Mutual regulation underlies paralogue functional diversification

Daniela Gurska [1], Iris M. Vargas Jentzsch [1], Kristen A. Panfilio [1,2]*

[1] Institute of Zoology: Developmental Biology, University of Cologne, 50674 Cologne, Germany
[2] School of Life Sciences, University of Warwick, Coventry, CV4 7AL, United Kingdom

* Corresponding author: Kristen.Panfilio@alum.swarthmore.edu

ABSTRACT

Gene duplication followed by functional divergence eliminates potential redundancy, but to what extent does either paralogue retain the ancestral function? Insect *Hox3/zen* genes represent an evolutionary hotspot, with orthologues required either for early specification or late morphogenesis of the protective extraembryonic tissues. The *zen* paralogues of the beetle *Tribolium castaneum* present a unique opportunity to investigate both functions in a single species. We show that despite high sequence similarity the paralogues have diverged substantially in function. High-resolution analyses of expression dynamics (transcript and protein) and transcriptional targets (RNA-seq after RNAi) demonstrate that the paralogues act non-redundantly, specifically in the serosal tissue. Together, they comprise an evolutionarily novel regulatory unit, with an unexpected early role whereby Tc-Zen2 inhibits its own activator, Tc-Zen1. We further link persistent Tc-Zen2 protein with ongoing roles in the serosa that culminate