{RNAi} during the late developmental stages by western blots. Tc-Zen2 expression is successfully reduced in early developmental stages, in which the transcript is expressed (8-24 h AEL, see also Fig. 4.7B) and the silencing effect persists also after one day (48-52 and 52-56 h AEL) (Fig. 4.17).

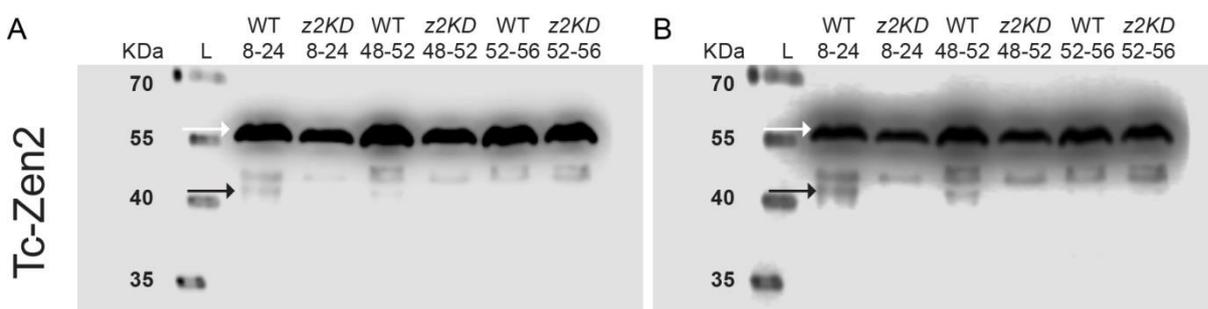


Figure 4.17. Tc-Zen2 expression in wild type and *Tc-zen2* knockdown samples. The silencing effect of *Tc-zen2*^{RNAi} verified in the late stages by western blot. The expression of Tc-Zen2 is silenced after the *Tc-zen2*^{RNAi} in the stages from 8-24 h AEL and the silencing effect persists until the pre-rupture (48-52 h AEL) and the rupture/post-rupture (52-56 h AEL) stages (A, black arrow). Tc-Zen2 expression in the WT post-rupture stage (52-56 h AEL) is only detectable with the higher intensities (B, black arrow). White arrow indicates internal loading control (Tc-Tub) present in each well of the western blot. Temporal stages are represented in h AEL.

4.3.5 Potential regulatory interactions between *Tc-zen1* and *Tc-zen2*

To get insight into the possible regulatory interactions between *Tc-zen1* and *Tc-zen2*, I examined the expression levels of *Tc-zen* genes in the KD samples of its respective paralogue by RT-qPCR (*Tc-zen1* KD samples: 6-10 h AEL, *Tc-zen2* KD samples 10-14 h AEL). After KD of *Tc-zen1*, the expression of *Tc-zen2* decreases 4-fold compared to its WT expression (Fig. 4.18A), consistent with loss of the presumptive serosal tissue domain. On the other hand, after *Tc-zen2* KD, the expression of *Tc-zen1* increases on average 2-fold of its WT expression (Fig. 4.18B). These results suggest that both genes are mutual downstream targets.

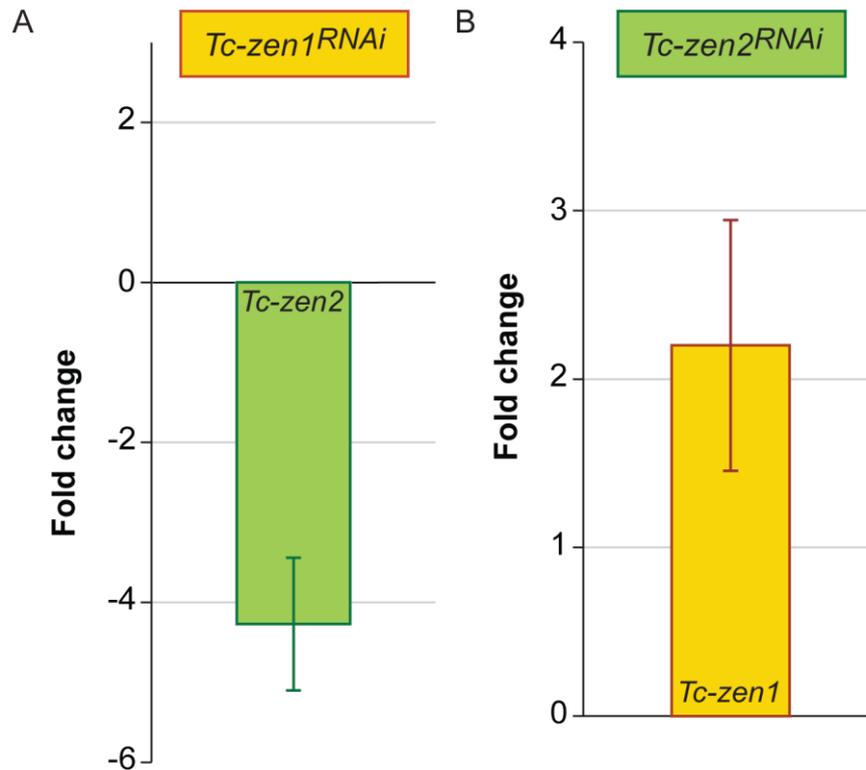


Figure 4.18. Expression levels of each *Tc-zen* gene in the knockdown samples of its paralogue. Expression of *Tc-zen2* in the *Tc-zen1* KD samples decreases 4-fold compared to its WT expression (A). Expression of *Tc-zen1* in the *Tc-zen2* KD samples increases more than 2-fold compared to its WT expression (B). Bars indicate mean values; error bars (standard deviation) represent variance across three biological replicates.

To further investigate the latter regulatory interactions, I have examined the expression pattern of *Tc-zen1* in *Tc-zen2*^{RNAi} embryos. In the early and the late differentiated blastoderm stages, *Tc-zen1* is ubiquitously expressed in the serosa, however the expression levels appear to be much higher after *Tc-zen2* KD than in the WT embryos (Fig. 4.19, compare A and D, and B and E). Furthermore, in the late serosal window stage, *Tc-zen1* expression domain expands. The expression is no longer retracted to the border of the serosal window, but remains throughout the whole serosa (Fig. 4.19, compare C and F). This pattern resembles the expression pattern of *Tc-zen2* at the same embryonic stage (Fig. 4.6G'). This result implicates that *Tc-zen2* ubiquitous expression in the serosal window stage represses *Tc-zen1* expression and restricts its expression only to the serosal window border.

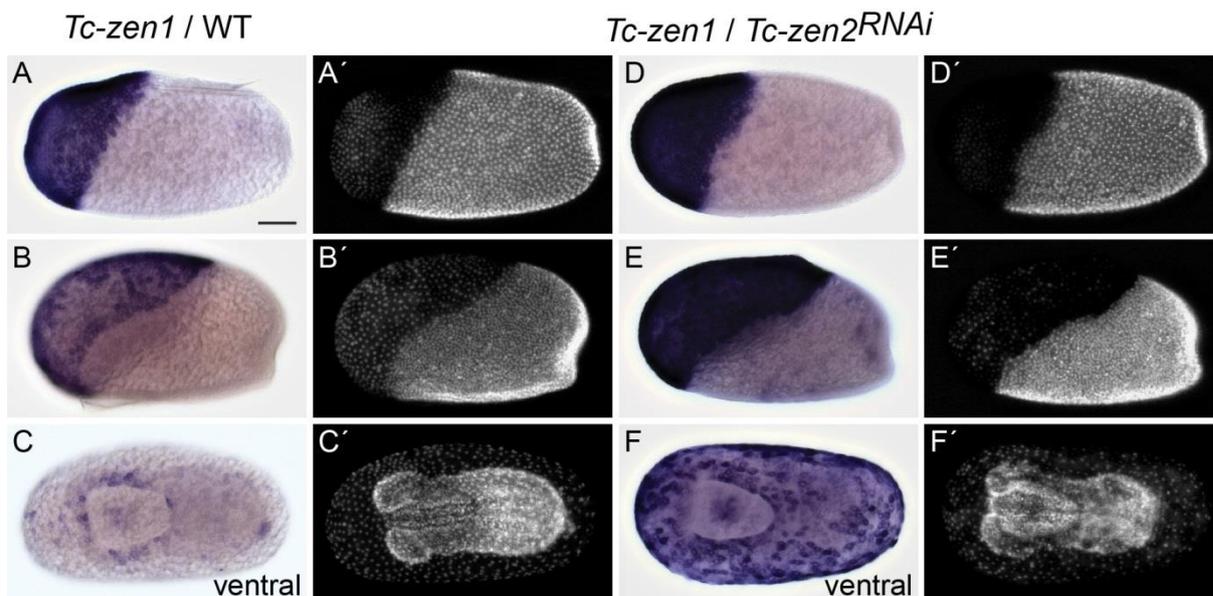


Figure 4.19. Comparison of *Tc-zen1* expression in the wild type and *Tc-zen2*^{RNAi} embryos. *Tc-zen1* expression in the WT embryos is ubiquitous in the serosa in the early differentiated blastoderm stage (A). In the late differentiated blastoderm stage, the expression becomes patchy across the serosa (B). In the late serosal window stage, *Tc-zen1* expression retracts to the border of the serosal window (C). During the early and the late differentiated blastoderm stages, *Tc-zen1* expression levels in *Tc-zen2*^{RNAi} embryos appear to be much higher compared to WT (D-E). In the late serosal window stage, the expression of *Tc-zen1* remains in the whole serosa (F). Unless stated otherwise, the views are lateral with anterior left. Images with the same letter are of a single embryo. Scale bar in A represents 100 μ m and applies to all images.

4.3.6 Possible off target knockdown effects of *Tc-zen1* long dsRNA fragment on *Tc-zen2* expression

Tc-zen1 and *Tc-zen2* are the result of a recent gene duplication and share, already on a nucleotide level, a highly conserved region within the coding sequence (Fig. 4.20). Therefore, I investigated the possible off target effect of the long *Tc-zen1* dsRNA fragment, which spans this conserved region (Fig. 4.20, yellow arrows: FW-REV-long), on *Tc-zen2* expression by RT-qPCR. It has been previously reported that, in *T. castaneum*, the length of dsRNA fragment could affect the KD efficiency of the gene, as the usage of longer dsRNA fragment of several genes resulted in stronger KD phenotypes (Wang et al., 2013). Therefore, in order to exclude that *Tc-zen2* expression levels could decrease after *Tc-zen1*^{RNAi} performed with the long *Tc-zen1* dsRNA fragment due to the higher *Tc-zen1* KD efficiency, I additionally compared *Tc-zen1* KD efficiency after RNAi performed with both short (Fig. 4.20, yellow arrows: FW-REV-short) and the long dsRNA fragment. Expression levels of *Tc-zen1* and *Tc-zen2* were compared between the *Tc-zen1* WT expression peak (6-10 h AEL) and *Tc-zen1*^{RNAi} samples in three biological replicates after the RNAi with both the short and the long *Tc-zen1* dsRNA fragments. Each biological replicate contained several technical replicates and the mean value was calculated. The percentage representation of KD efficiency is shown (Fig. 4.21).

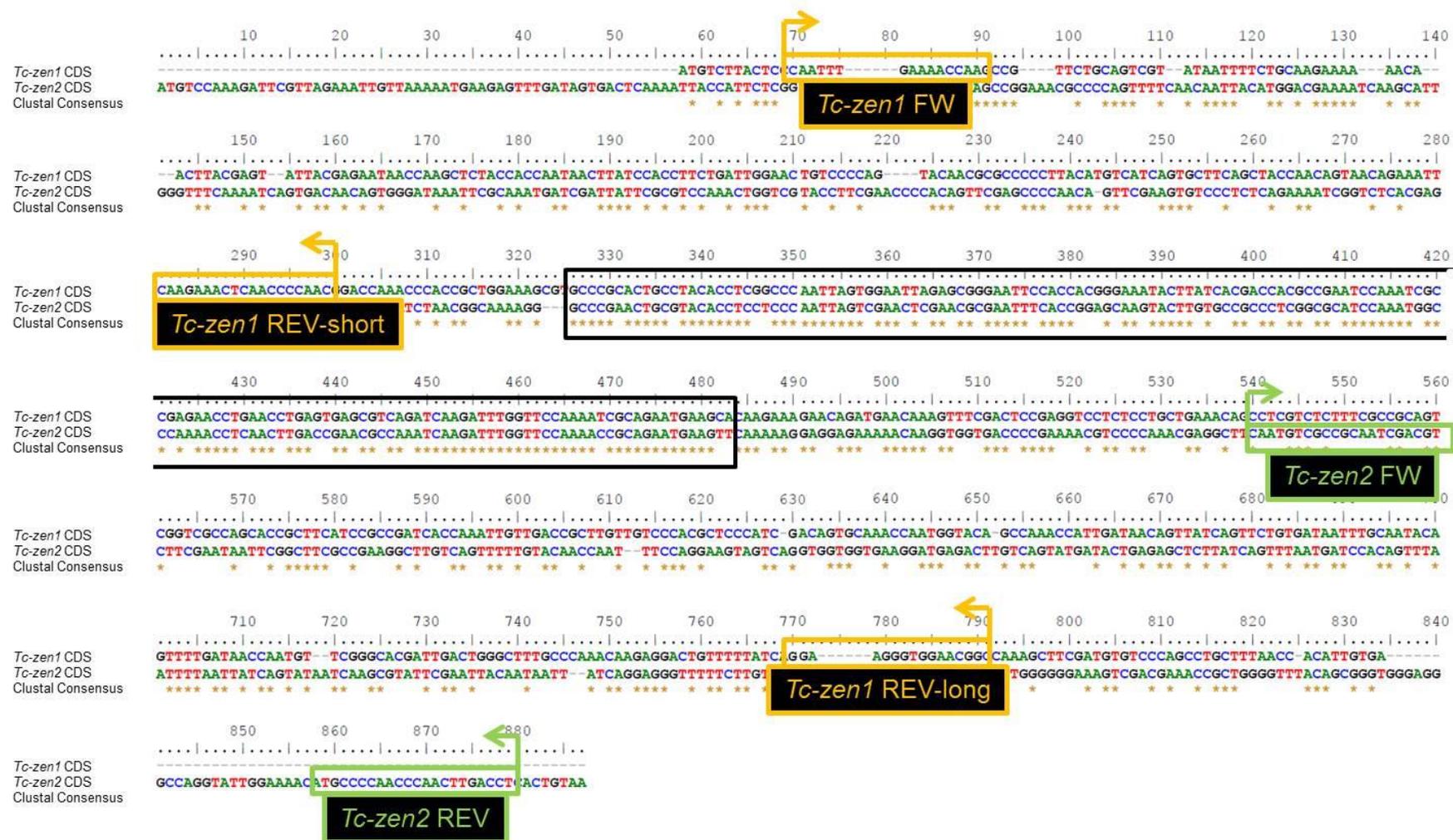


Figure 4.20. *Tc-zen1* long dsRNA fragment spans highly conserved region between *Tc-zen1* and *Tc-zen2*. *Tc-zen1* and *Tc-zen2* dsRNA primer position in regards to the sequence conservation between *Tc-zen* genes is shown. *Tc-zen1* and *Tc-zen2* share highly conserved homeobox region (black rectangle). Two primer pairs were designed for *Tc-zen1* dsRNA synthesis – short and long version (orange). The long version is spanning the highly conserved region. Only the short fragment avoiding the highly conserved region was designed for *Tc-zen2* dsRNA synthesis (green).

After *Tc-zen1*^{RNAi} is performed with the short dsRNA fragment, *Tc-zen1* expression decreases on average to 11% of its WT expression. When *Tc-zen1*^{RNAi} is performed with the long dsRNA fragment, the expression of *Tc-zen1* decreases on average to 10% of its WT expression (Fig. 4.21). This 1% decrease of *Tc-zen1* expression level after *Tc-zen1*^{RNAi} performed with the long dsRNA fragment suggests that the *Tc-zen1* KD efficiency is not affected by the length of *Tc-zen1* dsRNA.

On the other hand, when *Tc-zen1* KD is performed with the short dsRNA fragment, the expression of *Tc-zen2* decreases on average to 35% of its WT expression, while after usage of the long dsRNA fragment, the expression of *Tc-zen2* decreases on average to 19% (Fig. 4.21). This result suggests that the long *Tc-zen1* dsRNA fragment causes an off target effect on *Tc-zen2* expression and decreases its expression levels by 26%. Other than that, no morphological changes resembling *Tc-zen2* phenotype were observed in *Tc-zen1*^{RNAi} embryos.

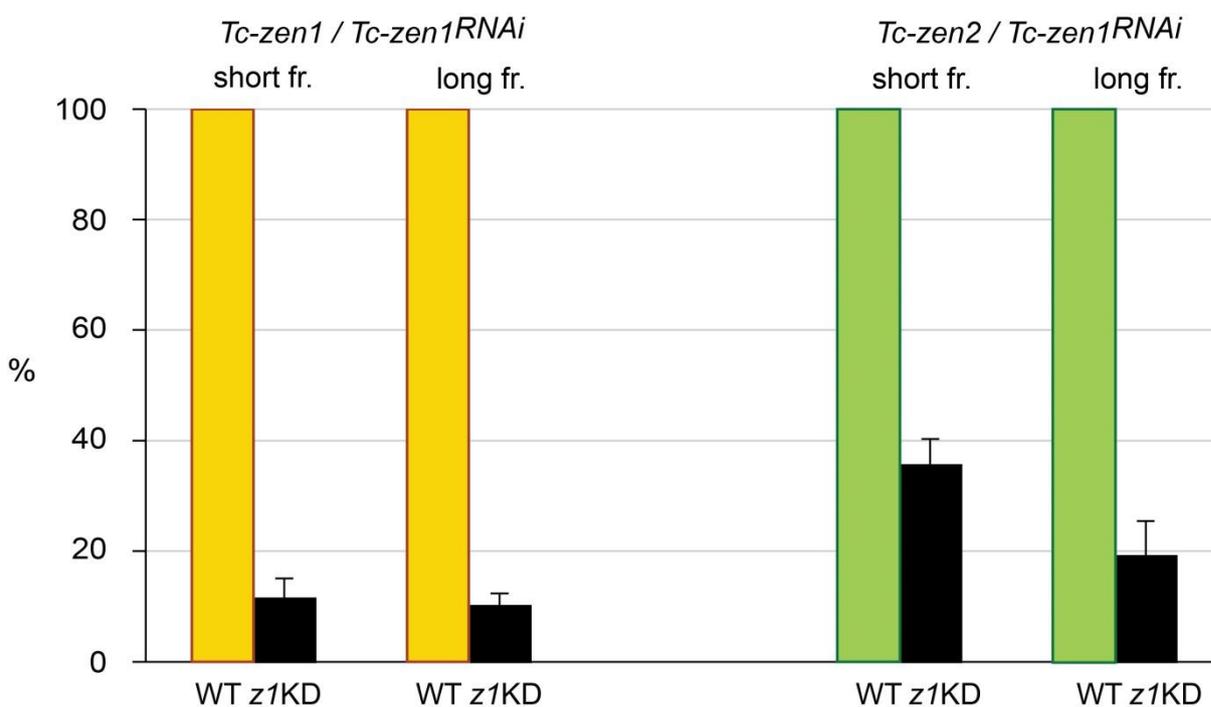


Figure 4.21. *Tc-zen1* long dsRNA fragment causes decrease in the *Tc-zen2* expression. The expression of *Tc-zen1* decreases to 11% and 10% of its WT expression, when *Tc-zen1*^{RNAi} is performed with the short and the long dsRNA fragment, respectively. The expression of *Tc-zen2* decreases to 35% of its WT expression in the *Tc-zen1*^{RNAi} samples when *Tc-zen1*^{RNAi} is performed with the short *Tc-zen1* dsRNA fragment, but when *Tc-zen1*^{RNAi} is performed with the long dsRNA fragment, the expression levels of *Tc-zen2* decrease to 19% of its WT expression. WT expression of *Tc-zen1* and *Tc-zen2* is considered to be 100%. Bars indicate mean values; error bars (standard deviation) represent variance across three biological replicates.

4.4 Global evaluation of *Tc-zen* genes' targets by RNA-sequencing after RNA interference

4.4.1 Variance between wild type and knockdown samples of early developmental stages

To identify and subsequently compare *Tc-zen1* and *Tc-zen2* candidate target genes during early embryogenesis, I performed RNA-seq after RNAi. *Tc-zen1* and *Tc-zen2* KD and corresponding WT samples in the stages of *Tc-zen* genes' expression peaks (WT1: 6-10 h AEL and WT2: 10-14 h AEL, respectively; Fig. 4.7B) were sequenced. To exclude any possibility of the off target effect, *Tc-zen1*^{RNAi} was performed only with the short *Tc-zen1* dsRNA fragment.

The degree of difference between the sequenced samples was first explored by principal component analysis (PCA) that clusters samples together according to their similarity. The first principal component clearly separates *Tc-zen1* KD samples and the corresponding WT1 samples from *Tc-zen2* KD samples and the corresponding WT2 samples. On the other hand, while the second principal component separates *Tc-zen1* KD samples from its corresponding WT1 samples, it fails to clearly separate *Tc-zen2* KD samples from its corresponding WT2 samples, which are grouped in close proximity to each other. This indicates little to no difference between *Tc-zen2* KD and WT2 samples obtained during early embryogenesis. Additionally, PCA revealed large difference between WT1 and WT2 samples, underlying the fact that the large number of developmental events is taking place within these embryonic stages (Fig. 4.22).

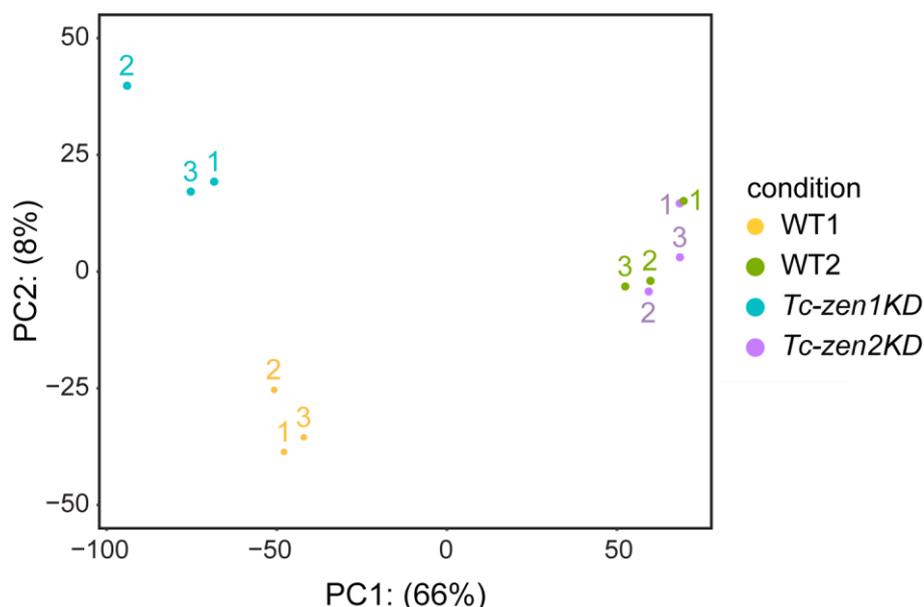


Figure 4.22. Principal component analysis of the RNA-sequencing samples in the early stages. PCA shows clear separation between *Tc-zen1*KD samples (blue) and the corresponding WT samples (WT1, orange). On the other hand, *Tc-zen2*KD samples (purple) were clustered in the close vicinity to the WT samples in the same stage (WT2, green). In the case of biological replicate 1, *Tc-zen2*KD and WT2 samples have overlapping positions. Numbers above dots represent the number of biological replicate. Variance explained by the first and the second principal component is shown in parentheses.

4.4.2 Identification of candidate target genes of *Tc-zen1* and *Tc-zen2* during early embryogenesis

I analyzed the RNA-seq data from the early developmental stages by using the pipeline, which has been developed within this project (see section 3). The final differential expression (DE) analysis generated lists of *Tc-zen1* and *Tc-zen2* candidate target genes. Using the lowest cut-off criteria regarding fold change (FC) and *P* adjusted values ($FC \geq [2], P_{adj} \leq 0.1$), 341 genes were identified as differentially expressed after *Tc-zen1* KD, whereas after KD of *Tc-zen2*, only 26 genes were differentially expressed (Table 4.5). Overall, 13-times more differentially expressed genes were identified after *Tc-zen1* KD. This result is consistent with those obtained from the PCA, which showed clear separation of the WT1 and *Tc-zen1* KD samples, but poor separation of the WT2 and *Tc-zen2* KD samples. Further, the low number of differentially expressed genes after *Tc-zen2* KD suggests that during early embryogenesis *Tc-zen2* has subtle functions.

Table 4.5. Count of differentially expressed genes during early embryogenesis. The number of differentially expressed genes after *Tc-zen1RNAi* and *Tc-zen2RNAi* under different cut-off conditions. FC-fold change, P_{adj} -p adjusted value. Number in parentheses represents the fact that mathematically it is not possible to divide by zero.

cut-off	<i>Tc-zen1</i> ^{RNAi}	<i>Tc-zen2</i> ^{RNAi}	ratio
$FC \geq [10], P_{adj} \leq 0.01$	20	0	(20)
$FC \geq [5], P_{adj} \leq 0.01$	75	3	25
$FC \geq [2], P_{adj} \leq 0.01$	338	26	13
$FC \geq [10], P_{adj} \leq 0.05$	20	0	(20)
$FC \geq [5], P_{adj} \leq 0.05$	75	3	25
$FC \geq [2], P_{adj} \leq 0.05$	341	26	13.1
$FC \geq [10], P_{adj} \leq 0.1$	20	0	(20)
$FC \geq [5], P_{adj} \leq 0.1$	75	3	25
$FC \geq [2], P_{adj} \leq 0.1$	341	26	13.1

4.4.3 Evaluation of potential target genes of *Tc-zen1* and *Tc-zen2*

To get insight into the functions of the potential targets of *Tc-zen1* and *Tc-zen2*, I performed a miniscreen of the chosen candidates. Initial choice of the candidate target genes was based on the results from DE analysis performed by quickNGS pipeline (Wagle et al., 2015), in which 43 *Tc-zen1* candidate targets had FC values higher than ± 10 . From these 43 genes I prioritized 16 based on their FC values (the highest/lowest was prioritized), direction of the DE (if

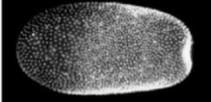
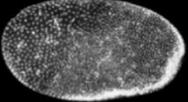
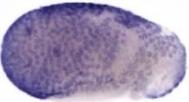
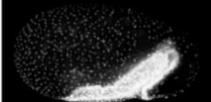
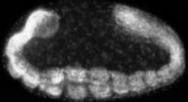
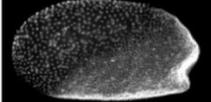
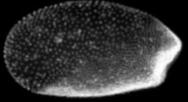
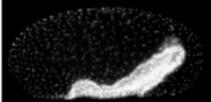
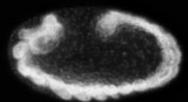
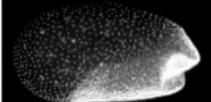
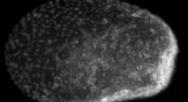
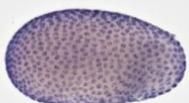
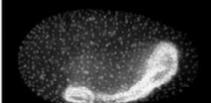
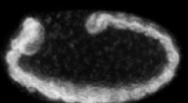
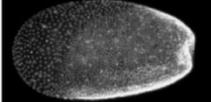
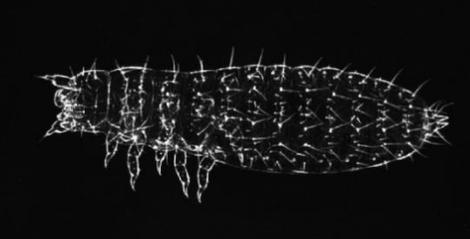
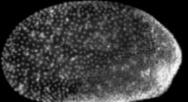
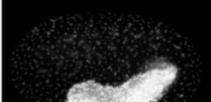
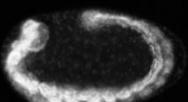
possible upregulated genes were prioritized due to the possible function in amnion; see section 4.3.1) and gene ontology (GO) terms (e.g.: potential function in cuticle structure/synthesis, genes with DNA binding domains). Only 16 genes were identified as differentially expressed with FC values higher than ± 10 after *Tc-zen2* KD. Therefore, I decided to evaluate all of them and no prioritizing was necessary for *Tc-zen2* candidate targets.

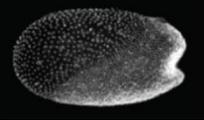
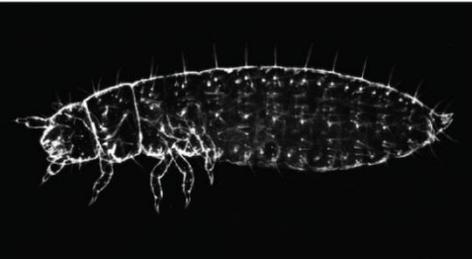
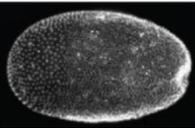
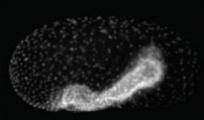
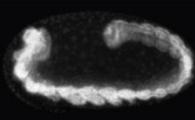
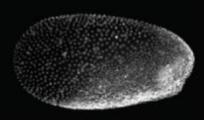
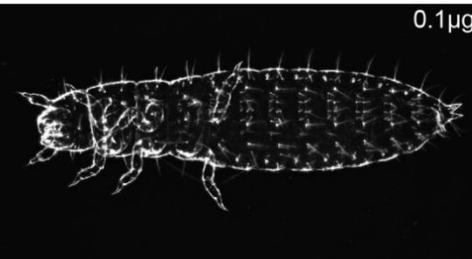
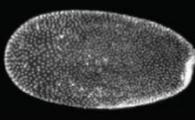
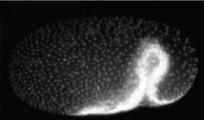
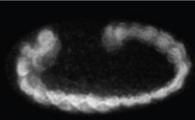
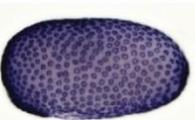
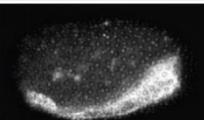
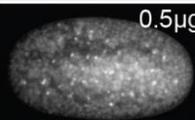
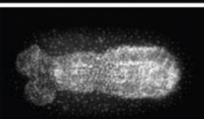
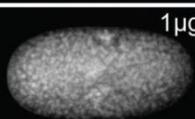
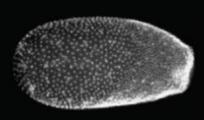
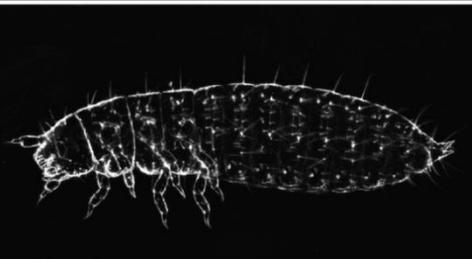
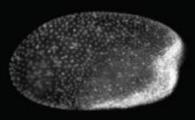
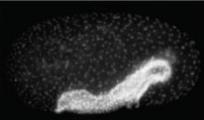
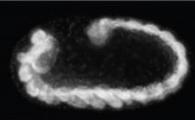
In the miniscreen I first investigated the expression pattern of all 32 candidate target genes by *in situ* hybridization. To investigate the function of some of these candidate targets, I performed pRNAi. Phenotypes were afterwards scored using three different methods: cuticle preparations, nuclear staining and serosal cuticle integrity determination. *Drosophila melanogaster* orthologues and the closest homologues of the candidate target genes were searched for in the Flybase (Gramates et al., 2017). Hints about the possible function of the candidate targets were looked for in the iBeetle-Base (Donitz et al., 2015).

4.4.4 *Tc-zen1* candidate target genes

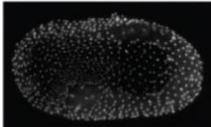
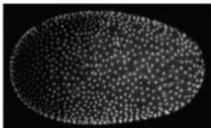
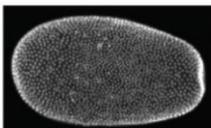
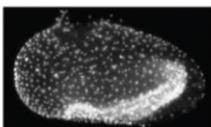
Among the 16 prioritized *Tc-zen1* candidate targets, nine genes were downregulated and had the lowest negative FC values, three were designated by GO terms as structural components of cuticle (serosa secretes chitin cuticle) and four genes were upregulated and had the highest positive FC values at the same time. To sum up, out of the 16 candidate target genes, 10 showed specific serosal expression pattern (Fig. 4.23). One gene showed embryonic expression pattern (*TC031198*), however the staining looked rather unspecific. I silenced expression of these 11 genes via pRNAi. KD of only one of the 11 genes caused a phenotype: after *TC015555^{RNAi}* I observed empty egg phenotype by performing all the three phenotype scoring methods mentioned above. Therefore, in order to describe the morphological defects, I first lowered the concentration to the half ($0.5 \mu\text{g}/\mu\text{l}$) and to the tenth ($0.1 \mu\text{g}/\mu\text{l}$) of the standard dsRNA concentration. Nonetheless, I again observed either the same empty egg phenotype ($0.5 \mu\text{g}/\mu\text{l}$), or no altered phenotype at all ($0.1 \mu\text{g}/\mu\text{l}$). After KD of the other 10 genes, the cuticles (larval and serosal) and the embryonic morphology were comparable with WT (Fig. 4.23). No defects in the serosal cuticle integrity were observed either (data not shown).

The other five genes showed unspecific staining and were marked as false positives (see section 3.1), hence KD was not pursued. The overview of the expression patterns, KD cuticle phenotypes, nuclear staining of the KD embryos and possible functions of the candidate targets is shown in the Fig. 4.23.

	Expression in WT		KD phenotype		Possible function
	ISH	DAPI	Cuticle preparations	DAPI	
TC000107					<p><i>Dm</i> orthologue: CG30427 - Fatty acyl-CoA reductase</p> <p><i>Dm</i> closest homologue: FAR1 - Fatty acyl-CoA reductase</p> <p>iBeetle screen: pupal phenotype: darker tanning of urogomphi no embryonic phenotype observed</p>
					
TC007258					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: CG9812 - no further information</p> <p>iBeetle screen: embryonic phenotype: muscle, anterior head and abdomen defects</p>
					
TC006727					<p><i>Dm</i> orthologue: CG33978 - contains EGF-like domain/ Ca-binding domain</p> <p><i>Dm</i> closest homologue: Pawn- EGF-like Ca-binding domain</p> <p>iBeetle screen: not screened</p>
					
TC015108					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: Desaturase 1 - contains fatty acid desaturase domain</p> <p>iBeetle screen: no embryonic phenotype observed</p>
					

	Expression in WT		KD phenotype		Possible function
	ISH	DAPI	Cuticle preparations	DAPI	
TC013480					<p><i>Dm</i> orthologue: Dcd - Dopa decarboxylase - aromatic-L-amino-acid decarboxylase activity</p> <p><i>Dm</i> closest homologue: Tdc2 - Tyrosine decarboxylase 2</p> <p>iBeetle screen: not screened</p>
					
TC015555					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: none existing</p> <p>iBeetle screen: not screened</p>
					
					
					
TC008400					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: Cpr72Ec - Cuticular protein 72Ec - structural constituent of chitin-based cuticle</p> <p>iBeetle screen: not screened</p>
					

	Expression in WT		KD phenotype		Possible function
	ISH	DAPI	Cuticle preparations	DAPI	
TC011141					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: Obst-A - Obstructor-A - contains chitin binding domain</p> <p>iBeetle screen: not screened</p>
TC031198					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: Grass - Gram-positive specific serine protease</p> <p>iBeetle screen: no embryonic phenotype observed</p>
TC011283					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: Wat - Waterproof - long-chain-fatty-acyl-CoA reductase activity</p> <p>iBeetle screen: no embryonic phenotype observed</p>
TC013404					<p><i>Dm</i> orthologue: CG10407 - contains haemolymph juvenile hormone binding domain</p> <p><i>Dm</i> closest homologue: CG10264 - contains haemolymph juvenile hormone binding domain</p> <p>iBeetle screen: not screened</p>

	Expression in WT		KD phenotype		Possible function
	ISH	DAPI	Cuticle preparations	DAPI	
TC014502			Not pursued: low read counts - false positive		<p><i>Dm</i> orthologue: CG1136 - structural constituent of cuticle</p> <p><i>Dm</i> closest homologue: L(3)mbn - Lethal (3) malignant blood neoplasm</p> <p>iBeetle screen: embryonic phenotype: no cuticle formation</p>
TC013320			Not pursued: low read counts - false positive		<p><i>Dm</i> orthologue: CG17167 - Na/Ca/P exchanger</p> <p><i>Dm</i> closest homologue: Zyd - Zydeco - Na/Ca/P exchanger</p> <p>iBeetle screen: not screened</p>
TC016348			Not pursued: low read counts - false positive		<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: CG4367 - chitin-binding protein</p> <p>iBeetle screen: no embryonic phenotype observed</p>
TC034701			Not pursued: low read counts - false positive		<p><i>Dm</i> orthologue: Ppk11 - Pickpocket 11 - epithelial Na channel</p> <p><i>Dm</i> closest homologue: Ppk16 - Pickpocket 16 - epithelial Na channel</p> <p>iBeetle screen: not screened</p>

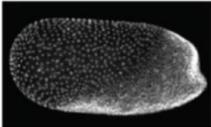
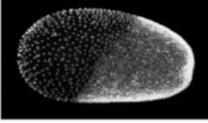
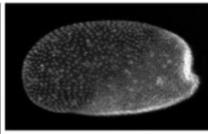
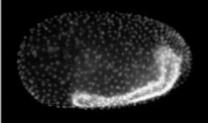
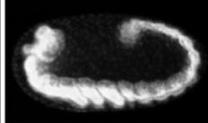
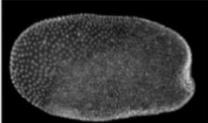
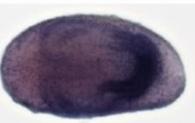
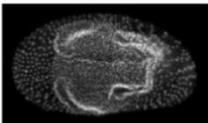
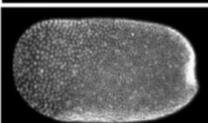
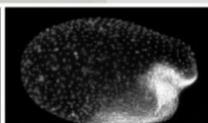
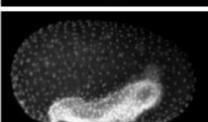
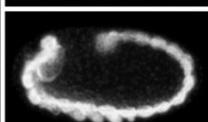
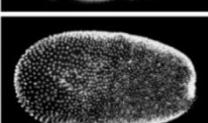
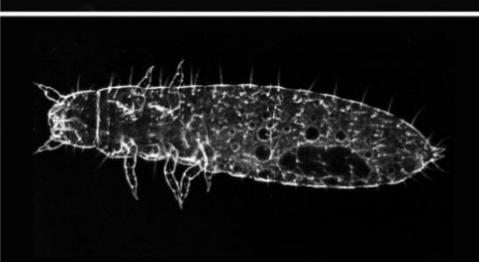
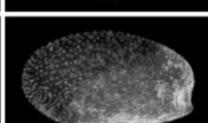
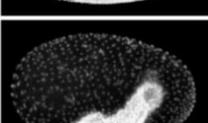
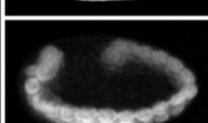
	Expression in WT		KD phenotype		Possible function
	ISH	DAPI	Cuticle preparations	DAPI	
TC012744			Not pursued: low read counts - false positive		<i>Dm</i> orthologue: none existing <i>Dm</i> closest homologue: none existing iBeetle screen: not screened

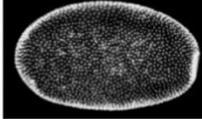
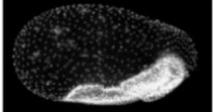
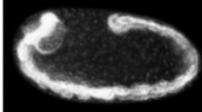
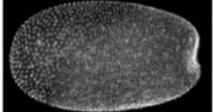
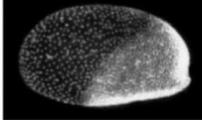
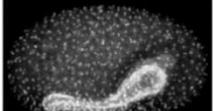
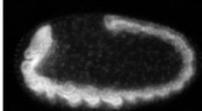
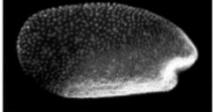
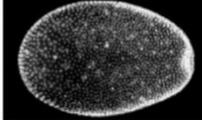
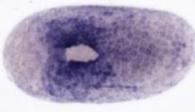
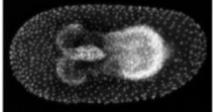
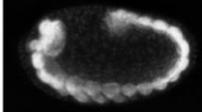
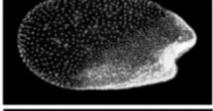
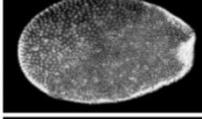
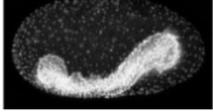
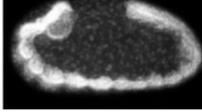
Figure 4.23. Miniscreen 1 - *Tc-zen1* candidate target genes. Summary of the *Tc-zen1* candidate target genes screened after the RNA-seq experiment in early embryogenesis. Candidate target genes were screened by *in situ* hybridization for their expression domain and by cuticle preparations and nuclear staining (DAPI) for the morphological defects of the KD embryos. Possible function of the target genes was inferred from *Drosophila melanogaster* (*Dm*) orthologues and the closest homologues and from the information available in the iBeetle database. Scale bar in the first TC000107 *in situ* hybridization image represents 100 μm and applies to all the *in situ* hybridization and DAPI images. Scale bar in the TC000107 cuticle preparation image represents 100 μm and applies to all the cuticle preparation images.

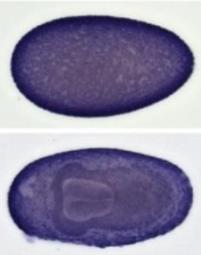
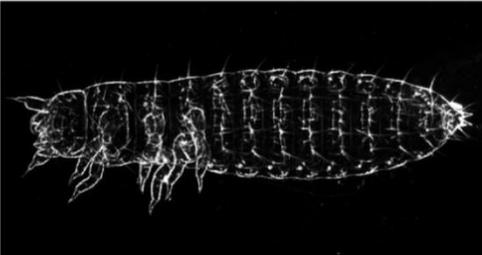
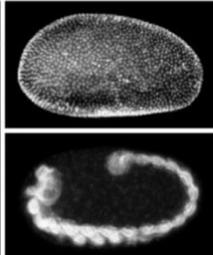
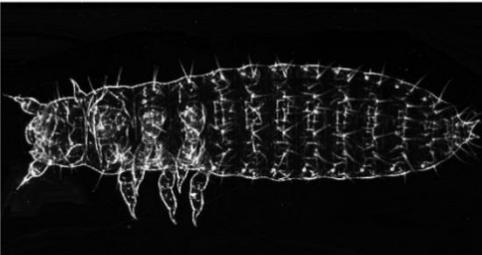
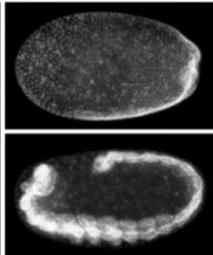
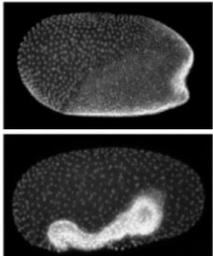
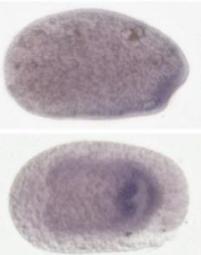
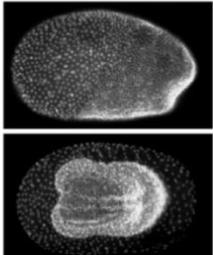
4.4.5 *Tc-zen2* candidate target genes

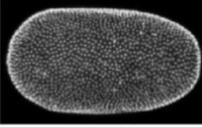
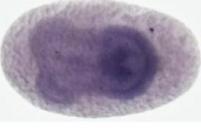
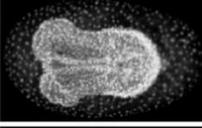
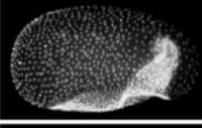
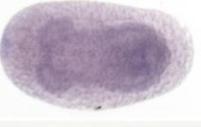
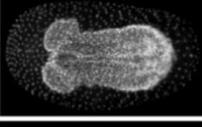
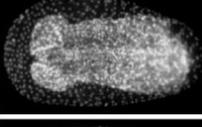
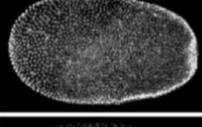
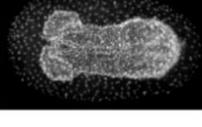
16 *Tc-zen2* candidate target genes were screened. These genes were designated as both up- and downregulated in the dataset obtained from the quickNGS pipeline. However, only one of the genes showed specific expression pattern in the serosa (*TC000511*) (Fig. 4.24). The other 15 genes showed either no, or unspecific staining. Nine of these 15 candidate target genes were marked as false positives (see section 3.1) and their KD was not pursued. The staining of the other six genes was simply unsuccessful, considering their expression (based on raw read count) in the sequenced samples was detectable by *in situ* hybridization (approx. 100 reads per gene per biological replicate, personal observation). Due to the fact, that I only identified one expression pattern as specific (*TC000511*), I added four new *Tc-zen2* candidate targets to the miniscreen. The new candidates were chosen from the datasets generated by custom pipeline we developed in this project (see section 3). Two of these genes showed specific serosal pattern (*TC008204-RB* and *TC0033464-RA*) and the other two (*TC011724-RA* and *TC000446-RA*) showed ubiquitous unspecific staining within the first five minutes of the staining reaction, suggesting unsuccessful staining (Fig. 4.24).

To sum up, I knocked down only genes that were not marked as false positives (10 genes in total), despite the fact that not all of them showed specific expression pattern. I did not observe any embryonic phenotypes, as the cuticle preparations and nuclear staining of KD embryos were comparable with WT. No defects in the serosal cuticle integrity were observed either (data not shown). The only phenotype was observed for the *TC011068* gene, which caused 100% pupal lethality after injection. This result was consistent with the information retrieved from the iBeetle screen database. The overview of the expression patterns, KD cuticle phenotypes, nuclear staining of the KD embryos and the possible functions of the *Tc-zen2* candidate targets is shown in the Fig. 4.24.

	Expression in WT		KD phenotype		Possible function
	ISH	DAPI	Cuticle preparations	DAPI	
TC007326					<p><i>Dm</i> orthologue: CG16959 - no further information</p> <p><i>Dm</i> closest homologue: not described</p> <p>iBeetle screen: not screened</p>
					
TC011068					<p><i>Dm</i> orthologue: Pot - Papillote - chitin-based cuticle attachment to epithelium</p> <p><i>Dm</i> closest homologue: not described</p> <p>iBeetle screen: high lethality in all stages (100% as pupa)</p>
					
TC000511					<p><i>Dm</i> orthologue: Hdly - Hadley - negative regulation of female receptivity, post-mating</p> <p><i>Dm</i> closest homologue: not described</p> <p>iBeetle screen: no embryonic phenotype observed</p>
					
TC000853					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: Cpr 92A - Cuticular protein 92A - structural constituent of cuticle</p> <p>iBeetle screen: no embryonic phenotype observed</p>
					

	Expression in WT		KD phenotype		Possible function
	ISH	DAPI	Cuticle preparations	DAPI	
TC002837					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: CG12164 - belongs to Ferritin-like superfamily</p> <p>iBeetle screen: not screened</p>
					
TC011635					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: Mag - Magro - lipase</p> <p>iBeetle screen: previtellogenic egg chamber number increased embryonic phenotype: no cuticle formation</p>
					
TC008204-RB					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: Dat - Dopamine N acetyltransferase</p> <p>iBeetle screen: no embryonic phenotype observed</p>
					
TC033464-RA					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: CG13618 - contains haemolymph juvenile hormone binding domain</p> <p>iBeetle screen: no embryonic phenotype observed</p>
					

	Expression in WT		KD phenotype		Possible function
	ISH	DAPI	Cuticle preparations	DAPI	
TC011724-RA					<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: CG13643 - contains chitin binding domain</p> <p>iBeetle screen: defects in eclosion no embryonic phenotype observed</p>
TC000446-RA					<p><i>Dm</i> orthologue: CG7530 - receptor activity</p> <p><i>Dm</i> closest homologue: not described</p> <p>iBeetle screen: embryonic phenotype: full or partial inside-out</p>
TC015615			Not pursued: low read counts - false positive		<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: CG30339 - contains lipid binding domain</p> <p>iBeetle screen: no embryonic phenotype observed</p>
TC002092			Not pursued: low read counts - false positive		<p>gene model in OGS3 split into 5 genes</p> <p><i>Dm</i> orthologue of TC031642-RA: Ance-3 - angiotensin converting enzyme, metaloprotease activity</p> <p><i>Dm</i> closest homologue of TC031642-RA: not described</p> <p>iBeetle screen of TC031642-RA: not screened</p>

	Expression in WT		KD phenotype		Possible function
	ISH	DAPI	Cuticle preparations	DAPI	
TC002831			Not pursued: low read counts - false positive		<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: not described</p> <p>iBeetle screen: not screened</p>
					
TC008236			Not pursued: low read counts - false positive		<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: not described</p> <p>iBeetle screen: no embryonic phenotype observed</p>
					
TC014361			Not pursued: low read counts - false positive		<p><i>Dm</i> orthologue: none existing</p> <p><i>Dm</i> closest homologue: Cyp6a2 - Cytochrome P450-6a2</p> <p>iBeetle screen: no embryonic phenotype observed</p>
					
TC001119			Not pursued: low read counts - false positive		<p><i>Dm</i> orthologue: Dhc 16F - Dynein heavy chain 16F, ATPase activity</p> <p><i>Dm</i> closest homologue: not described</p> <p>iBeetle screen: no embryonic phenotype observed</p>

	Expression in WT		KD phenotype		Possible function
	ISH	DAPI	Cuticle preparations	DAPI	
TC013146			Not pursued: low read counts - false positive		<i>Dm</i> orthologue: none existing <i>Dm</i> closest homologue: Hr38 - Hormone receptor - like in 38 iBeetle screen: embryonic phenotype: defects in cuticle
TC000546			Not pursued: low read counts - false positive		<i>Dm</i> orthologue: none existing <i>Dm</i> closest homologue: CG11836 - serine protease iBeetle screen: no embryonic phenotype observed
TC002942			Not pursued: low read counts - false positive		<i>Dm</i> orthologue: Zip42C.1 - Zn/Fe transmembrane transporter protein 42C.1 <i>Dm</i> closest homologue: Zip89B - Zn/Fe transmembrane transporter protein 89B iBeetle screen: no embryonic phenotype observed

Figure 4.24. Miniscreen 1 - *Tc-zen2* candidate target genes. Summary of the *Tc-zen2* candidate target genes screened after the RNA-seq experiment in early embryogenesis. Candidate target genes were screened by *in situ* hybridization for their expression domain and by cuticle preparations and nuclear staining (DAPI) for the morphological defects of the KD embryos. Possible function of the target genes was inferred from *Drosophila melanogaster* (*Dm*) orthologues and the closest homologues and from the information available in the iBeetle database. Scale bar in the first TC007326 *in situ* hybridization image represents 100 μ m and applies to all *in situ* hybridization and DAPI images. Scale bar in the TC007326 cuticle preparation image represents 100 μ m and applies to all the cuticle preparation images.

4.4.6 Does *Tc-zen2* copy *Tc-zen1* function during early embryogenesis?

Due to the fact that *Tc-zen2* reaches expression peak during early embryogenesis, while its morphogenesis function takes place during the late developmental stage, we assumed that *Tc-zen2* could potentially have early function as well. Within this project I have already identified early *Tc-zen2* role in *Tc-zen1* repression (Fig. 4.18, 4.19). In addition, we hypothesize that due to the fact that *Tc-zen* genes arose from a recent gene duplication, *Tc-zen2* might copy function of *Tc-zen1* during early embryogenesis, possibly with the lower threshold, because, based on the RNA-seq results, the early *Tc-zen2* regulatory function seems to be subtle. In order to investigate this hypothesis, I decided to identify number of target genes, which are shared by both *Tc-zen* genes during early embryogenesis.

To identify the number of target genes that *Tc-zen1* and *Tc-zen2* share, I performed comparative analyses of differentially expressed genes. Due to the fact that samples were sequenced in two different time points (WT1 and *Tc-zen1* KD samples: 6-10 h AEL and WT2 and *Tc-zen2* KD samples: 10-14 h AEL, hereafter referred to as WT shift), during which many important developmental events take place, I based all data comparisons on the group of differentially expressed and non-differentially expressed genes within the WT shift. Out of 18536 isoforms (15222 gene models in OGS3) 7391 isoforms were filtered out prior to the DE analysis between WT1 and WT2 datasets, because they did not fulfil the requirement of minimum ten reads per gene (see section 3.1). Out of the remaining 11145 isoforms, only those with P_{adj} value less or equal to 0.1 (6844 isoforms) were considered for the further analysis of shared target genes. I have divided these 6844 isoforms into three groups: not strongly differentially expressed ($-2 > FC < 2$) within the WT shift (4275), strongly upregulated ($FC \geq 2$) (1205), and strongly downregulated isoforms ($FC \leq -2$) (1364). Due to the facts that under WT conditions *Tc-zen1* expression decreases and *Tc-zen2* expression increases within the WT shift, all the downregulated genes within the WT shift are potentially genes that *Tc-zen1* activates and *Tc-zen2* represses. On the other hand, all the upregulated genes within the WT shift are potentially repressed by *Tc-zen1* and activated by *Tc-zen2*.

At first, I investigated how many genes that were not differentially expressed within the WT shift, *Tc-zen1* and *Tc-zen2* share. Out of 4275 isoforms, 88 are strongly differentially expressed after *Tc-zen1* KD and zero after *Tc-zen2* KD. When lowering the DE threshold for *Tc-zen2* targets (from $\geq [2]$ to $> [1]$), 12 genes are shared by both *Tc-zen* genes (Fig. 4.25A).

Next, of 1364 strongly downregulated isoforms within the WT shift, 62 are strongly differentially expressed after *Tc-zen1* KD and zero after *Tc-zen2* KD. After lowering the DE threshold for *Tc-zen2* targets, only one gene is shared between *Tc-zen1* and *Tc-zen2* (Fig. 4.25B).

Further, out of all the strongly upregulated isoforms within the WT shift (1205), 16 are strongly differentially expressed after *Tc-zen2* KD and one of them is also strongly differentially expressed after *Tc-zen1* KD. However, this one shared gene is *Tc-zen2*. Nonetheless, after lowering the DE threshold for *Tc-zen2* differentially expressed isoforms, 192 isoforms are differentially expressed after *Tc-zen2* KD and 34 of these are also targets of *Tc-zen1* (Fig. 4.25C). Interestingly, 29 out of these 34 shared genes, changed the direction of

DE from being downregulated after *Tc-zen1* KD to being upregulated after *Tc-zen2* KD, although with the lower FC values.

Next, out of all the strongly upregulated isoforms within the WT shift (1205), only one gene is differentially expressed after *Tc-zen1* KD and none after *Tc-zen2* KD. With the lower DE threshold for *Tc-zen2* targets, still no genes are shared between *Tc-zen1* and *Tc-zen2* (Fig. 4.25D).

Finally, out of 1364 strongly downregulated isoforms within the WT shift, zero are differentially expressed after the KD of both *Tc-zen* genes. With the lower DE threshold for *Tc-zen2* targets, 57 isoforms are differentially expressed and one of them is shared with *Tc-zen1* (Fig. 4.25E).

Collectively, after the removal of duplicates and isoforms, and with the lower threshold criteria for *Tc-zen2* targets, *Tc-zen1* and *Tc-zen2* share 45 genes (Table S1). For 14 genes (out of these 45) either no homologue in *D. melanogaster* exists, or only CG identification number, without any GO term assigned, is available. These 14 genes were either not screened within the iBeetle screen, or no further information after the screen is provided. For the remaining 31 genes, information about GO terms of homologues from *D. melanogaster* is available, however these GO terms represent various functions. Out of these 31 genes, 13 were screened within iBeetle screen, but KD of only one gene caused, among others, partial inside-out phenotype. Altogether, these results suggest that *Tc-zen2* does not regulate substantial amount of *Tc-zen1* targets, even on the lower threshold levels, and those genes that are shared between paralogues are possibly involved in many different functions.

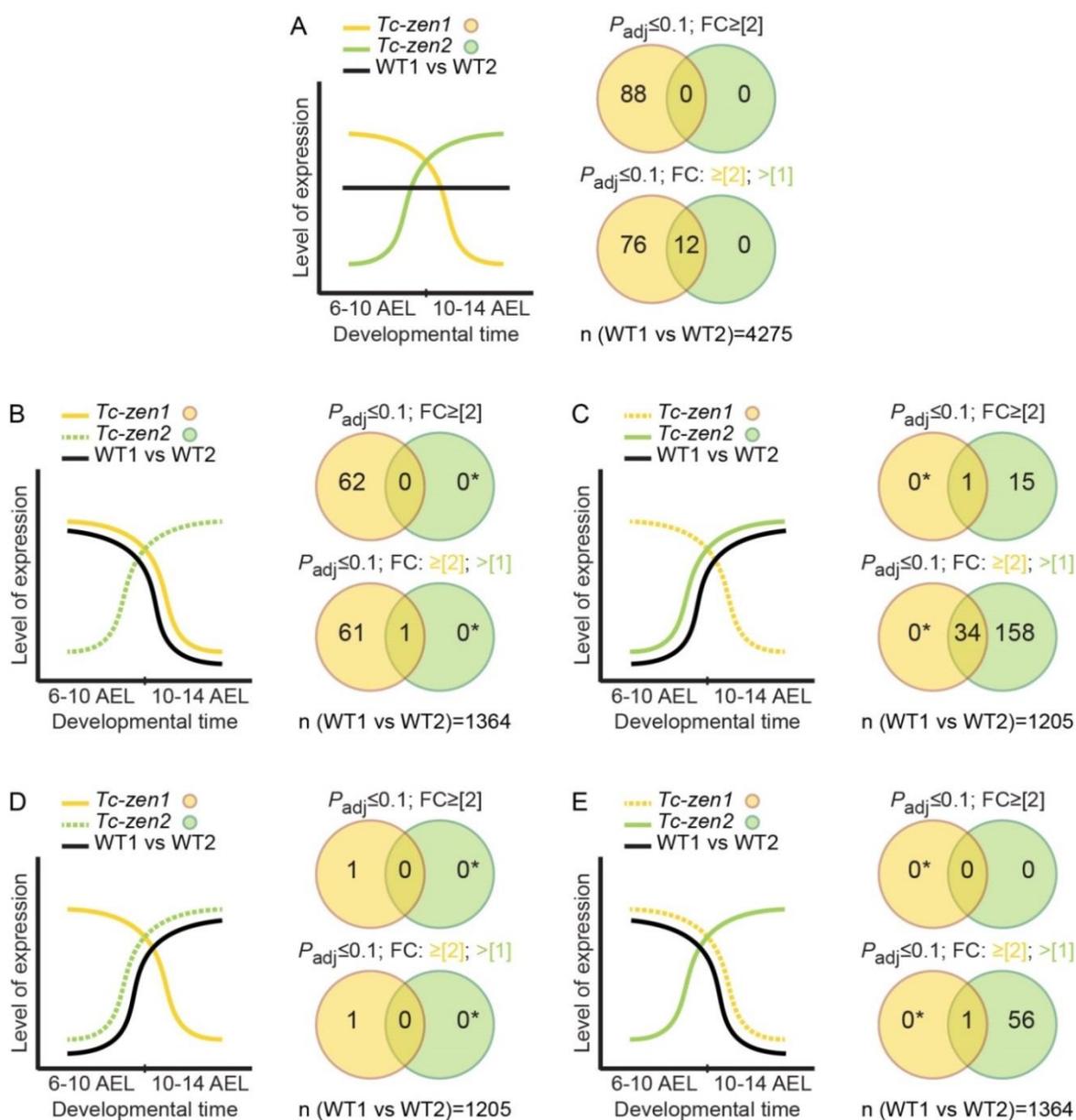


Figure 4.25. Shared candidate target genes between *Tc-zen1* and *Tc-zen2*. Overlap of differentially expressed candidate target genes after *Tc-zen1* and *Tc-zen2* KD. Out of 4275 isoforms that are not strongly differentially expressed within the WT shift, 88 are strongly differentially expressed after *Tc-zen1* KD and 0 genes are strongly differentially expressed after *Tc-zen2* KD. After lowering the cut-off for *Tc-zen2* targets, 12 genes are shared by *Tc-zen1* and *Tc-zen2* (A). Out of 1364 strongly downregulated isoforms within the WT shift, 62 are strongly differentially expressed after *Tc-zen1* KD. Out of these 62 isoforms, 0 genes are strongly differentially expressed after *Tc-zen2* KD. With the lower cut-off for *Tc-zen2* targets, 1 gene is shared by *Tc-zen1* and *Tc-zen2* (B). Out of 1205 strongly upregulated isoforms within the WT shift, 16 are strongly differentially expressed after *Tc-zen2* KD. Out of these 16 isoforms, 1 gene is strongly differentially expressed after *Tc-zen1* KD and shared with *Tc-zen2*. With the lower cut-off for *Tc-zen2* targets, 192 genes are differentially expressed and out of these 34 are shared with *Tc-zen1* (C). Out of 1205 strongly upregulated genes within the WT shift, 1 is strongly differentially expressed after *Tc-zen1* KD and this gene is not strongly differentially expressed after *Tc-zen2* KD. After lowering the cut-off for *Tc-zen2* targets, 0 genes are shared by *Tc-zen1* and *Tc-zen2* (D). Out of 1364 strongly downregulated isoforms within the WT shift, 0 genes are strongly differentially expressed after *Tc-zen2* KD. With the lower cut-off for *Tc-zen2* targets, 57 isoforms are differentially expressed and out of these 1 is shared with *Tc-zen1* (E). 0* represents 0 genes by definition based on the analysis; dashed line represents the fact that data were compared to this dataset in the second step of the particular analysis (see Methods, section 2.12.9 for details).

4.4.7 Variance between wild type and knockdown samples of late developmental stages

To identify downstream genes, which *Tc-zen2* effects during its late morphogenesis function, which takes place during the membrane rupture stage, I performed the second RNA-seq after RNAi experiment. In this case, *Tc-zen2* KD and WT samples right before the rupture (pre-rupture, 48-52 h AEL) and during, as well as, right after the rupture (post-rupture, 52-56 h AEL) were sequenced.

To reveal the degree of difference between the *Tc-zen2* KD and the WT samples collected before and after the membrane rupture, I performed PCA. The first principal component clearly separates *Tc-zen2* KD and the WT samples in the respective pre- and post-rupture stages (Fig. 4.26). This separation suggests that the stages I have examined are indeed relevant for detecting the transcriptional underpinnings of *Tc-zen2*'s morphogenesis function. In addition, the first principal component clusters in the close proximity *Tc-zen2* KD samples of the post-rupture stage (purple) (biological replicate 1, 2 and 4) and the WT samples of the pre-rupture stage (red) (Fig. 4.26). This result may reflect a delay in the development after *Tc-zen2* KD caused by the stress from the ectopic membrane rupture.

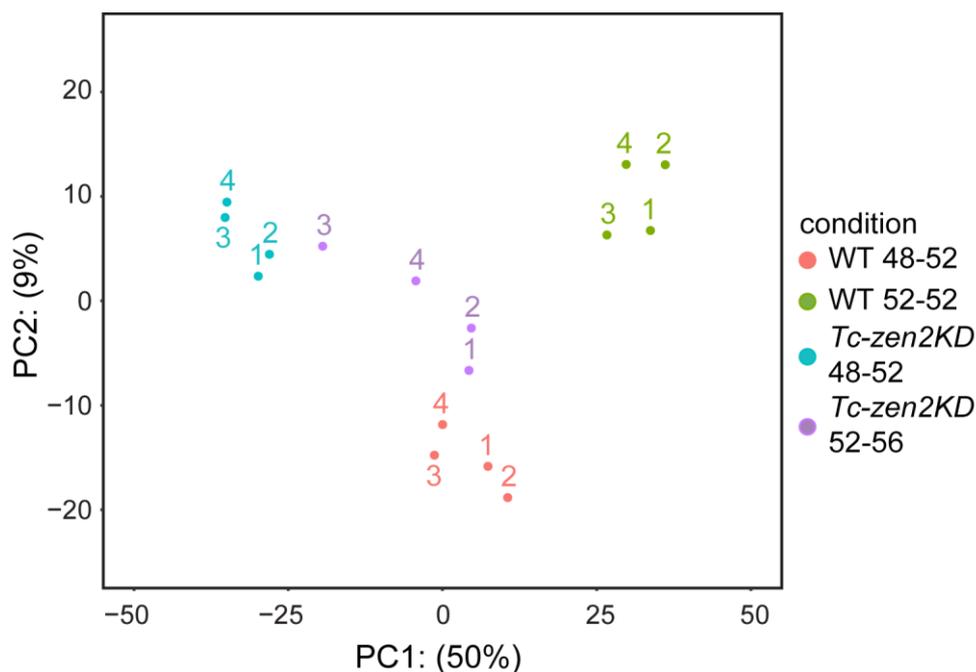


Figure 4.26. Principal component analysis of the RNA-sequencing samples in the late stages. *Tc-zen2* KD samples of the pre-rupture stage (48-52 h AEL, blue) cluster in different positions than the WT samples in the same stage (red). The same is true for the post-rupture *Tc-zen2* KD samples (52-56 h AEL, purple), which cluster in different positions compared to the respective WT samples (green). In addition, the post-rupture *Tc-zen2* KD samples (purple) cluster in the close vicinity of the WT pre-rupture samples (red) (all except biological replicate 3, which groups rather with the KD pre-rupture samples). Numbers represent the number of biological replicate. Variance explained by the first and the second principal component is shown in parentheses.

4.4.8 Identification of *Tc-zen2* candidate target genes during late embryogenesis

I analyzed the RNA-seq data from the late developmental stages by using the pipeline, which we developed within this project (see section 3). In order to identify potential target genes of *Tc-zen2* during late development, I performed DE analysis and generated the list of differentially expressed genes after *Tc-zen2* KD during the late development. With our low-stringency cut-off criteria ($FC \geq [2]$, $P_{adj} \leq 0.1$), 481 genes differentially expressed after *Tc-zen2* KD in the pre-rupture stage and 431 genes differentially expressed after the *Tc-zen2* KD in the post-rupture stage (Table 4.6) were identified. This high number of differentially expressed genes is consistent with the PCA results (Fig. 4.26), which show clear separation of the WT and the *Tc-zen2* KD samples in the respective pre- and post-rupture stages.

Table 4.6. Count of differentially expressed genes during late embryogenesis. The number of differentially expressed genes after *Tc-zen2*^{RNAi} in the pre-rupture and the post-rupture stages under the different cut-off conditions. FC-fold change, P_{adj} -p adjusted value.

cut-off	<i>Tc-zen2</i> ^{RNAi} pre-rupture	<i>Tc-zen2</i> ^{RNAi} post-rupture
$FC \geq [10], P_{adj} \leq 0.01$	9	3
$FC \geq [5], P_{adj} \leq 0.01$	42	22
$FC \geq [2], P_{adj} \leq 0.01$	481	431
$FC \geq [10], P_{adj} \leq 0.05$	9	3
$FC \geq [5], P_{adj} \leq 0.05$	42	22
$FC \geq [2], P_{adj} \leq 0.05$	481	431
$FC \geq [10], P_{adj} \leq 0.1$	9	3
$FC \geq [5], P_{adj} \leq 0.1$	42	22
$FC \geq [2], P_{adj} \leq 0.1$	481	431

4.4.9 Functional profile of *Tc-zen2* candidate target genes of late development

In order to verify that the differentially expressed genes identified after *Tc-zen2* KD in the late developmental stages play roles in extraembryonic development and not in embryonic development, I performed a large-scale GO term analysis. Using the Blast2GO program (Conesa et al., 2005) four datasets of differentially expressed genes were used as input for the analysis: genes differentially expressed in the pre-rupture stage (48-52 h AEL), genes differentially expressed in the post-rupture stage (52-56 h AEL), genes differentially expressed between the WT pre- and post-rupture stages and differentially expressed genes between *Tc-zen2* KD pre- and post-rupture stages. These datasets were blasted against two different blast databases available in NCBI: *Drosophila* and non-redundant (nr) databases. It is important to note that only differentially expressed genes with successful mapping and annotation were assigned GO terms (see Methods, section 2.12.10). On average, 51-54% of the differentially expressed genes were considered for the further GO term analysis when blasted against the *Drosophila* and 61-63% when blasted against the nr databases.

After I blasted the four above mentioned datasets against the two databases, I searched the results for the specific GO names that could represent the ongoing processes in the

extraembryonic tissues. I looked specifically for the GO terms that fell into the following categories: “stress”, “cuticle”, “cytoskeleton”, “morphogenesis and epithelium”, “wing disc and pupal and metamorphosis”, “extracellular matrix (ECM) and adhesion”, “regulation of gene expression” and “transmembrane transport” (for a detailed list of GO terms grouped to the categories of interest see Methods, Table 2.4) (Fig. 4.27).

The four datasets were first blasted against the nr database. 20-31% of all the sequences fall into the categories of interest within the biological process GO type. Within the molecular function GO type, 20-29% of the sequences fall into the categories of interest. Due to the small number of the sequences generally falling into the cellular component GO type, large percent representation of sequences falls into the categories of interest: 42-73% (Fig. 4.27A, Table 4.7A).

Since, after blasting against the nr database, the number of sequences falling into the categories of interests within biological process and the molecular function GO type was rather low, we decided to perform the second GO term analysis by blasting the four datasets against the *Drosophila* database. 57-62% of all the sequences fall into the categories of interest within the biological process GO type. Within the molecular function GO type, 34-41% of all the sequences fall into the categories of interest and within the cellular component GO type, 43-58% of all the sequences represent the categories of interest (Fig. 4.27B, Table 4.7B). The results from the two GO term analyses with two different databases suggest that the choice of the database can drastically influence the result of the GO term analysis and therefore, the conclusions about the datasets of differentially expressed genes and ultimately the interpretation of underlying biological events.

Nonetheless, with the focus on the biological process GO type of the latter analyses, the most abundant categories of interest are “morphogenesis and epithelium” and “wing disc and pupal and metamorphosis” with percent representation of 16-17% and 13-15% of all the sequences, respectively (Table 4.7B). The highest percent representation in these two categories was rather expected, as *Tc-zen2* has morphogenesis function in the epithelial tissue (EEMs), and the tissues during metamorphosis and wing disc formation are of epithelial character (Aldaz et al., 2010; Hilbrant et al., 2016). Only 5-7% of all the sequences fell into the category “regulation of gene expression”, which might reflect that during late development, as a result of *Tc-zen2* KD, we observe a morphogenesis phenotype manifestation, rather than a transcriptional effect in terms of altered transcriptional regulation. Overall, after blasting the four datasets against the *Drosophila* database, on average only about 50% of sequences were assigned GO terms from the categories of interest. However, complete information about the differentially expressed genes, which functional profile was not retrieved using Blast2GO program, is missing and altogether represents 40% of the dataset. In order to retrieve full information about the four datasets, a further comparative analysis is required. Employment of a functional analysis of the *Tc-zen2* candidate target genes would certainly be necessary in the future.

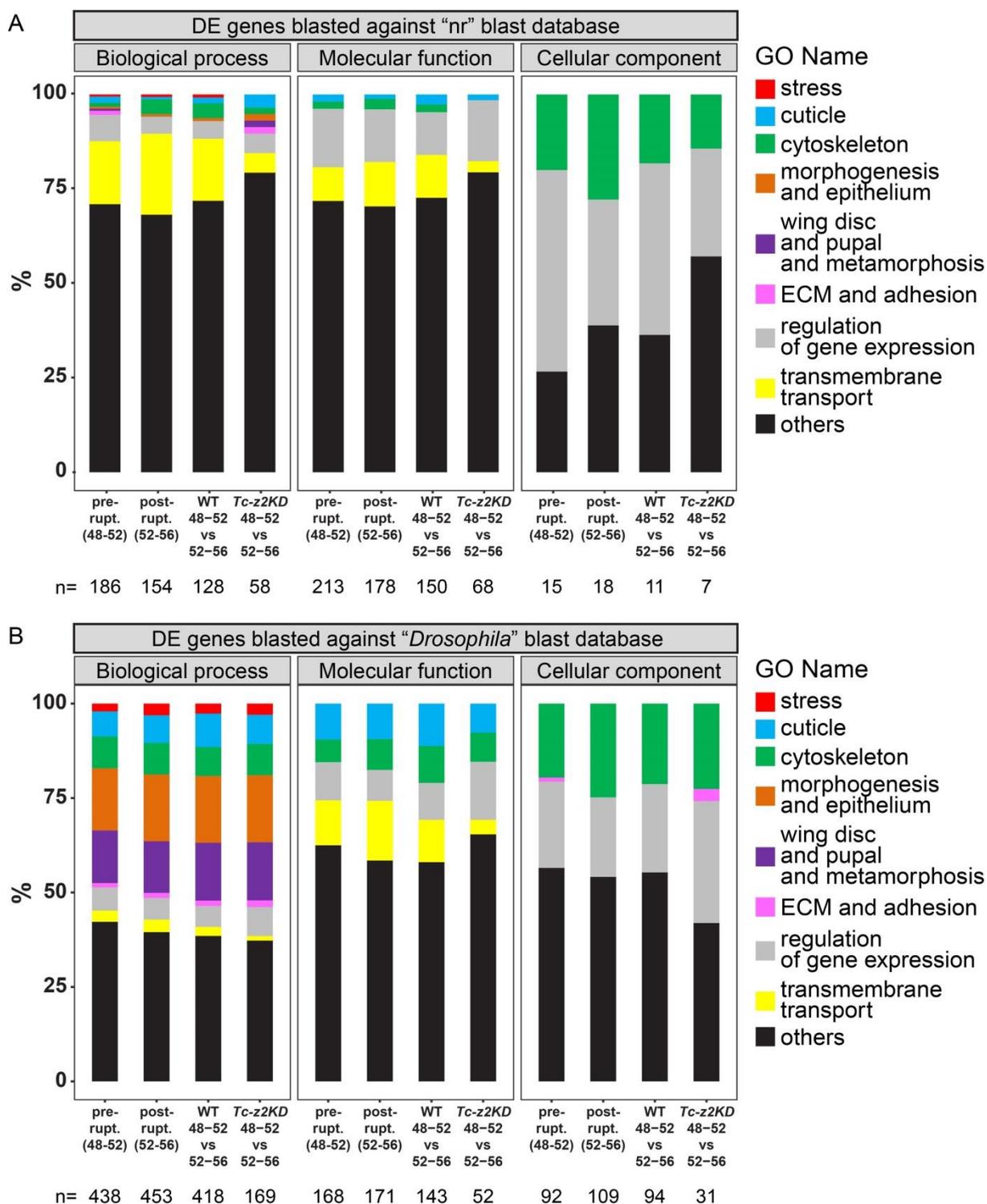


Figure 4.27. Percentage of the sequences in the GO term categories of interest. Approximately 25% of the total number of sequences fall into the categories of interest in the biological process and molecular function GO type, while in the cellular component GO type it is up to 73% of the sequences (mind the low sample size), when blasted against the nr database (**A**). After blasting against the *Drosophila* database, 60% of the total sequence number falls into the categories of interest into the biological process GO type, 30-40% of sequences fall into the molecular function GO type and 50% into the cellular component GO type (**B**) (see also Table 4.7). n represents sample size of *T. castaneum* gene sequences. DE-differentially expressed.

Table 4.7A. Sequence percent representation in the categories of interest after blasting against the nr database. DE-differentially expressed.

DE genes blasted against the “nr” blast database												
GO Name	Biological process [%]				Molecular function [%]				Cellular component [%]			
	Pre-rupture	Post-rupture	WT 48-52 vs 52-56	<i>Tc-zen2</i> KD 48-52 vs 52-56	Pre-rupture	Post-rupture	WT 48-52 vs 52-56	<i>Tc-zen2</i> KD 48-52 vs 52-56	Pre-rupture	Post-rupture	WT 48-52 vs 52-56	<i>Tc-zen2</i> KD 48-52 vs 52-56
stress	0.54	0.65	0.78	NA	NA	NA	NA	NA	NA	NA	NA	NA
cuticle	1.61	0.65	1.56	3.45	1.88	1.12	2.67	1.47	NA	NA	NA	NA
cytoskeleton	1.08	3.9	3.91	1.72	1.88	2.79	2	NA	20	27.78	18.18	14.29
morphogenesis and epithelium	0.54	0.65	0.78	1.72	NA	NA	NA	NA	NA	NA	NA	NA
wing disc and pupal and metamorphosis	0.54	NA	NA	1.72	NA	NA	NA	NA	NA	NA	NA	NA
ECM and adhesion	1.08	NA	NA	1.72	NA	NA	NA	NA	NA	NA	NA	NA
regulation of gene expression	6.99	4.55	4.69	5.17	15.49	13.97	11.33	16.18	53.33	33.33	45.45	28.57
transmembrane transport	16.67	21.43	16.41	5.17	8.92	11.73	11.33	2.94	NA	NA	NA	NA
others	70.97	68.18	71.88	79.31	71.83	70.39	72.67	79.41	26.67	38.89	36.36	57.14

Table 4.7B. Sequence percent representation in the categories of interest after blasting against the *Drosophila* database. DE-differentially expressed.

DE genes blasted against the “ <i>Drosophila</i> “ blast database												
GO Name	Biological process [%]				Molecular function [%]				Cellular component [%]			
	Pre-rupture	Post-rupture	WT 48-52 vs 52-56	<i>Tc-zen2</i> KD 48-52 vs 52-56	Pre-rupture	Post-rupture	WT 48-52 vs 52-56	<i>Tc-zen2</i> KD 48-52 vs 52-56	Pre-rupture	Post-rupture	WT 48-52 vs 52-56	<i>Tc-zen2</i> KD 48-52 vs 52-56
stress	2.05	3.09	2.63	2.96	NA	NA	NA	NA	NA	NA	NA	NA
cuticle	6.62	7.28	8.85	7.69	9.52	9.36	11.19	7.69	NA	NA	NA	NA
cytoskeleton	8.45	8.39	7.66	8.28	5.95	8.19	9.79	7.69	19.57	24.77	21.28	22.58
morphogenesis and epithelium	16.44	17.66	17.7	17.75	NA	NA	NA	NA	NA	NA	NA	NA
wing disc and pupal and metamorphosis	13.93	13.69	15.31	15.38	NA	NA	NA	NA	NA	NA	NA	NA
ECM and adhesion	1.14	1.32	1.44	1.78	NA	NA	NA	NA	1.09	NA	NA	3.23
regulation of gene expression	6.16	5.74	5.5	7.69	10.12	8.19	9.79	15.38	22.83	21.1	23.4	32.26
transmembrane transport	2.97	3.31	2.39	1.18	11.9	15.79	11.19	3.85	NA	NA	NA	NA
others	42.24	39.51	38.52	37.28	62.5	58.48	58.04	65.38	56.52	54.13	55.32	41.94

Besides searching for the GO terms that represent events ongoing in the extraembryonic tissues, I also looked for the terms that most differentially expressed genes were assigned to in all the three GO types after blasting against both databases. When blasted against the nr database, in all the four datasets blasted, the highest number of sequences is assigned to proteolysis, ion transport and aminosugar metabolic process GO term within the biological process GO type. Within the molecular function GO type the highest number of sequences is assigned to metal ion binding and peptidase activity GO term. Finally, within the cellular component GO type, the highest number of sequences is assigned to mitochondrion, spliceosome and centrosome complex GO term (Table 4.8A).

On the other hand, when blasted against the *Drosophila* database, the highest number of sequences is assigned to cell differentiation and nervous system development GO term. This is true for all the four datasets for the biological process GO type. Next, within the molecular function GO type, similarly to the results obtained after the blasting against the nr database, the highest number of sequences is assigned to metal ion binding, peptidase activity and channel activity GO term. Lastly, in the cellular component GO type, the highest number of sequences is assigned to lipid particle, nuclear lumen and cytosol GO term (Table 4.8B).

Table 4.8A. The first two sequence most rich GO terms assigned in each GO type in 4 different datasets after blasting against the nr database. DE-differentially expressed.

DE genes blasted against the “nr” blast database				
GO type	Pre-rupture	Post-rupture	WT (48-52 vs 52-56)	<i>Tc-zen2KD</i> (48-52 vs 52-56)
Biological process	proteolysis	proteolysis	proteolysis	proteolysis
	aminosugar metabolic process	ion transport	aminoglycan metabolic process	aminoglycan metabolic process
Molecular function	metal ion binding	metal ion binding	metal ion binding	metal ion binding
	peptidase activity, acting on L-amino acid peptides			
Cellular component	mitochondrion	mitochondrion	mitochondrion	mitochondrion
	spliceosome complex	vacuole	mitochondrial outer membrane	centrosome

Table 4.8B. The first two sequence most rich GO terms assigned in each GO type in 4 different datasets after blasting against the *Drosophila* database. DE-differentially expressed.

DE genes blasted against the “ <i>Drosophila</i> “ blast database				
GO type	Pre-rupture	Post-rupture	WT (48-52 vs 52-56)	<i>Tc-zen2KD</i> (48-52 vs 52-56)
Biological process	cell differentiation	cell differentiation	cell differentiation	cell differentiation
	nervous system development	nervous system development	nervous system development	nervous system development
Molecular function	peptidase activity, acting on L-amino acid peptides	metal ion binding	metal ion binding	hydrolase activity
	serine peptidase activity	channel activity	peptidase activity, acting on L-amino acid peptides	metal ion binding
Cellular component	cytosol	lipid particle	nuclear lumen	nuclear lumen
	lipid particle	nuclear lumen	lipid particle	cytoplasmic vesicle

4.4.10 Evaluating differential expression of *Tc-zen2* candidate targets in late development

As previously mentioned, it is not possible to investigate expression pattern of genes by *in situ* hybridization in the stages of late development due to the serosal cuticle, which is not permeable for hybridization probes. Therefore, I validated transcriptomic data results by a RT-qPCR miniscreen. I tested, whether the genes in general are differentially expressed, and whether the direction of the DE corresponds to the results from the RNA-seq data analysis.

Based on the GO term analysis results obtained after the blasting against the *Drosophila* database, I chose several genes for miniscreen. The choice of the genes was based on the GO term category, FC values and the direction of DE. I chose seven genes from each pre-rupture stage and the post-rupture stage. Two of these 14 genes were chosen on the basis that they are differentially expressed in both pre- and post-rupture stages and the direction of the DE changed between the stages (Table 4.9). The DE itself as well as the direction of DE of the chosen genes was confirmed by the RT-qPCR miniscreen (Fig. 4.28).

Table 4.9. Summary of all the genes screened after RNA-sequencing in late embryogenesis. Listed genes were screened in the RT-qPCR miniscreen after the RNA-seq during the pre- and post-rupture stages. *Dmel-Drosophila melanogaster*.

TC ID	pre-/post-rupture	GO Descriptor	<i>Dmel</i> homologue	FC
TC010840	pre-rupture	no information available	no homologue	-2.16
	post-rupture			2.56
TC011665	pre-rupture	morphogenesis and epithelia	Orisis	-2.76
	post-rupture			2.03
TC005982	pre-rupture	cytosol	Major facilitator superfamily transporter 3	4.22
TC006575	pre-rupture	transmembrane transport	Pickpocket 26	12.68
TC014041	pre-rupture	cell differentiation	Z band alternatively spliced PDZ-motif	-4.74
TC014143	pre-rupture	stress	Protein kinase, cAMP-dependent, catalytic subunit 3	-2.36
TC014497	pre-rupture	cuticle	Cuticular protein 65Av	-5.25
TC007162	post-rupture	morphogenesis and epithelia	Ribbon	2.37
TC031481	post-rupture	regulation of gene expression	Nuclear factor interleukin-3-regulated protein-like protein	2.90
TC033856	post-rupture	cytoskeleton	CG8213	2.54
TC008606	post-rupture	transmembrane transport	Excitatory amino acid transporter 3-like protein	-2.89
TC011349	post-rupture	cuticle	Chondroitin proteoglycan 2-like protein	-3.32

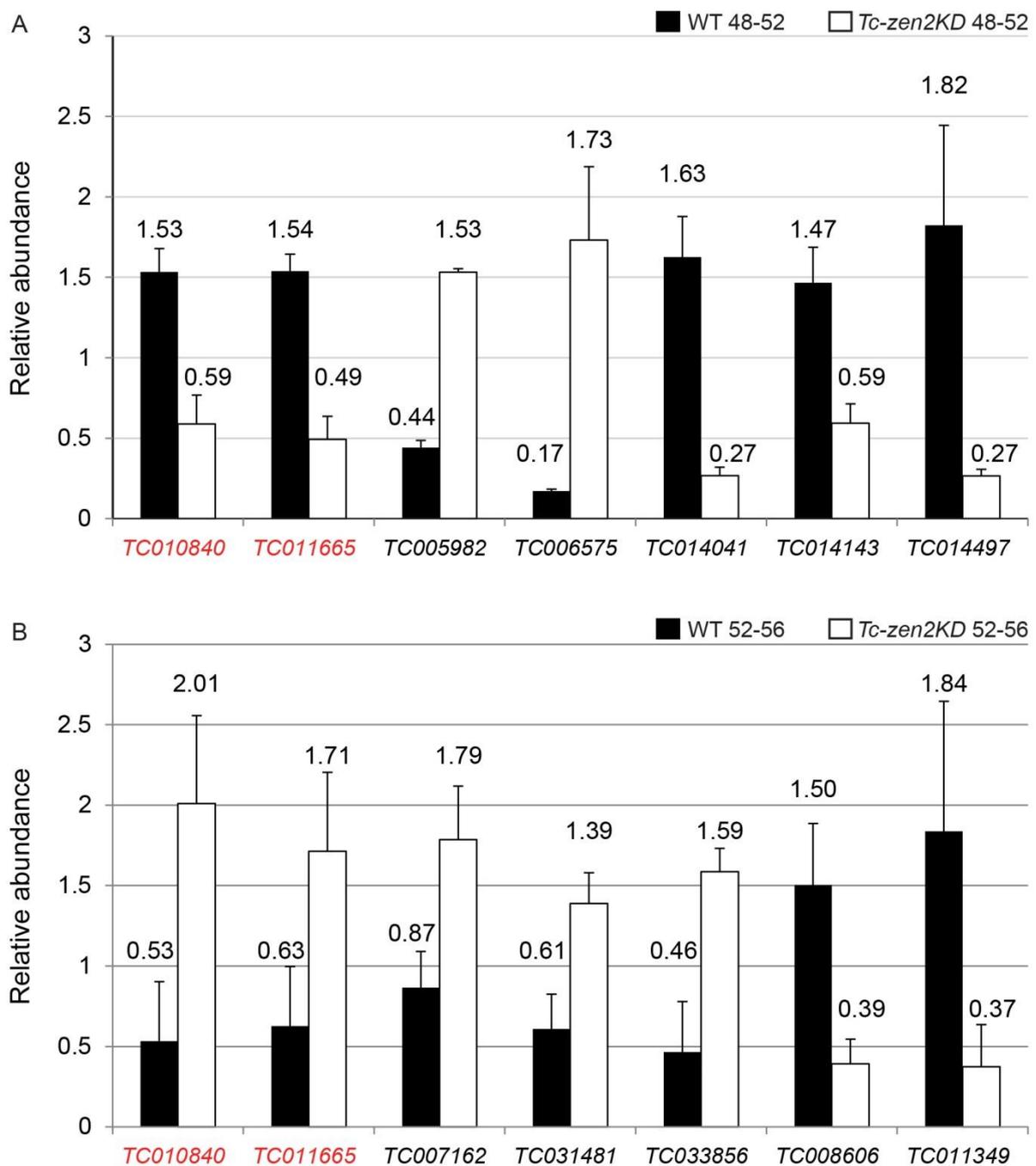


Figure 4.28. Miniscreen 2 - *Tc-zen2* candidate target genes during late embryogenesis. RT-qPCR miniscreen of 14 genes after the RNA-seq of the samples in late stages. Seven genes were chosen from each pre-rupture (**A**) and post-rupture (**B**) stage. Two genes are differentially expressed in both stages (red TC IDs) and have opposite directions of DE in each stage. DE itself, as well as the DE direction of all the screened genes, correspond to the transcriptomic datasets (see also Table 4.9). Bars indicate mean values; error bars (standard deviation) represent variance across three biological replicates.

5 DISCUSSION

5.1 Conservation levels of non-coding regions between *zen* genes

5.1.1 Promoters of *Tc-zen1* and *Tc-zen2* differ in sequence

In *T. castaneum*, *Tc-zen1* and *Tc-zen2* arose from lineage specific tandem duplication. The *Tc-zen* paralogues are located in Hox3 locus and are separated from each other only by 172 bp. Nevertheless, they perform two distinct functions during two developmental stages, separated from each other by two days. Therefore, we assume that their expression is regulated through different regulatory regions, or the restriction of regulatory crosstalk is required in order to allow for a fine-tuned expression of both *Tc-zen* genes.

By investigation of conservation levels between genomic DNA sequences of *Tc-zen1* and *Tc-zen2*, I have identified highly conserved regions only within the coding sequence, particularly in the homeobox region (Fig. 4.1 and 4.2). Highly conserved non-coding regions were not identified between the core promoters of *Tc-zen* paralogues, suggesting that the gene expression of *Tc-zen1* and *Tc-zen2* is controlled from different proximal regulatory regions. Change of regulatory regions followed by alteration of upstream transcription factors could explain, to a certain degree, paralogue's acquirement of distinct functions (True and Carroll, 2002) and their subsequent separation to two different developmental stages.

However, with lower threshold of conservation I identified two conserved non-coding regions. One of them is positioned between the promoter sequence of *Tc-zen1* and the second intron of *Tc-zen2* (Fig. 4.2A). This region is only conserved when DNA strands of *Tc-zen1* and *Tc-zen2* are in opposite direction. The change of the direction of the functional conserved regulatory region is rather atypical, but nonetheless occurs (Bradley et al., 2010).

The other conserved non-coding region was identified between the second intron of *Tc-zen1* and the first intron of *Tc-zen2* (Fig. 4.2A). Regulatory regions within introns have been described in diverse species ranging from protozoans, plants to vertebrates, although usually they are present within the first intron of the gene (Mack and Owens, 1999; Kim et al., 2002; Calderwood et al., 2003). What is rather intriguing is the position of the identified conserved regions. The change in the position of regulatory regions between orthologues in respect to the first exon is possible, but it usually occurs within distal promoter region (the first 500 bp upstream of transcriptional start site) (Davidson, 2001; Stone and Wray, 2001). A jump of regulatory regions between different exons of orthologues has not been reported, thus the identified conserved non-coding regions are most likely not relevant.

However, before any conclusions can be made, the functional relevance of these two conserved non-coding regions needs to be investigated further. The most obvious necessity is a search for transcription factor binding sites (TFBS) within the identified conserved non-coding regions. Both of these regions are 70% conserved within 30 nt long sequence (21 nt identical), which is rather a short conserved region. On the other hand, TFBS are indeed very short (4-8 nt), especially those involved in transcriptional regulation of tissue specific genes,

as they are controlled through focused transcription initiation (Juven-Gershon and Kadonaga, 2010).

5.1.2 Several conserved non-coding regions between four *Tribolium* species were identified

Investigation of conserved non-coding regions between the promoters of *Tc-zen1* and *Tc-zen2* could enlighten the conservation level of only *cis*-regulatory regions. In order to identify regulatory regions potentially responsible for long-distance transcriptional regulation of the *Tc-zen* paralogues, I performed comparative analysis of Hox3 loci of four closely related *Tribolium* species. Several conserved non-coding regions were identified among four species (both up- and downstream of *zen* genes, as well as in their intergenic region) (Fig. 4.3 and 4.4). In order to evaluate the significance of these conserved non-coding regions, further analysis of TFBS is necessary. Since the identified conserved regions could potentially play role in regulating expression of both *Tc-zen1* and *Tc-zen2*, a search for chromatin insulator binding sites might be of high relevance. Particularly investigation of insulator CTCF (CCCTC-binding factor) binding site should be performed, since CTCF regulates colinear expression within the Hox cluster of *D. melanogaster*. This Hox-CTCF interaction has been shown to provide mechanism for body patterning and has been conserved across Bilateria (Heger et al., 2012). Therefore, it might be possible that CTCF is involved in selective regulation of the fine-tuned expression of the *Tc-zen* genes.

5.2 Transcriptional and translational regulation differ between the *Tc-zen* paralogues

5.2.1 *Tc-zen1* is transiently expressed in embryo

Consistent with their function in extraembryonic (EE) development, *Tc-zen1* and *Tc-zen2* are expressed in the serosa (Fig. 4.5 and 4.6). However, transient expression of *Tc-zen1* was repeatedly observed in the embryo during the serosal window stage (Fig. 4.5G-I' and 5.1A). Since *zen* has derived from canonical *Hox3* gene, whose function is in axial patterning, it might be possible that the observed small transient *Tc-zen1* embryonic expression domain is a residue of its ancestral *Hox3* expression. The expression domain of *Tc-zen1* in embryo is localized within the expression domain of *Tc-Hox4* gene, *Deformed* (*Dfd*) (Fig. 5.1, compare A and B). One of the hypotheses of the insect *Hox3* gene ability to abandon the canonical function without deleteriously affecting embryonic development suggests that the loss was possible due to its redundancy, which arose as a consequence of the overlapping expression domains with the neighboring Hox genes (*Hox2* and *Hox4*) (Telford and Thomas, 1998a, b). Therefore, the observation that *Tc-zen1* is expressed in embryo within the domain of *Tc-Dfd* could suggest the reduced ancestral expression of *Tc-Hox3*.

On the other hand, embryonic *zen* expression has not been observed in any basally branching (more ancestral) insect species studied so far, which is in contrast to the speculations about the ancestral residual embryonic *Hox3* expression of *Tc-zen1*.

Another possible simple reason for the observed embryonic *Tc-zen1* expression could be explained by cross-reactivity of *Tc-zen1* hybridization probe with *Tc-Dfd* due to the high level of sequence homology between Hox genes. The probe designed within this project spans almost the entire mRNA sequence of *Tc-zen1* and therefore raises the chance of cross-reaction. On the other hand, even *zen* and *Hox3* genes do not share highly conserved region within full mRNA sequences or even within the homeodomain itself (Fig. 1.1) (Panfilio et al., 2006; Panfilio and Akam, 2007), therefore it is not likely that the probe would cross-react with *Tc-Hox4* (*Dfd*). The question about transient embryonic expression of *Tc-zen1* remains opened and further investigations are necessary. To start, one could perform sequence pairwise alignment of *Tc-zen1* and *Tc-Dfd* in order to identify potential cross-reacting regions and afterwards exploit expression patterns with different hybridization probes (different in length) avoiding potential conserved regions.

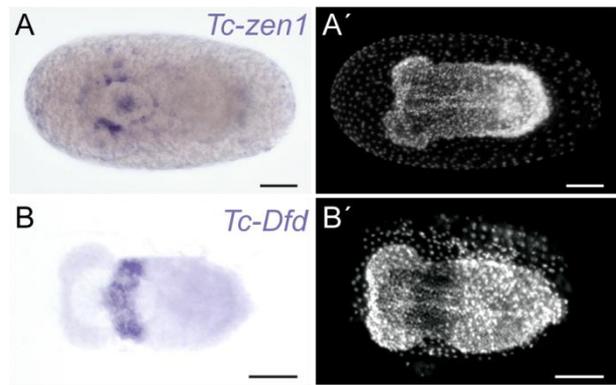


Figure 5.1. Expression of *Tc-zen1* and *Tc-Dfd* during the serosal window stage. *Tc-zen1* is expressed in the border of the serosal window and in the ventral-medial domain of the embryo (A). *Tc-Dfd* is expressed in the embryo with typical stripe pattern of Hox genes (B). Views are lateral with anterior left. Images labeled with the same letter are of a single embryo. Scale bars in images represent 100 μ m. Images B and B' were provided by N. Frey.

5.2.2 *Tc-zen1* and *Tc-zen2* are not maternally supplied

In order to decipher how did the two distinct roles of *Tc-zen1* and *Tc-zen2* arise, I investigated differences in their transcriptional and translational regulation. Particularly, I described transcript and protein expression profiles of both *Tc-zen* genes.

It has been previously reported that *Tc-zen1* is maternally expressed (van der Zee et al., 2005). However, expression of *Tc-zen1* and *Tc-zen2* during early developmental stages (from egg lay until 4 h AEL) was not detected (Fig. 4.7A). Maternal determination of *zen* varies among studied species. While no maternal supply of *zen* was described for bug *O. fasciatus* (Panfilio et al., 2006) or flies *M. abdida*, *E. balteatus* and *D. melanogaster* (Rafiqi et al., 2008), in locust *S. gregaria* *zen* is maternally determined. Moreover, expression of *zen* in *S. gregaria* oocyte is localized to cytoplasm and suggestions about specific control of *zen*'s exclusion from nucleus have been proposed (Dearden et al., 2000).

Yet another remarkable case of maternal supply of *zen* is represented by butterfly *Pararge aegeria*. In this species four extra copies of *zen* genes have been recently discovered in Hox3 locus (Ferguson et al., 2014). The four *zen* copies have been termed *Special homeobox A-D* (*ShxA-D*) genes and their expression has been detected in extraembryonic membranes

(EEMs). However, only two of the *Shx* genes (*C* and *D*) are maternally supplied, while the other two *Shx* genes and bona fide *zen* are supplied zygotically. It seems that whether *zen* is maternally supplied or not is rather a species specific characteristic and no correlation between maternal determination of *zen* and any of the two functions described for *zen* has been proposed.

5.2.3 *Tc-zen1* and *Tc-zen2* expression peak is only during early embryogenesis

Investigation of transcript expression profiles of *Tc-zen1* and *Tc-zen2* revealed the highest expression of both *Tc-zen* genes only during early embryogenesis (Fig. 4.7B and 4.8). *Tc-zen1* expression peak is reached during the uniform blastoderm and the primitive pit stages. Later, during the differentiation of blastoderm and gastrulation its expression decreases and abruptly switches off after the serosal window closure (Fig. 4.7B). Although *Tc-zen2* morphogenesis role takes place during late development, its expression peak was observed already during early embryogenesis. *Tc-zen2* expression starts later than *Tc-zen1*, in the differentiated blastoderm stage, peaks during gastrulation and the serosal window closure stage, and gradually wanes of during germband extension (Fig. 4.7B). Since the expression of *Tc-zen2* has only been detected in the serosa, the slight offset in *Tc-zen2*'s expression could be explained by the necessity of its expression domain establishment by *Tc-zen1*. Further, the early *Tc-zen2* expression suggests that *Tc-zen2* is functionally relevant also during early development (discussed later).

5.2.4 Minimal *Tc-zen2* expression is sufficient for protein turnover through late embryogenesis

Due to the fact that *Tc-zen2* function was described during late development (van der Zee et al., 2005), I investigated the expression of *Tc-zen2* before its function in membrane rupture takes place. Minimal and decreasing expression of *Tc-zen2* was observed at least 10 h before the membrane rupture occurs (Fig. 4.8). Further, the expression of Tc-Zen2 protein was detected throughout the whole embryogenesis until the late developmental stages (Fig. 4.10B). These results suggest that, although *Tc-zen2* expression peak occurs during early embryogenesis, the minimal expression of *Tc-zen2* continues and, moreover, is sufficient for the maintenance of protein expression until the time point when its function is required (42 h later). However, what is still unknown is the function of *Tc-zen2* during those 42 hours.

5.2.5 Tc-Zen2 is localized only to nucleus during early and late embryogenesis

Since Tc-Zen2 function takes place during late development, but its expression was detected throughout early and late embryogenesis, we wondered whether the initial trigger of *Tc-zen2* function could occur earlier than right before the membrane rupture. We hypothesized that the initial trigger could be detected by possible change of the cellular localization of Tc-Zen2 from cytoplasm to nucleus. Exclusion from nucleus and cytoplasmic localization has been

previously reported for Zen orthologue of locust *S. gregaria* (Dearden et al., 2000). However, cytoplasmic localization of Tc-Zen2 was not observed either during early, or during late embryogenesis (Fig. 4.9M-P and 4.11).

Another possibility for detection of potential trigger of late *Tc-zen2* function could be to investigate Tc-Zen2 protein modification changes during the entire developmental time, in which presence of Tc-Zen2 has been detected. Further, when exactly *Tc-zen2* expression needs to be silenced in order to induce inside-out phenotype could be examined by embryonic injections of *Tc-zen2* dsRNA in different consecutive developmental embryonic stages.

5.2.6 Tc-Zen2 is expressed exclusively in the serosa during entire lifespan of extraembryonic membranes

It has been previously reported that *Tc-zen2* plays role specifically in the amnion during the membrane rupture. Its amniotic expression was observed during the extended germband stage and its role in the fusion of EEMs before the rupture was proposed (van der Zee et al., 2005). Nonetheless, within this project, no amniotic expression was observed for the transcript (Fig. 4.6H-I), or for Tc-Zen2 protein (Fig. 4.9M-P, 4.11 and 4.12). *Tc-zen2* expression was solely observed in the serosa. Moreover, it was recently reported that the serosa and the amnion do not fuse before rupture, but rather form a bilayer and rupture independently from each other, although the role of the amnion in the rupture initiation was proposed (Hilbrant et al., 2016). However, since *Tc-zen2* expression was not observed in the amniotic cap (rupture competence zone of the amnion) (Fig. 4.12I, I'), another, yet unknown factor must be involved in the rupture of the amnion. Further, Tc-Zen2 expression was observed only in the serosa during the entire lifespan of EEMs, which suggests that the potential function of *Tc-zen2* during 42 h-long time period of its expression could be either in the serosal cuticle maintenance or innate immunity.

5.3 Knockdowns of the *Tc-zen* paralogues differ in strength and cause different phenotypes

5.3.1 *Tc-zen1* and *Tc-zen2* knockdowns result in two distinct phenotypes

Following up on the previous work (van der Zee et al., 2005), I characterized in detail knockdown (KD) phenotypes of the *Tc-zen* genes. Consistent with the two distinct functions *Tc-zen* genes acquired, two different phenotypes after KD were observed. In the case of *Tc-zen1* KD, the phenotype is unambiguous as the serosal tissue does not differentiate from the blastoderm cells (Fig. 4.13, 4.14), since *Tc-zen1* role is in the specification of serosal tissue identity. What is rather intriguing is the fact that although one complete EEM is missing, the embryonic development is not compromised. It has been shown that in *T. castaneum* serosa provides innate immunity and protection against pathogen and desiccation (through secretion of the serosal cuticle) and that indeed the embryogenesis is disrupted after *Tc-zen1* KD unless relatively clean and humid environment is provided (Jacobs et al., 2013; Jacobs and van der Zee, 2013; Jacobs et al., 2014). Nonetheless, although the “save” environment is provided,

due to the missing EEM the morphogenetic movements required for proper progression through embryogenesis are altered. After *Tc-zen1* KD, the blastoderm cells, which would have under wildtype (WT) conditions acquire serosal fate, are re-specified to embryonic and amniotic cells. The expansion of amnion is indeed what “saves” the embryo from dying, because by performing dorsal closure it allows the embryo to withdraw the lateral flanks and close dorsally. Therefore, what in the end enables *Tc-zen1* KD embryos to survive, is the ability of the expanded amnion to compensate for the loss of serosa (Horn et al., 2015).

On the other hand, KD of *Tc-zen2* results in several cuticle phenotypes. *Tc-zen2* late function is in morphogenesis of mature EEMs and it plays major role during the membrane rupture stage (van der Zee et al., 2005; Hilbrant et al., 2016). After *Tc-zen2* KD, the membranes either do not rupture, or they rupture ectopically. When no rupture occurs, the embryo closes ventrally and fully folds itself inside-out. Partial inside-out phenotypes are observed as well, due to the ectopic rupture in different locations of membranes (Fig. 4.15).

Although *Tc-zen2* expression starts and peaks during early embryogenesis, the phenotypic manifestation in a form of blocked or ectopic rupture does not occur until late embryogenesis. Until the membrane rupture stage, the progression of embryonic development resembles the one of WT embryo (Thorsten Horn and Kristen Panfilio, personal communication).

5.3.2 Knockdown strength of *Tc-zen1*^{RNAi} is higher than the one of *Tc-zen2*^{RNAi}

Validation of KD strength after RNAi of *Tc-zen1* and *Tc-zen2* by RT-qPCR showed that KD of *Tc-zen1* is stronger than KD of *Tc-zen2*. After *Tc-zen1* KD the expression of *Tc-zen1* was downregulated across three biological replicates (BRs) (each of several technical replicates) to 7-20% of its WT expression (Table 4.1), whereas the *Tc-zen2* expression after *Tc-zen2* KD was downregulated only to 14-57% (Table 4.3). KD strength depends on several factors (e.g.: length of dsRNA (Wang et al., 2013)), and therefore, it is not rare that KD strength of two different genes is not the same. However, KD of *Tc-zen1* is very strong and persistent, which might be explained by the fact that *Tc-zen1* is specifying its own expression domain. Therefore, if from the beginning of the embryogenesis serosa is not specified, although *Tc-zen1*'s expression is only silenced, its expression domain is completely missing and it cannot be expressed. The only option for expression of *Tc-zen1* after its KD would be to change the expression domain, which has not been observed. In fact, the expression of *Tc-zen2*, which is exclusively expressed in serosa, is downregulated 3.5-fold after *Tc-zen1* KD, while after KD of *Tc-zen2*, the expression is downregulated only 2.2-fold (data not shown).

Although in some cases the expression of *Tc-zen2* transcript went down only to 40-57% (Table 4.3), phenotypic scoring revealed that among three BRs investigated, 74-92% of the cuticle phenotypes investigated fell into the gene-specific phenotypic categories (Fig. 4.16 and Table 4.4). Moreover, after *Tc-zen2* KD, the expression of Tc-Zen2 protein was not observed in early, or in late embryogenesis (Fig. 4.17), which suggests that the KD strength of *Tc-zen2*, although not as strong as *Tc-zen1*, is sufficient for successful silencing of the protein expression.

5.3.3 *Tc-zen* paralougues are mutual downstream targets

Investigation of expression levels of *Tc-zen1* and *Tc-zen2* in the KD samples of its paralogue revealed rather surprising regulatory interactions between the *Tc-zen* genes. Although downregulation of *Tc-zen2* after *Tc-zen1* KD (Fig. 4.18A) was anticipated due to the missing serosa phenotype, the incomplete loss of *Tc-zen2* expression was unexpected. Despite the absence of its expression domain, *Tc-zen2* expression was downregulated only to 35% of its WT expression after *Tc-zen1* KD (Fig. 4.21). As mentioned above, in *Tc-zen1* KD samples, *Tc-zen2* expression level drops more than in *Tc-zen2* KD samples. Therefore, the remaining expression is intriguing, given the fact that the *Tc-zen2* expression domain (serosa) is completely missing. The pattern of the remaining *Tc-zen2* expression was investigated by *in situ* hybridization, but did not reveal any alternative expression domains, as no *Tc-zen2* expression pattern in *Tc-zen1*^{RNAi} embryos was observed (data not shown), probably due to undetectable levels. Therefore, the remaining expression of *Tc-zen2* without presence of its only expression domain remains still unclear.

The upregulation of *Tc-zen1* after *Tc-zen2* KD revealed unknown potential regulatory role of *Tc-zen2* during early embryogenesis. The expression of *Tc-zen1* is upregulated almost 3-fold after KD of its paralogue (Fig. 4.18B). *Tc-zen1* upregulated expression was investigated in the *Tc-zen2*KD embryos and showed that not only its expression levels are higher than in WT embryos, but its expression domain expands to the whole serosa during the serosal window stage (Fig. 4.19F). These results suggest that *Tc-zen2* is repressing *Tc-zen1* during early embryogenesis. Additionally, the repression of *Tc-zen1* by *Tc-zen2* explains the abrupt switching off of *Tc-zen1* expression during early embryogenesis (Fig. 4.7B). This surprising result suggests that during evolution *Tc-zen1* might have recruited *Tc-zen2* for its own repression.

5.4 *Tc-zen*s' functions are separated to early and late development

5.4.1 Knockdown of *Tc-zen2* in early stages does not have a robust transcriptional effect

Both *Tc-zen* genes reach the expression peak during early embryogenesis, but only *Tc-zen1* has early function, while *Tc-zen2* function takes place 42 h later. In order to identify transcriptional targets of the *Tc-zen* paralougues during developmental stages, when their expression is the highest, I performed RNA-sequencing (RNA-seq) after KD of both paralougues. Subsequent principal component and differential expression analyses (PCA and DE) revealed that KD of *Tc-zen1* has a profound effect on its downstream transcriptional control, what is consistent with *Tc-zen1* having early function. The separation between the KD and WT samples was clear (Fig. 4.22) and consistent with high number of identified candidate target genes (Table 4.5).

On the other hand, KD of *Tc-zen2* during early stages had very little impact on its downstream transcriptional control. According to the PCA, both the first and the second principal components failed to separate the *Tc-zen2* KD and WT samples (Fig. 4.22), suggesting little to no effect of *Tc-zen2*^{RNAi}. Further, low number of identified transcriptional

targets (13 times less than for *Tc-zen1*) (Table 4.5) is consistent with PCA results and ultimately with *Tc-zen2* having function during late embryogenesis.

5.4.2 *Tc-zen2* has subtle early regulatory role

The RNA-seq after RNAi results from stages of early embryogenesis showed that while *Tc-zen1* function takes place during the time its expression is the highest, time points of *Tc-zen2*'s highest expression and function do not coincide. This fact raises the question about *Tc-zen2* function during early embryogenesis. As previously discussed, I have observed a negative regulatory feedback loop between *Tc-zen1* and *Tc-zen2* when upon *Tc-zen2* KD the expression of *Tc-zen1* was upregulated almost 3-fold compared to its WT expression (Fig. 4.18B). Therefore, one possible reason for necessity to express *Tc-zen2* early could be explained by subtle early regulatory role it plays in the repression of *Tc-zen1*.

In addition, an early Tc-Zen2 function was described by Schoppmeier et al., 2009. It was reported that Tc-Zen2 together with Tc-Mex3 are repressing Tc-Caudal in the anterior of *T. castaneum* embryos during the establishment of the anterior-posterior axis: Tc-Mex3 in embryonic cells and Tc-Zen2 in EE cells (in the serosa). The stage, during which the translational repression function of Tc-Zen2 was described, corresponds to the stage of *Tc-zen2* expression peak. However, morphological defects along anterior-posterior axis after *Tc-zen2* KD during early embryogenesis were not described.

Nonetheless, discovery of the translational repression function of Tc-Zen2 has an exciting implication for the evolution of *zen* genes in general. As previously mentioned, *bicoid* (*bcd*) gene has evolved through two rounds of duplication of an ancestral *Hox3/zen* orthologue (reviewed in McGregor, 2005). In *D. melanogaster*, *bcd* mRNA is maternally supplied and the protein plays important role in axis patterning through repression of Caudal in the anterior of the embryo (Stauber et al., 1999). Interestingly, *D. melanogaster* Bcd is able to repress *T. castaneum* orthologue of Caudal (Wolff et al., 1998). This combined evidence suggests that the translational repression function of Tc-Zen2 and Dm-Bcd could have been a feature of Hox3/Zen homeodomain and preceded the divergence of Bcd (Schoppmeier et al., 2009).

5.4.3 Defects in extraembryonic development were not observed after knockdown of *Tc-zen1* and *Tc-zen2* candidate targets of early embryogenesis

In order to confirm that the genes designated by RNA-seq after RNAi with subsequent DE analysis approach, are indeed targets of either *Tc-zen1*, or *Tc-zen2*, I investigated their expression patterns and potential function. The list of the first investigated 32 candidates was based on the results from quickNGS pipeline, which generated high number of genes falsely designated as differentially expressed (false positives). Therefore, in the first round of the miniscreen, out of 32 chosen candidates, only 17 had the potential to be target genes of either *Tc-zen1*, or *Tc-zen2*. Out of these 17 genes, 11 showed expression pattern in the serosa,

suggesting that these genes are actual downstream target genes of *Tc-zen1* (10 genes) and *Tc-zen2* (1 gene) (Fig. 4.23 and 4.24).

Since the specific expression pattern of only one of the 6 potential *Tc-zen2* targets was observed in the serosa, I decided to add four genes to the list of investigated candidate targets. In order to avoid any complications with false positive results, selection of these four genes was based on the results obtained from the custom pipeline developed within this project (see section 3). Nonetheless, only two genes showed specific expression pattern in the serosa. Altogether, specific serosal expression patterns were observed for 13 (out of 21) candidate target genes of the *Tc-zen* paralogues. The rather low success rate might have been caused by non-functional hybridization probes. Many of these probes were redesigned, but usually the synthesis of the second probe failed already due to non-working primer pairs and design of the third primer pair was not pursued.

Nonetheless, the potential target genes' function was further investigated by RNAi. Unfortunately, out of the 21 genes investigated, only one showed phenotype (empty egg) and one caused 100% post-injection lethality of pupae (Fig. 4.23 and 4.24). I did not observe any phenotype connected with the defects in EE development, since all the KD larvae resembled WT morphology. However, the morphology of WT larvae was observed also after *Tc-zen1* KD, even though one of the EEMs is completely missing. Therefore, to further investigate detailed embryonic and EE development after KD of identified targets of *Tc-zen1* and *Tc-zen2*, live imaging should be employed in the future.

Another option why KD of potential target genes did not result in any phenotype could be explained by the nature of the candidate target genes. By definition (target of *Tc-zen1* and *Tc-zen2*), these genes can function only in serosa, therefore presumably in its maintenance, or in the maintenance of the serosal cuticle, or in the defense against pathogens. However, many of the genes responsible for chitin cuticle synthesis have been shown to cooperate between each other, and therefore, their individual KD does not result in any phenotype (T. Horn, personal communication). Whether *Tc-zen1* and *Tc-zen2* candidate target genes are playing important roles in protecting the embryo from outer environment should be investigated in the future, preferably by exposing the KD embryos in a septic arid environment.

5.4.4 *Tc-zen2* does not copy *Tc-zen1*'s function during early embryogenesis

In order to evaluate, whether the early expression of *Tc-zen2* might also be necessary due to its involvement in the same functions as *Tc-zen1* has, I examined, to which degree the target genes of *Tc-zen1* and *Tc-zen2* are shared between the paralogues. Since *Tc-zen2* has late function and its KD did not have robust impact on its transcriptional control during early embryogenesis, I considered the option that *Tc-zen2* might copy *Tc-zen1*'s function, but on the lower levels. Therefore, I have investigated the overlap of differentially expressed genes between *Tc-zen1* and *Tc-zen2* with different cut-off criteria for *Tc-zen2* candidate targets (Fig. 4.25). Moreover, precautions due to the fact that two neighboring developmental stages were sequenced, had to be taken into account. All the genes, compared between *Tc-zen1* and *Tc-*

zen2, were either upregulated, downregulated or non-differentially expressed between the two sequenced stages (WT shift) with the P_{adj} values less than 0.1.

Comparative analysis showed that *Tc-zen1* and *Tc-zen2* do not share any strongly differentially expressed genes, which are upregulated, downregulated or not differentially expressed within the WT shift (Fig. 4.25). The only gene, which is strongly differentially expressed after KDs of both *Tc-zen1* and *Tc-zen2*, is in fact *Tc-zen2*, due to the silencing of *Tc-zen2* by *Tc-zen2*^{RNAi} and due to the absence of its expression domain (serosa) after *Tc-zen1*^{RNAi} (Fig. 4.18A). With the lower threshold criteria for *Tc-zen2* targets, 45 genes were designated as shared between the paralogues (Table S1). Retrieving functional profile about the shared genes was possible only for 31 genes, which homologues with assigned gene ontology (GO) terms exist in *D. melanogaster*. Nonetheless, based on the information about these GO terms, the 31 shared genes have potential function in various biological processes. GO terms that were assigned at least to two *D. melanogaster* homologues of the shared genes suggest that some of these genes are involved in transporter processes, have potential to bind ions and play role in chitin synthesis and oxidase/reductase events. These results suggest that *Tc-zen1* and *Tc-zen2* share only very small number of genes during early embryogenesis and that most likely *Tc-zen2* is not extensively involved in the same biological functions as *Tc-zen1*, but has rather acquired functions independent from its paralogue.

5.4.5 Functional profile of *Tc-zen2* candidate target genes of late development was retrieved

To identify transcriptional targets of *Tc-zen2* during late development, when its morphogenesis function takes place, I performed RNA-seq of WT and *Tc-zen2* KD samples collected before and during the membrane rupture stage. The subsequent DE analysis revealed high number of candidate target genes of *Tc-zen2* during late development (Table 4.6). In fact the number of differentially expressed genes in late developmental stages after *Tc-zen2* KD was 18 times higher than the number of differentially expressed genes in early developmental stages (Table 4.5). This result supports the separation of *Tc-zen2* morphogenesis function to late embryogenesis.

However, before and during the membrane rupture stage, the number of embryonic cells is considerably higher than the number of EE cells. Therefore, before it was possible to make any final conclusions, the functional profile of identified differentially expressed genes had to be obtained. GO term analysis of differentially expressed genes before and during the membrane rupture was performed by blasting against two different databases: non-redundant (nr) and *Drosophila* (Fig. 4.27 and Table 4.7). Unfortunately, it was not possible to retrieve functional profile of all the identified differentially expressed genes, because not all of them have their respective orthologues or homologues described and annotated in the blast databases, which were used. In fact, retrieving GO term information of almost 40% (blast against the nr database) and 50% (blast against the *Drosophila* database) of differentially expressed genes was not possible, which indeed represents a substantial information loss. Reason, why low percentage of blast hits were obtained when blasting against the *Drosophila*

database, might be explained by the fact that *D. melanogaster* genes are in general much longer, and therefore, the overall sequence homology to *Tribolium* genes is much lower.

Nonetheless, the functional profile of the differentially expressed genes, whose orthologues and homologues do exist in other insect species or *D. melanogaster*, was obtained. Under the assumption that after *Tc-zen2* KD genes with function in EE development were designated as differentially expressed, representation of certain GO terms was expected. Based on the observations of altered development undergoing after KD of *Tc-zen2*, GO terms describing biological processes corresponding to response to stress, cuticle formation, cytoskeleton reorganization, morphogenesis of epithelium, cell adhesion, regulation of gene expression and transmembrane transport were expected to be assigned to the identified differentially expressed genes. These GO terms were specifically searched for in the output results of GO term analysis. When blasted against the nr database, approx. 20-30% of the differentially expressed genes were assigned GO terms falling into the above mentioned categories of interest. On the other hand, when blasted against the *Drosophila* database, almost 60% of the differentially expressed genes fell into the categories of interest.

Speculations, about why higher percentage of the differentially expressed genes falls into the categories of interest, when blasted against the *Drosophila* database, are rather inconclusive. One possibility is that, in *D. melanogaster*, higher number of orthologues or homologues has potential EE roles, which reflects the fact that *D. melanogaster* possesses the amnioserosa. However, compared to *T. castaneum*, the amnioserosa represents only a reduced form of EEMs. Nonetheless, although “insect taxonomy filter” was applied during GO term analysis performed with the nr database, we cannot ensure that insect species, against which *Tribolium* genes were blasted, possess any kind of EE tissue.

So far, it was not possible to retrieve functional profile of complete datasets of the differentially expressed genes and, moreover, overt differences between results obtained by blasting against two different blast databases were observed. Nonetheless, based on the results from blasting against the *Drosophila* database, it is possible to conclude that, after *Tc-zen2* KD, almost one quarter of the differentially expressed genes reflects altered EE development.

In addition, identification of GO terms, to which most of the differentially expressed genes were assigned to by blasting against both databases (Table 4.8), revealed that many GO terms that were not grouped to the categories of interest might still represent the altered EE development. For example, many of the differentially expressed genes were assigned to GO terms describing metal ion binding and transport, or channel activity, which might reflect altered transmembrane transport through EEMs between embryo and outer environment, either due to the ectopic rupture or delayed embryonic development. Many genes were also assigned to GO terms with serine peptidase activity, which are known to be part of signaling pathways playing important roles in embryonic development, immune system or wound healing of arthropods (references cited in Veillard et al., 2016). Lastly, many genes were assigned to GO terms describing nervous system development and cell differentiation. Before and during the membrane rupture stage, the nervous system development is an ongoing process in the embryo, which could potentially involve higher levels of cell differentiation.

The reason, why these genes are differentially expressed, could be due to the delay in development caused by the stress from the ectopic rupture after *Tc-zen2* KD.

To conclude, it is possible that with the future detailed analysis of GO terms, which did not fall into the categories of interest, we might identify much more than one quarter of genes with potential role in EE development. Therefore, it is fair to assume that the genes, which were designated as differentially expressed after *Tc-zen2* KD, are indeed potential target genes of *Tc-zen2* during late development. Nonetheless, in order to draw any final conclusions, future analysis of function of the identified differentially expressed genes is necessary.

5.4.6 RT-qPCR miniscreen confirmed results obtained from differential expression analysis

Several *Tc-zen2* candidate target genes from the pre- and the post-rupture developmental stages, which were assigned to any of the GO terms from the categories of interest, were selected for further miniscreen. Due to the presence of the serosal cuticle, which blocks penetration of any hybridization probes in late developmental stages, it was not possible to describe expression domains of the selected *Tc-zen2* candidate target genes. In order to evaluate the obtained DE analysis data, I performed RT-qPCR miniscreen. Results from the miniscreen confirmed that chosen candidates are indeed differentially expressed and the obtained expression levels were consistent with the direction of DE assigned to the candidate targets during DE analysis. Particularly intriguing cases of two genes, which were downregulated in the pre-rupture stage and upregulated in the post-rupture stage, suggest that in the future it is important to investigate *Tc-zen2* late candidate target genes' function.

5.5 Distinct *Tribolium zen* functions most likely arose through sub-functionalization

Results obtained within this project suggest that two distinct functions of *Tc-zen1* and *Tc-zen2* might have arisen through the process of sub-functionalization. Besides the fact that KD of each of the paralogues results in distinct phenotype, the observation that only small number of genes is regulated by both *Tc-zen* paralogues during early embryogenesis suggest division of their functions during early development.

Sub-functionalization hypothesis assumes that duplication in the Hox3 locus preceded the split of functions and that subsequently one gene copy retained specification function along the anterior-posterior axis (although now in EE tissue, *Tc-zen1*) and the second copy acquired late morphogenesis function (*Tc-zen2*). This hypothesis further assumes that specification role was lost in basally branching species, while the morphogenesis function was lost in higher holometabolous species (Fig. 1.4).

Consistent with the hypothesis, in basally branching species (Hemiptera), which possess only one copy of *zen* gene, early expression of *zen* was observed, despite its late

morphogenesis function (Dearden et al., 2000; Panfilio, 2009). This observation implies that in basally branching species specification function was lost, but the early ancestral expression has been retained. Further, the loss of late morphogenesis function in higher holometabolous species (Diptera) could be explained by its redundancy due to the reduced EEMs. Since *D. melanogaster* possesses reduced form of EE tissue, the amnioserosa, which covers only the dorsal part of the yolk, the late morphogenesis function during membrane rupture is not required, simply, because no rupture occurs.

What remains still unclear is the extent of the sub-functionalization, in a sense, how disconnected the functions of *Tc-zen* paralogues are and to which extent can they substitute each other's functions. Based on the results from *Tc-zen1* KD, it is clear that *Tc-zen2* is not able to take over the specification function, because no serosal tissue was observed in *Tc-zen1*^{RNAi} embryos. This might stem from the fact, that *Tc-zen2* is expressed only in the serosa and unlike *Tc-zen1*, it is not able to specify its own expression domain. Double KD experiments of both *Tc-zen1* and *Tc-zen2* resulted in the phenotype indistinguishable from the one obtained after *Tc-zen1* KD alone (van der Zee et al., 2005), because of the re-specification of blastoderm to embryonic and amniotic cells and subsequent alteration of morphogenetic movements.

On the other hand, it is still necessary to further investigate *Tc-zen2*'s function dependency on *Tc-zen1*. The discovery of surprising regulatory feedback loop between the *Tc-zen* paralogues implies that the transcriptional regulation of *Tc-zen* genes during early embryogenesis is connected. Since I showed that *Tc-zen2* represses *Tc-zen1* expression, it would be of high interest to investigate potential effects of ectopically overexpressed *Tc-zen1*. If the excess of *Tc-zen1* mRNA shows the ability to drive the inside-out phenotype, then the *Tc-zen2* function could be considered dependent on *Tc-zen1* and would, to a certain degree, explain necessity of early expression of both *Tc-zen* paralogues.

5.6 Changes in protein sequence and features enabled switch from Hox3 to Zen

Based on the results obtained from the research in the evo-devo field in the last four decades, it is fair to assume that what drives the gene function change, and ultimately evolution of morphological diversity, are the changes in transcriptional regulation through alteration of *cis*-regulatory regions (Carroll, 2000; Tautz, 2000; Davidson, 2001). Therefore, it is also fair to assume that what partially drove the switch from canonical embryonic *Hox3* to insect EE *zen* and the subsequent functional divergence of the paralogues, were the changes in regulatory regions of *Hox3* genes.

However, in the case of the switch of *Hox3* to *zen*, it is necessary to consider the role of coding sequence divergence as well. Even subtle changes in protein sequences can have strong impact on the function of the protein, especially if the protein is involved in transcriptional regulation (Hsia and McGinnis, 2003). The changes in the sequence could severely affect the binding potential of transcription factors and thus alter their downstream targets. An example of rather small deletion in the protein sequence, which caused strong effect on limb development, is the example of Ultrabithorax in hexapod *Drosophila* and

multi-limbed crustacean *Artemia*. In *Drosophila*, Ultrabithorax is responsible (together with Abdominal-A) for repression of limb development. It is believed that the loss of threonines and serines from C-terminus of Ultrabithorax proteins, during the transition from *Artemia*-like ancestor to insects, contributed to the macroevolutionary change in limb number between hexapods and multi-limbed Arthropods. By losing two main amino acids, which possess phosphorylation sites, hexapods altered the ancestral function of Ultrabithorax, which can now repress the limb formation (Averof and Akam, 1995; Ronshaugen et al., 2002). The example of Ultrabithorax represents case of modest divergence in function after the change of protein sequence.

The example of complete functional divergence after sequence alteration is in fact the example of Hox3, Zen and Bcd. Together with the alteration of regulatory regions, substantial changes in the amino acid sequence of these transcription factors (Fig. 1.1) contributed to the possibility of switching developmental roles. The change of the sequence resulted in loss of the hexapeptide motif, typical of Hox genes, through which binding of cofactor Extradenticle is possible and whose loss correlates with EE expression of *zen* (Falciani et al., 1996; Panfilio and Akam, 2007). Interestingly, the sequence of Bcd diverged even more than Zen (Fig. 1.1) and acquired yet another function, now back in the specification of anterior-posterior axis (Stauber et al., 1999), however only in higher Diptera.

It is understandable that the change in regulatory regions together with protein sequence alteration is able to drive the switch of developmental functions, but what is rather intriguing is the fact that loss of a canonical Hox function did not cause a deleterious phenotype, but became advantageous, got fixed during evolution of winged insects and allowed them to colonize land. A proposed hypothesis of possibility to lose one of the most important genes functioning in body patterning without drastically altering embryonic development assumes that this function must have been redundant (Telford and Thomas, 1998a). The redundancy in function can be explained by overlapping expression domains with neighboring Hox genes in species with canonical *Hox3* function. In fact, it has been shown that expression domain of *Hox3* gene in mite *Archezogozetes longisetosus* overlaps with that of *Hox2* (*Proboscipedia*) (Telford and Thomas, 1998a) and *Hox4* (*Deformed*) (Telford and Thomas, 1998b). This overlapping expression of domains could have resulted in *Hox3* functional redundancy and might have triggered new function emergence.

5.7 Conclusion

Detailed analyses of transcriptional and translation regulation of the *Tc-zen* paralogues presented in this thesis revealed several differences in regulation of *Tc-zen1* and *Tc-zen2* expression. Although, there is only a slight offset in the expression of the *Tc-zen* paralogues, with *Tc-zen1* being expressed first, after the peak expression of both *Tc-zen* genes, *Tc-zen1* switches off its expression abruptly, while *Tc-zen2*'s expression gradually wanes. Moreover, Tc-Zen2 protein profiles throughout embryogenesis imply that this waning transcript expression is persistent and sufficient for retention of the protein expression until the stage, when its function takes place. Differences in the timing and character of *Tc-zen1* and *Tc-zen2* expression suggest that the *Tc-zen* paralogues are regulated by different regulatory inputs. This was further confirmed by *in silico* analysis of their promoter sequences, which showed no conservation of these regions. Therefore, it seems that the changes on upstream level, in the regulation of *Tc-zen* genes' transcription, contributed to the acquirement of paralogues' distinct functions, through alteration of the regulation of their spatial-temporal expression.

Identification of downstream transcriptional targets of *Tc-zen1* and *Tc-zen2* during early embryogenesis revealed strong difference in impact of transcriptional control by the paralogues during early embryonic development. Substantially higher number of target genes was identified after KD of *Tc-zen1* than after KD of *Tc-zen2*, what is consistent with the separation of their functions to early (*Tc-zen1*) and late (*Tc-zen2*) development. The results from subsequent analysis of level of overlap between *Tc-zen* genes' identified targets suggest that the paralogues do not share profound number of target genes during early embryonic development. Further, identification of high number of candidate target genes of *Tc-zen2* during late development is consistent with its late function, which was separated from its early expression. The GO term analysis suggests that many of these identified candidate targets play roles in EE development. The high impact on transcriptional control of *Tc-zen1* during early embryogenesis and of *Tc-zen2* during late development implies that the distinct functions of *Tc-zen* paralogues were separated to two different developmental stages. These results support the sub-functionalization hypothesis, however the level of dependency of *Tc-zen2*'s function on *Tc-zen1*'s expression needs to be investigated further. Nonetheless, the difference in the downstream transcriptional targets of *Tc-zen1* and *Tc-zen2* suggests that alterations on downstream levels contributed to the acquirement of paralogues' distinct functions, as they regulate expression of different genes in two developmental stages.

Protein sequence alignment clearly shows that Tc-Zen1 and Tc-Zen2 diverged in their sequence, which might have affected their binding specificity and ultimately possibility of regulating the same downstream targets. Therefore, the changes in the sequence of paralogues could have as well contributed to their acquirement of two distinct functions.

In conclusion, these diverse lines of evidence suggest that acquirement of the two distinct roles of the *Tc-zen* paralogues was most likely possible due to the combination of changes that occurred on three different levels: the changes in upstream regulation of *Tc-zen* genes expression, in regulation of downstream transcriptional targets and the changes within the paralogues themselves collectively contributed to the functional divergence of *Tc-zen1* and *Tc-zen2*.

6 OUTLOOK

Results obtained within the presented project provide a ground foundation for future comparative studies of *Hox3* insect orthologues and their transcriptional targets. But first, it would be of high relevance to continue investigations in *T. castaneum*. Particularly, description of function of genes, which were identified as potential targets of *Tc-zen2* during late embryogenesis, will enable possible discovery of factors playing roles during membrane rupture. Although detailed description of membrane rupture process was recently reported (Hilbrant et al., 2016), it is not yet known, which genes are involved in this important developmental event.

Further, investigation of functional relevance of identified conserved non-coding regions between the *Hox3* loci of three closely related congenics of *T. castaneum* would permit to assess, which fine-tuned transcriptional regulation of *Tc-zen1* and *Tc-zen2* could be responsible for the acquirement of the *Tc-zen* paralogues' diverse functions, which were separated to two developmental stages. Recent establishment of *T. castaneum* immortalized cell lines (Silver et al., 2014) would provide a homologues system for functional tests of the identified regulatory regions. It has been reported that these cell lines were established from epidermal tissues, which express genes involved in chitin synthesis as well as genes involved in immunity. Serosa expresses the same type of genes along with the *Tc-zen* paralogues, therefore it would be meaningful to explore whether the transcriptional apparatus regulating *Tc-zen* genes is potentially present in these cell lines. If so, new possibilities of functional testing of regulatory regions could be explored.

To place the presented study in evolutionary perspective, comparisons with other species will be necessary. Fortunately, direct comparison with *O. fasciatus* is possible on many different levels, mainly due to the extensive knowledge about its EE development and *Of-zen* orthologue, which has been already obtained (eg.: Panfilio et al., 2006; Panfilio, 2008, 2009; Panfilio and Roth, 2010). One immediate possibility is to investigate behavior of *O. fasciatus* orthologues of *Tc-zen1* and *Tc-zen2* target genes. Nowadays it is fairly easy to obtain orthologue information due to the well-established OrthoDB protein orthology database (Waterhouse et al., 2013). Moreover, vast amount of transcriptome information from several developmental stages of *O. fasciatus* was recently obtained by our lab. The custom pipeline developed within this project is transferable to *O. fasciatus* data. Therefore, based on the availability of the above mentioned resources, it should be straight forward to obtain information about developmental stages, in which *O. fasciatus* orthologues of *Tc-zen* genes' targets are expressed.

Besides the bioinformatic approach, obtaining transcript and protein expression profiles of *Of-zen* would enable a direct comparison with expression profiles of the *Tc-zen* genes, which were generated within the presented project. RT-qPCR in *O. fasciatus* has been recently established through my joined efforts with other colleagues and preliminary RT-qPCR runs have been performed, suggesting successful method establishment. Therefore, obtaining transcript expression profile of *Of-zen* during early and late development should be straight forward. Due to the fact that *Of-zen* function was described in late morphogenesis, but

it's expression was observed during early embryogenesis (like for *Tc-zen2*) (Panfilio et al., 2006; Panfilio, 2009), a determination of exact start of *Of-zen* expression could provide insights to whether early expression of *zen* is of ancestral origin. Attempts to raise the Of-Zen antibody have been performed within this project, however the preliminary tests showed rather unspecific signal (data not shown) and further investigations of immune sera still need to be performed.

Another potential experimental idea would offer an insight into the functional interchangeability of *Of-zen* orthologue. One could KD the endogenous *Tc-zen1* via pRNAi and investigate whether the specification function of *Tc-zen1* could be rescued by *Of-zen* after the ectopic embryonic injection of its capped mRNAs.

In order to further conceptually place the presented study in larger evolutionary perspective, it would be necessary to perform orthology inference of *Tc-zen1* and *Tc-zen2* target genes across the gene sets of hemipteran species, which show no evidence of *Hox3* duplication. In the next step, this analysis can be extended to other species, whose genomes were recently sequenced within the i5K project (i5K Consortium, 2013) and are currently of draft-quality. OrthoDB (Waterhouse et al., 2013) provides protein orthology information for more than 100 species and serves as a resourceful platform for multi-species comparisons. Ultimately, results from this large-scale comparative analysis should provide insights into the functional divergence of targets of transcription factors encoded by Hox3 loci.

7 REFERENCES

- Aldaz, S., Escudero, L.M., Freeman, M., 2010. Live imaging of *Drosophila* imaginal disc development. Proceedings of the National Academy of Sciences of the United States of America 107, 14217-14222.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402.
- Andrews, S., 2014. FASTQC. A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Averof, M., Akam, M., 1995. Hox genes and the diversification of insect and crustacean body plans. Nature 376, 420-423.
- Bao, R., Fischer, T., Bolognesi, R., Brown, S.J., Friedrich, M., 2012. Parallel duplication and partial subfunctionalization of beta-catenin/armadillo during insect evolution. Molecular biology and evolution 29, 647-662.
- Benton, M.A., Akam, M., Pavlopoulos, A., 2013. Cell and tissue dynamics during *Tribolium* embryogenesis revealed by versatile fluorescence labeling approaches. Development 140, 3210-3220.
- Berghammer, A.J., Klingler, M., Wimmer, E.A., 1999. A universal marker for transgenic insects. Nature 402, 370-371.
- Berghammer, A.J., Weber, M., Trauner, J., Klingler, M., 2009. Red flour beetle (*Tribolium*) germline transformation and insertional mutagenesis. Cold Spring Harb Protoc 2009, pdb prot5259.
- Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1, 307-310.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.
- Bradley, R.K., Li, X.Y., Trapnell, C., Davidson, S., Pachter, L., Chu, H.C., Tonkin, L.A., Biggin, M.D., Eisen, M.B., 2010. Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related *Drosophila* species. PLoS Biol 8, e1000343.
- Brown, S.J., Shippy, T.D., Miller, S., Bolognesi, R., Beeman, R.W., Lorenzen, M.D., Bucher, G., Wimmer, E.A., Klingler, M., 2009. The red flour beetle, *Tribolium castaneum* (Coleoptera): a model for studies of development and pest biology. Cold Spring Harb Protoc 2009, pdb emo126.
- Bucher, G., Scholten, J., Klingler, M., 2002. Parental RNAi in *Tribolium* (Coleoptera). Current biology : CB 12, R85-86.

- Busby, M.A., Stewart, C., Miller, C.A., Grzeda, K.R., Marth, G.T., 2013. Scotty: a web tool for designing RNA-Seq experiments to measure differential gene expression. *Bioinformatics* 29, 656-657.
- Calderwood, M.S., Gannoun-Zaki, L., Wellems, T.E., Deitsch, K.W., 2003. *Plasmodium falciparum* var genes are regulated by two regions with separate promoters, one upstream of the coding region and a second within the intron. *J Biol Chem* 278, 34125-34132.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421.
- Carroll, S.B., 2000. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* 101, 577-580.
- Chen, G., Handel, K., Roth, S., 2000. The maternal NF-kappaB/dorsal gradient of *Tribolium castaneum*: dynamics of early dorsoventral patterning in a short-germ beetle. *Development* 127, 5145-5156.
- Coleman, K.G., Poole, S.J., Weir, M.P., Soeller, W.C., Kornberg, T., 1987. The *invected* gene of *Drosophila*: sequence analysis and expression studies reveal a close kinship to the *engrailed* gene. *Genes & development* 1, 19-28.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674-3676.
- Damen, W.G.M., Tautz, D., 1998. A Hox class 3 orthologue from the spider *Cupiennius salei* is expressed in a Hox-gene like fashion. *Development genes and evolution* 208, 586-590.
- Davidson, E.H., 2001. Genomic regulatory systems: development and Evolution. Elsevier Science.
- Dearden, P., Akam, M., 1999. Developmental evolution: Axial patterning in insects. *Current biology* : CB 9, R591-594.
- Dearden, P., Grbic, M., Falciani, F., Akam, M., 2000. Maternal expression and early zygotic regulation of the *Hox3/zen* gene in the grasshopper *Schistocerca gregaria*. *Evolution & development* 2, 261-270.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21.
- Donitz, J., Schmitt-Engel, C., Grossmann, D., Gerischer, L., Tech, M., Schoppmeier, M., Klingler, M., Bucher, G., 2015. iBeetle-Base: a database for RNAi phenotypes in the red flour beetle *Tribolium castaneum*. *Nucleic Acids Res* 43, D720-725.
- Du Val, P.N.C.J., Fairmaire, L., Migneaux, J., 1863. Manuel entomologique: genera des coléoptères d'Europe comprenant leur classification en familles naturelles, la description de tous les genres. A. Deyrolle.

- Falciani, F., Hausdorf, B., Schroder, R., Akam, M., Tautz, D., Denell, R., Brown, S., 1996. Class 3 Hox genes in insects and the origin of *zen*. *Proceedings of the National Academy of Sciences of the United States of America* 93, 8479-8484.
- Farnesi, L.C., Menna-Barreto, R.F., Martins, A.J., Valle, D., Rezende, G.L., 2015. Physical features and chitin content of eggs from the mosquito vectors *Aedes aegypti*, *Anopheles aquasalis* and *Culex quinquefasciatus*: Connection with distinct levels of resistance to desiccation. *J Insect Physiol* 83, 43-52.
- Ferguson, L., Marletaz, F., Carter, J.M., Taylor, W.R., Gibbs, M., Breuker, C.J., Holland, P.W., 2014. Ancient expansion of the hox cluster in lepidoptera generated four homeobox genes implicated in extra-embryonic tissue formation. *PLoS Genet* 10, e1004698.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531-1545.
- Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I., 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32, W273-279.
- Gilles, A.F., Schinko, J.B., Averof, M., 2015. Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle *Tribolium castaneum*. *Development* 142, 2832-2839.
- Gramates, L.S., Marygold, S.J., Santos, G.D., Urbano, J.M., Antonazzo, G., Matthews, B.B., Rey, A.J., Tabone, C.J., Crosby, M.A., Emmert, D.B., Falls, K., Goodman, J.L., Hu, Y., Ponting, L., Schroeder, A.J., Strelets, V.B., Thurmond, J., Zhou, P., the FlyBase, C., 2017. FlyBase at 25: looking to the future. *Nucleic Acids Res* 45, D663-D671.
- Grimaldi, D., Engel, M.S., 2005. *Evolution of the Insects*. Cambridge University Press.
- Guillen, I., Mullor, J.L., Capdevila, J., Sanchez-Herrero, E., Morata, G., Guerrero, I., 1995. The function of *engrailed* and the specification of *Drosophila* wing pattern. *Development* 121, 3447-3456.
- Gustavson, E., Goldsborough, A.S., Ali, Z., Kornberg, T.B., 1996. The *Drosophila engrailed* and *invected* genes: partners in regulation, expression and function. *Genetics* 142, 893-906.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41, 95-98.
- Handel, K., Grunfelder, C.G., Roth, S., Sander, K., 2000. *Tribolium* embryogenesis: a SEM study of cell shapes and movements from blastoderm to serosal closure. *Development genes and evolution* 210, 167-179.
- Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., Wiehe, T., 2012. The chromatin insulator CTCF and the emergence of metazoan diversity. *Proceedings of the National Academy of Sciences of the United States of America* 109, 17507-17512.
- Hidalgo, A., 1994. Three distinct roles for the *engrailed* gene in *Drosophila* wing development. *Current biology : CB* 4, 1087-1098.
- Hilbrant, M., Horn, T., Koelzer, S., Panfilio, K.A., 2016. The beetle amnion and serosa functionally interact as apposed epithelia. *Elife* 5.

- Hinton, H.E., 1948. A synopsis of the genus *Tribolium Macleay*, with some remarks on the evolution of its species-groups (Coleoptera, Tenebrionidae). *Bull Entomol Res* 39, 13-55.
- Horn, T., Hilbrant, M., Panfilio, K.A., 2015. Evolution of epithelial morphogenesis: phenotypic integration across multiple levels of biological organization. *Front Genet* 6, 303.
- Hsia, C.C., McGinnis, W., 2003. Evolution of transcription factor function. *Curr Opin Genet Dev* 13, 199-206.
- <http://mussa.caltech.edu/mussa>.
- <https://usegalaxy.org/>.
- Hughes, C.L., Kaufman, T.C., 2002a. Exploring the myriapod body plan: expression patterns of the ten Hox genes in a centipede. *Development* 129, 1225-1238.
- Hughes, C.L., Kaufman, T.C., 2002b. Hox genes and the evolution of the arthropod body plan. *Evolution & development* 4, 459-499.
- Hughes, C.L., Liu, P.Z., Kaufman, T.C., 2004. Expression patterns of the rogue Hox genes *Hox3/zen* and *fushi tarazu* in the apterygote insect *Thermobia domestica*. *Evolution & development* 6, 393-401.
- i5K Consortium, 2013. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J Hered* 104, 595-600.
- Illumina ©Inc, 2010. Illumina Sequencing Technology. Highest data accuracy, simple workflow, and a broad range of applications. in ed, 5.
- Jacobs, C.G., Rezende, G.L., Lamers, G.E., van der Zee, M., 2013. The extraembryonic serosa protects the insect egg against desiccation. *Proceedings. Biological sciences / The Royal Society* 280, 20131082.
- Jacobs, C.G., Spaink, H.P., van der Zee, M., 2014. The extraembryonic serosa is a frontier epithelium providing the insect egg with a full-range innate immune response. *Elife* 3.
- Jacobs, C.G., van der Zee, M., 2013. Immune competence in insect eggs depends on the extraembryonic serosa. *Developmental and comparative immunology* 41, 263-269.
- Juven-Gershon, T., Kadonaga, J.T., 2010. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental biology* 339, 225-229.
- Kim, S.Y., Lee, J.H., Shin, H.S., Kang, H.J., Kim, Y.S., 2002. The human elongation factor 1 alpha (EF-1 alpha) first intron highly enhances expression of foreign genes from the murine cytomegalovirus promoter. *J Biotechnol* 93, 183-187.
- Koelzer, S., Kolsch, Y., Panfilio, K.A., 2014. Visualizing late insect embryogenesis: extraembryonic and mesodermal enhancer trap expression in the beetle *Tribolium castaneum*. *PLoS One* 9, e103967.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-359.

- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-2948.
- Lehnert, B., 2015. BlandAltmanLeh: Plots (Slightly Extended) Bland-Altman Plots. R package version 0.3.1. <https://CRAN.R-project.org/package=BlandAltmanLeh>.
- Leng, N., Dawson, J.A., Thomson, J.A., Ruotti, V., Rissman, A.I., Smits, B.M., Haag, J.D., Gould, M.N., Stewart, R.M., Kendziorski, C., 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics* 29, 1035-1043.
- Lewis, E.B., 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565-570.
- Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323.
- Liao, Y., Smyth, G.K., Shi, W., 2013. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 41, e108.
- Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923-930.
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550.
- Mack, C.P., Owens, G.K., 1999. Regulation of smooth muscle alpha-actin expression in vivo is dependent on CARG elements within the 5' and first intron promoter regions. *Circ Res* 84, 852-861.
- Malicki, J., Schughart, K., McGinnis, W., 1990. Mouse *Hox-2.2* Specifies Thoracic Segmental Identity in *Drosophila* Embryos and Larvae. *Cell* 63, 961-967.
- Marie, B., Bacon, J.P., 2000. Two *engrailed*-related genes in the cockroach: cloning, phylogenetic analysis, expression and isolation of splice variants. *Development genes and evolution* 210, 436-448.
- Mayor, C., Brudno, M., Schwartz, J.R., Poliakov, A., Rubin, E.M., Frazer, K.A., Pachter, L.S., Dubchak, I., 2000. VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16, 1046-1047.
- McGinnis, N., Kuziora, M.A., McGinnis, W., 1990. Human *Hox-4.2* and *Drosophila deformed* encode similar regulatory specificities in *Drosophila* embryos and larvae. *Cell* 63, 969-976.
- McGregor, A.P., 2005. How to get ahead: the origin, evolution and function of *bicoid*. *Bioessays* 27, 904-913.
- Panfilio, K.A., 2008. Extraembryonic development in insects and the acrobatics of blastokinesis. *Developmental biology* 313, 471-491.
- Panfilio, K.A., 2009. Late extraembryonic morphogenesis and its *zen*^{RNAi}-induced failure in the milkweed bug *Oncopeltus fasciatus*. *Developmental biology* 333, 297-311.

- Panfilio, K.A., Akam, M., 2007. A comparison of *Hox3* and *Zen* protein coding sequences in taxa that span the *Hox3/zen* divergence. *Development genes and evolution* 217, 323-329.
- Panfilio, K.A., Liu, P.Z., Akam, M., Kaufman, T.C., 2006. *Oncopeltus fasciatus zen* is essential for serosal tissue function in katatrepsis. *Developmental biology* 292, 226-243.
- Panfilio, K.A., Oberhofer, G., Roth, S., 2013. High plasticity in epithelial morphogenesis during insect dorsal closure. *Biol Open* 2, 1108-1118.
- Panfilio, K.A., Roth, S., 2010. Epithelial reorganization events during late extraembryonic development in a hemimetabolous insect. *Developmental biology* 340, 100-115.
- Papillon, D., Telford, M.J., 2007. Evolution of *Hox3* and *ftz* in arthropods: insights from the crustacean *Daphnia pulex*. *Development genes and evolution* 217, 315-322.
- Passner, J.M., Ryoo, H.D., Shen, L., Mann, R.S., Aggarwal, A.K., 1999. Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. *Nature* 397, 714-719.
- Peel, A.D., Telford, M.J., Akam, M., 2006. The evolution of hexapod *engrailed*-family genes: evidence for conservation and concerted evolution. *Proceedings. Biological sciences / The Royal Society* 273, 1733-1742.
- Peterson, M.D., Popadic, A., Kaufman, T.C., 1998. The expression of two *engrailed*-related genes in an apterygote insect and a phylogenetic analysis of insect *engrailed*-related genes. *Development genes and evolution* 208, 547-557.
- Pultz, M.A., Diederich, R.J., Cribbs, D.L., Kaufman, T.C., 1988. The *proboscipedia* locus of the Antennapedia complex: a molecular and genetic analysis. *Genes & development* 2, 901-920.
- R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria URL <https://www.R-project.org/>.
- Rafiqi, A.M., Lemke, S., Ferguson, S., Stauber, M., Schmidt-Ott, U., 2008. Evolutionary origin of the amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of *zen*. *Proceedings of the National Academy of Sciences of the United States of America* 105, 234-239.
- Rafiqi, A.M., Lemke, S., Schmidt-Ott, U., 2010. Postgastrular *zen* expression is required to develop distinct amniotic and serosal epithelia in the scuttle fly *Megaselia*. *Developmental biology* 341, 282-290.
- Rezende, G.L., Martins, A.J., Gentile, C., Farnesi, L.C., Pelajo-Machado, M., Peixoto, A.A., Valle, D., 2008. Embryonic desiccation resistance in *Aedes aegypti*: presumptive role of the chitinized serosal cuticle. *BMC Dev Biol* 8, 82.
- Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Bucher, G., 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature* 452, 949-955.
- Rieden, P.M.J.I.D., Mainguy, G., Woltering, J.M., Durston, A.J., 2004. Homeodomain to hexapeptide or PBC-interaction-domain distance: size apparently matters. *Trends Genet* 20, 76-79.

- Ronshaugen, M., McGinnis, N., McGinnis, W., 2002. Hox protein mutation and macroevolution of the insect body plan. *Nature* 415, 914-917.
- Ruijter, J.M., Ramakers, C., Hoogaars, W.M., Karlen, Y., Bakker, O., van den Hoff, M.J., Moorman, A.F., 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res* 37, e45.
- Rushlow, C., Doyle, H., Hoey, T., Levine, M., 1987. Molecular characterization of the *zerknüllt* region of the Antennapedia gene complex in *Drosophila*. *Genes & development* 1, 1268-1279.
- Rushlow, C., Levine, M., 1990. Role of the *zerknüllt* gene in dorsal-ventral pattern formation in *Drosophila*. *Advances in genetics* 27, 277-307.
- Sarrazin, A.F., Peel, A.D., Averof, M., 2012. A segmentation clock with two-segment periodicity in insects. *Science* 336, 338-341.
- Schinko, J.B., Hillebrand, K., Bucher, G., 2012. Heat shock-mediated misexpression of genes in the beetle *Tribolium castaneum*. *Development genes and evolution* 222, 287-298.
- Schinko, J.B., Weber, M., Viktorinova, I., Kiupakis, A., Averof, M., Klingler, M., Wimmer, E.A., Bucher, G., 2010. Functionality of the GAL4/UAS system in *Tribolium* requires the use of endogenous core promoters. *BMC Dev Biol* 10, 53.
- Schmidt-Ott, U., 2000. The amnioserosa is an apomorphic character of cyclorrhaphan flies. *Development genes and evolution* 210, 373-376.
- Schmidt-Ott, U., Rafiqi, A.M., Lemke, S., 2010. *Hox3/zen* and the evolution of extraembryonic epithelia in insects. *Adv Exp Med Biol* 689, 133-144.
- Schmitt-Engel, C., Schultheis, D., Schwirz, J., Strohlein, N., Troelenberg, N., Majumdar, U., Dao, V.A., Grossmann, D., Richter, T., Tech, M., Donitz, J., Gerischer, L., Theis, M., Schild, I., Trauner, J., Koniszewski, N.D., Kuster, E., Kittelmann, S., Hu, Y., Lehmann, S., Siemanowski, J., Ulrich, J., Panfilio, K.A., Schroder, R., Morgenstern, B., Stanke, M., Buchholz, F., Frasch, M., Roth, S., Wimmer, E.A., Schoppmeier, M., Klingler, M., Bucher, G., 2015. The iBeetle large-scale RNAi screen reveals gene functions for insect development and physiology. *Nat Commun* 6, 7822.
- Schoppmeier, M., Fischer, S., Schmitt-Engel, C., Lohr, U., Klingler, M., 2009. An ancient anterior patterning system promotes *caudal* repression and head formation in ecdysozoa. *Current biology : CB* 19, 1811-1815.
- Scott, M.P., Tamkun, J.W., Hartzell, G.W., 3rd, 1989. The structure and function of the homeodomain. *Biochimica et biophysica acta* 989, 25-48.
- Syednasrollah, F., Laiho, A., Elo, L.L., 2015. Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief Bioinform* 16, 59-70.
- Silver, K., Jiang, H., Fu, J., Phillips, T.W., Beeman, R.W., Park, Y., 2014. The *Tribolium castaneum* cell line TcA: a new tool kit for cell biology. *Sci Rep* 4, 6840.

- Simmonds, A.J., Brook, W.J., Cohen, S.M., Bell, J.B., 1995. Distinguishable functions for *engrailed* and *invected* in anterior-posterior patterning in the *Drosophila* wing. *Nature* 376, 424-427.
- St Johnston, D., Driever, W., Berleth, T., Richstein, S., Nusslein-Volhard, C., 1989. Multiple steps in the localization of *bicoid* RNA to the anterior pole of the *Drosophila* oocyte. *Development* 107 Suppl, 13-19.
- Stappert, D., Frey, N., von Levetzow, C., Roth, S., 2016. Genome-wide identification of *Tribolium* dorsoventral patterning genes. *Development* 143, 2443-2454.
- Stauber, M., Jackle, H., Schmidt-Ott, U., 1999. The anterior determinant *bicoid* of *Drosophila* is a derived Hox class 3 gene. *Proceedings of the National Academy of Sciences of the United States of America* 96, 3786-3789.
- Stauber, M., Prell, A., Schmidt-Ott, U., 2002. A single *Hox3* gene with composite *bicoid* and *zerknüllt* expression characteristics in non-Cyclorrhaphan flies. *Proceedings of the National Academy of Sciences of the United States of America* 99, 274-279.
- Stone, J.R., Wray, G.A., 2001. Rapid evolution of *cis*-regulatory sequences via local point mutations. *Molecular biology and evolution* 18, 1764-1770.
- Strobl, F., Stelzer, E.H., 2014. Non-invasive long-term fluorescence live imaging of *Tribolium castaneum* embryos. *Development* 141, 2331-2338.
- Tautz, D., 2000. Evolution of transcriptional regulation. *Curr Opin Genet Dev* 10, 575-579.
- Telford, M.J., Thomas, R.H., 1998a. Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proceedings of the National Academy of Sciences of the United States of America* 95, 10671-10675.
- Telford, M.J., Thomas, R.H., 1998b. Of mites and *zen*: expression studies in a chelicerate arthropod confirm *zen* is a divergent Hox gene. *Development genes and evolution* 208, 591-594.
- Trauner, J., Schinko, J., Lorenzen, M.D., Shippy, T.D., Wimmer, E.A., Beeman, R.W., Klingler, M., Bucher, G., Brown, S.J., 2009. Large-scale insertional mutagenesis of a coleopteran stored grain pest, the red flour beetle *Tribolium castaneum*, identifies embryonic lethal mutations and enhancer traps. *BMC Biol* 7, 73.
- True, J.R., Carroll, S.B., 2002. Gene co-option in physiological and morphological evolution. *Annu Rev Cell Dev Biol* 18, 53-80.
- Tuomi, J.M., Voorbraak, F., Jones, D.L., Ruijter, J.M., 2010. Bias in the Cq value observed with hydrolysis probe based quantitative PCR can be corrected with the estimated PCR efficiency value. *Methods* 50, 313-322.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res* 40, e115.
- van der Zee, M., Berns, N., Roth, S., 2005. Distinct functions of the *Tribolium zerknüllt* genes in serosa specification and dorsal closure. *Current biology : CB* 15, 624-636.

- Veillard, F., Troxler, L., Reichhart, J.M., 2016. *Drosophila melanogaster* clip-domain serine proteases: Structure, function and regulation. *Biochimie* 122, 255-269.
- von Charpentier, T., 1825. *Horae entomologicae, adjectis tabulis novem coloratis*. Wratistaviae.
- Wagle, P., Nikolic, M., Frommolt, P., 2015. QuickNGS elevates Next-Generation Sequencing data analysis to a new level of automation. *Bmc Genomics* 16, 487.
- Wakimoto, B.T., Turner, F.R., Kaufman, T.C., 1984. Defects in embryogenesis in mutants associated with the *antennapedia* gene complex of *Drosophila melanogaster*. *Developmental biology* 102, 147-172.
- Wang, J., Wu, M., Wang, B., Han, Z., 2013. Comparison of the RNA interference effects triggered by dsRNA and siRNA in *Tribolium castaneum*. *Pest Manag Sci* 69, 781-786.
- Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57-63.
- Waterhouse, R.M., Tegenfeldt, F., Li, J., Zdobnov, E.M., Kriventseva, E.V., 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Res* 41, D358-365.
- Wilhelm, B.T., Landry, J.R., 2009. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* 48, 249-257.
- Wolff, C., Schroder, R., Schulz, C., Tautz, D., Klingler, M., 1998. Regulation of the *Tribolium* homologues of *caudal* and *hunchback* in *Drosophila*: evidence for maternal gradient systems in a short germ embryo. *Development* 125, 3645-3654.
- Zeh, D.W., Zeh, J.A., Smith, R.L., 1989. Ovipositors, Amnions and Eggshell Architecture in the Diversification of Terrestrial Arthropods. *The Quarterly Review of Biology* 64, 147-198.
- Zhao, J.J., Lazzarini, R.A., Pick, L., 1993. The mouse *Hox-1.3* gene is functionally equivalent to the *Drosophila Sex combs reduced* gene. *Genes & development* 7, 343-354.

8 SUPPLEMENT

Table S1. List of candidate target genes, which are shared by *Tc-zen1* and *Tc-zen2* during early embryogenesis. *Dm-Drosophila melanogaster*.

TC ID	<i>Dm</i> orthologue/closest homologue	Beetle-Base phenotype
TC003982	none existing	not screened
TC012208	none existing	not screened
TC015555	none existing	not screened
TC032847	none existing	not screened
TC033185	none existing	not screened
TC033543	none existing	not screened
TC033623	none existing	not screened
TC008005	none existing	screened, but no information
TC009377	CG34115 - no further information	not screened
TC014634	Kvk - Krotzkopf verkehrt - chitin synthase	not screened
TC003371	CG13510 - no further information	screened, but no information
TC014345	CG3777 - no further information	screened, but no information
TC015598	CG14439 - no further information	screened, but no information
TC034364	CG5928 - no further information	screened, but no information
TC003085	A10 - Antennal protein 10, insect odorant-binding protein	not screened
TC003708	CG4115 - lectin like domain predicted	not screened
TC006727	Pwn - Pawn - Ca binding domain	not screened
TC010653	Knk - knickknopf-chitin organization protein	screened, but no information
TC010675	Skeletor	screened, but no information
TC010825	Exp - expansion - contains SMAD domain, involved in trachea development	screened, but no information

TC011140	Obst-A - Obstructor-A - chitin binding domain	not screened
TC011141	Obst-A - Obstructor-A - chitin binding domain	not screened
TC013464	CG5958 - transporter activity	not screened
TC014517	CG4089 - Tubuline-tyrosine ligase	screened, but no information
TC015554	CG9360 - Short-chain dehydrogenase/reductase (SDR)	screened, but no information
TC015721	CG9503 - Glucose-methanol-choline oxidoreductase	screened, but no information
TC031718	CG14205 - Acyltransferase 3	screened, but no information
TC031823	CG7330 - no further information	not screened
TC032053	Kaz1-ORFB - Kazal domain (potential serine protease inhibitor)	not screened
TC033053	Ndae1 - Na ⁺ -driven anion exchanger 1	screened, but no information
TC033106	CG15497 - Haemolymph juvenile hormone binding	screened, but no information
TC033244	Mthl15 - Methuselah-like 15; GPCR, family 2, secretin-like	not screened
TC034444	CG9990 - ABC-2 transporter	screened, but no information
TC000520	Spz - Spaetzle	muscle pattern potentially obscured by segmentation defect
TC004438	TI - Toll	Tube-like phenotype, fate shift
TC010864	CG3246 - lipid binding domain	mature eggs not present
TC011791	AnxB11 - Annexin B11 - actin/Ca ion binding	pupal molt delayed
TC012027	CG7675 - oxidoreductase activity	muscle pattern potentially obscured by segmentation defect
TC014100	Serp - Serpentine	muscle pattern potentially obscured by segmentation defect; cuticle crumbs
TC014346	Cyp6a23 - heme/Fe ion binding; oxidoreductase activity; Cytochrome P450	severe defects during embryogenesis - multiple cuticle phenotypes including partial inside-out
TC015481	Cht7 - Chitinase 7 - chitin binding	larva dorsally bent, muscle pattern irregular

TC015712	CG9503 - Glucose-methanol-choline oxidoreductase	antennal flagellum size decreased, urogomphi orientation irregular
TC033746	CG9514 - Glucose-methanol-choline oxidoreductase	mature egg deposition blocked number, ovariole decreased
TC034834	Ade-3 - Adenosine-3 - Phosphoribosylglycinamide synthetase	eclosion not fulfilled, empty egg phenotype
TC034861	Pmp70 - Peroxisomal Membrane Protein 70 kDa - ABC transporter-like	severe defects during embryogenesis - multiple cuticle phenotypes

ZUSAMMENFASSUNG

Hox-Gene kodieren für Transkriptionsfaktoren, die für die Determination der axialen Muster der Embryonen von bilateralen Lebewesen verantwortlich sind. Die Insekten-Orthologe des *Hox3*-Gens, genannt *zerknüllt* (*zen*), haben jedoch ihre Funktion mehrmals gewandelt. Das hat zum Verlust der kanonischen *Hox*-Funktion geführt und zu einer Verschiebung ihrer funktionalen Domäne von embryonalem Gewebe zu extraembryonalem Gewebe. Alle bis heute beschriebenen *zen*-Gene haben eine Rolle in extraembryonalen Membranen (EEM). Diese Membranen schützen den Embryo vor äußeren Einflüssen und erlauben Insekten, in verschiedenen Nischen Eier abzulegen. Das hat ihnen letztendlich erlaubt auch das Land zu kolonisieren. Die Evolution von der EEMs ist mit der Evolution von *Hox3/zen* eng verknüpft. Gleichzeitig mit der Entstehung der EEMs hat sich die Rolle von *Hox3* in der Entwicklung des Embryos allmählich zu der Funktion von *zen* in extraembryonalen Membranen gewandelt. Nur in geflügelten Insekten ist diese Transition vollständig und vollständige EEMs können beobachtet werden. Neben der Verlagerung der Funktion von *zen* haben diese Gene in ihrer neuen Expressionsdomäne zwei neue Rollen übernommen: eine in der frühen Gewebespezifikation sowie eine in der späten Morphogenese. Bisher ist jedoch wenig über die Gründe für die Wandlung von *Hox3* zu *zen* bekannt, genauso wenig wie über die funktionelle Divergenz von *zen*. Um die Auslöser für die funktionelle Wandlung des *Hox3*-Gens zu untersuchen, habe ich mich bei meinen Untersuchungen auf den holometabolen Käfer *Tribolium castaneum* konzentriert, in dem zwei funktionell divergente Paraloge von *zen* beschrieben sind: eines mit einer Funktion in der frühen Embryogenese (*Tc-zen1*) und das zweite mit einer Funktion in der späten Embryogenese (*Tc-zen2*).

Um zu erforschen, wie die zwei divergenten Funktionen von *Tc-zen1* und *Tc-zen2* erworben wurden, habe ich die Expressions- und Translationsregulation durch beide Gene während der frühen und späten Embryogenese untersucht. Ich konnte zeigen, dass beide Paraloge höchste Expression in der frühen Embryogenese zeigen, obwohl bisher nur für *Tc-zen1* eine frühe Funktion beschrieben ist. Um den Grad der Divergenz von Zielgenen zwischen beiden Paralogen in der frühen Embryonalentwicklung zu zeigen, habe ich die Translation der *Tc-zen*-Gene mittels parentaler RNAi unterdrückt und RNA-Expressionsanalyse in den Nachkommen durchgeführt. Die Analyse der differentiellen Expression und nachfolgende vergleichende Analysen der identifizierten, potentiellen Zielgene von *Tc-zen1* und *Tc-zen2* deuten darauf hin, dass beide Paraloge keine wesentliche Menge von Zielgenen in der frühen Embryonalentwicklung teilen. Außerdem deckte eine Hauptkomponentenanalyse auf, dass trotz der frühen Expression beider Gene, der *Knockdown* von *Tc-zen2* wesentlich weniger Auswirkungen auf die frühe Transkriptionskontrolle hat, als der *Knockdown* von *Tc-zen1*. Dieses Ergebnis stimmt mit der beschriebenen, späten Funktion von *Tc-zen2* überein. Die Untersuchung der Expressionsniveaus beider *Tc-zen*-Gene in RNAi-Embryonen des jeweiligen Paralogs zeigte jedoch eine subtile regulatorische Funktion von *Tc-zen2*, vor allem in der Repression von *Tc-zen1*.

Weitere Analysen der Expressionsregulation von *Tc-zen2* zeigten, dass eine niedrige Expression des Transkripts bis in die späte Embryonalentwicklung bestehen bleibt, obwohl die höchste Expression von *Tc-zen2* bereits in der frühen Entwicklung vorherrscht. Passend zu

der Expression von *Tc-zen2*-mRNA ist auch das Tc-Zen2-Protein bis in die späte Embryonalentwicklung präsent, wo dessen bisher einzige Funktion beschrieben war. Um Zielgene von *Tc-zen2* zu identifizieren habe ich eine weitere RNA-Expressionsanalyse durchgeführt. Ein Ergebnis dieser Analyse war, dass *Tc-zen2* eine viel höhere Wichtigkeit bei der Transkriptionskontrolle in der späten als in der frühen Embryonalentwicklung hat. Durch gründliche Gen-Ontologie-Analysen wurde ein Funktionsprofil potentieller Zielgene von *Tc-zen2* in der späten Embryonalentwicklung angefertigt. Passend zu der bereits beschriebenen Funktion von *Tc-zen2* in der späten Embryogenese, konnten vielen der identifizierten Kandidaten Gen-Ontologien zugewiesen werden, die eine Funktion in epithelialer Morphogenese aufweisen.

Zusammenfassend für Ergebnisse dieses Projekts kann also gesagt werden, dass die unterschiedlichen Funktionen der *Tc-zen*-Paraloge durch die unterschiedlichen transkriptionalen Signaturen begründet werden können. Während die Funktion von *Tc-zen1* mit dessen höchster Expression und dessen transkriptioneller Regulation seiner Zielgene korreliert, hat *Tc-zen2* nur einen vergleichsweise geringeren Einfluss auf die transkriptionale Regulation in der frühen Embryogenese. Da außerdem beide *Tc-zen*-Paraloge wenige Zielgene teilen und die regulatorische Funktion von *Tc-zen2* in der Frühentwicklung gering ist, kann eine Unterscheidung der Rollen von *Tc-zen1* und *Tc-zen2* in die frühe und späte Embryonalentwicklung angenommen werden. Diese Annahme wird durch den Nachweis von Tc-Zen2 während der gesamten Embryonalentwicklung bis zu dessen beschriebener Funktion unterstützt. Außerdem konnte ich eine viel größere Anzahl an potentiellen Zielgenen für *Tc-zen2* während der späten Embryogenese identifizieren wovon viele eine mögliche Rolle in der Morphogenese von Epithelien haben könnten. Diese unterschiedlichen Ergebnisse deuten darauf hin, dass die divergenten Funktionen der *Tc-zen*-Paraloge zunächst durch die Regulation unterschiedlicher Zielgene entstanden sein könnten und dann dadurch in unterschiedliche Funktionen in der frühen und späten Embryonalentwicklung unterschieden wurden.

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Kristen Panfilio for giving me the opportunity to work on this challenging project. Thank you for designing this project in an exciting way, which kept me always intrigued and hungry for new discoveries, and during which I could learn incredible amount of diverse techniques. Thank you for all the scientific discussions we had over the past three and a half years, because they kept me motivated throughout the whole period of this PhD. Lastly, thank you for critically revising this thesis and ignoring my disability to use the articles.

I would also like to thank Prof. Siegfried Roth for regularly sharing his wisdom and providing scientific ideas for this project during our lab and thesis committee meetings. Also, I would like to thank Prof. Milos Tsiantis for keeping me from forgetting about the large point of view on the project and for reminding me to keep my focus on what I want to discover. I would also like to thank Prof. Thomas Wiehe, who, with rather short notice, accepted to chair my defense.

I have met many incredible people in the last three and a half years and I honestly don't know where to start. Thank you, all lovely people from Siegfried's and Kristen's lab!!! It has always felt as if we were only one group!!! Thanks for the great working environment full of laughter and for selflessly sharing all your knowledge and for all the suggestions and advices!!!

My dear Nadine, thank you for being a great colleague, for sharing all your endless ISH wisdom and knowledge, and for always putting your things aside in order to help me immediately. Thank you for all the scientific discussions and for invaluable advices you gave me. Thank you for being the best roomie on all the conferences and thank you for surviving Venice with me! But most importantly, thank you for becoming my friend. I will never forget how you first invited me to join for a beer in the park :). I admire your strength, with which you go through each day! You are such an amazing smart beautiful person and I am very lucky to have spent this crazy journey with you and I will miss you ever so much!!!

Dennisino, you joined our lab like a flash of lightning and you brought with you so much positive energy and laughter. Your positive attitude and constant smile on your face were always very cheerful and made the working atmosphere simply perfect. You, my co-founder of "Friday Beer Club" (which nobody else joined), became such a dear friend of mine. I don't know many people that would drive through the whole city during the night just to pick me up and drive me for one minute to the emergency, because I stupidly cut my finger. I have always enjoyed our conversations and discussions (both scientific and personal) very much, whether over the Friday beer, or over the "spritzig" beverage at "Rhein Terrace". I admire your enthusiasm and your drive! I am indeed very happy to have met you and will miss you super a lot!!!

My dear Salimino, I am so thankful I have met you in my life. I have learnt so much from you! You are such a sweet, kind and honest person and your chicken masala is simply amazing. I will always remember our "Abenteuer" during German classes. Also, I have tremendously enjoyed many of our extremely focused scientific discussions.

Dear Kai, how many tons of sugar have we consumed over the past years? Thank you for constant supply of sweets of various kinds and textures and for picking up a late dinner for me in a form of Steinofen baguette and mango chili spread. But also thank you for countless laughter we had during the time spent in the office and lab. Thanks for all the scientific

discussions, especially about different RNA-seq tools. I have to admit that your daily sharp challenging questions in the end broaden my knowledge more than I expected.

Thorsteino, thank you for everything!!! You are such an incredibly smart person. Thank you for countless scientific discussions, which always ended up with you giving me a suggestion, which turned out to be great and useful. Every time I exhausted all my ideas about certain obstacles, I knew, I just have to discuss the issue with Thorsteino and he will give me the answer. You were not just my postdoc colleague, you were like my co-supervisor. Also, thank you for critically revising the results chapter. It was very kind of you to make time for it. Oh, and thanks for all the corks :)!

Iris, you have made impossible possible. You made me lose my fear of the “black box” and actually made me liking it. Thank you for being so incredibly patient while spending endless amount of time with explaining to me the most basic stuff, which I couldn't understand, over and over and over again. I know it has not always been a (poolparty) with me. Thank you for spending months working on this project and developing the pipeline. The cool results obtained within this project are also your accomplishments! It would have not been possible without you!!! Also, thank you for finding the time to critically revise this thesis and for, once again, explaining to me stuff I seemed not yet to understand. Lastly, but very importantly, thank you for all the amazing cakes and breads you baked and shared :)!

Buru and Waldi, the pinky and the brain of the lab. You guys made me laugh so much. It was so much fun to spent time with you in the lab. Waldi, thanks for all the help with the westerns. I think, without you, I would be still developing the ladder. Buru, thanks for your countless cheering jokes. I will always remember that I should never put my baby in the corner :)!

Moja milovaná Mamka, ďakujem aj Tebe, za psychickú podporu, za všetky návštevy, za všetky veci, čo si sem navláčila, za všetky farebné zubné kefky, za všetky nostalgické pamlsky a za všetky krásne priania. Si mi veľkou oporou a viem, že vždy budeš! Ďakujem!!!

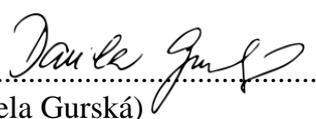
Last, but definitely not least, I would like to thank you, mio dolce Puppolino. Thank you for letting me go to do what I love without ever reproaching it. Thank you for taking countless number of flights to keep US going. Thank you for constant mental, but also scientific support. Thank you for refusing to stop helping me, although I insisted. Thank you for all the R scripts you wrote for me. Without you, until today, I would not know how much variance is explained by the first and the second PC. Thank you for being my “Tim of statisticians”. Thank you for critically revising the results chapter and sorry for not letting you revise more, but I had to mitigate the tension. Thank you for always taking care of me and thank you for supporting me, especially during the time before this PhD. Ti amo, amore mio, per sempre, tutti frutti!!! You are my favorite person and I would have not managed to start, or to finish this PhD without you!!!

ERKLÄRUNG

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Frau Dr. Kristen Panfilio betreut worden.

Köln, 18.12.2017

.....
(Ort, Datum)


.....
(Daniela Gurská)

CURRICULUM VITAE

NAME: Daniela Gurská
DATE AND CITY OF BIRTH: 08.11.1986, Košice
NATIONALITY: Slovak

EDUCATION

May 2014 – present **Ph.D. student**, Institute for Zoology, Department of Developmental Biology, **University of Cologne**, Germany
 Laboratory of Dr. Kristen A. Panfilio

Thesis project: “Changing roles of *Hox3* genes in insect evolution: characterizing the *zen* paralogues in the beetle *Tribolium castaneum*”

Oct. 2009 – June 2011 **M.Sc.** in Cell and Developmental Biology
 Faculty of Science, **Charles University in Prague**, Czech Republic, Laboratory of Dr. Zbyněk Kozmik

Thesis project: “The functional analysis of invertebrate (*Branchiostoma floridae*) promoters in heterologous systems”

Oct. 2006 – Sept. 2009 **B.Sc.** in Molecular Biology and Biochemistry of Organisms
 Faculty of Science, **Charles University in Prague**, Czech Republic, Laboratory of Dr. Zbyněk Kozmik

Thesis project: “Regulation of gene expression: *Hox Paradox*”

Sept. 2002 – June 2006 Gymnasium Šrobárova 1, Košice, Slovakia
 Field: General education with extended tuition in English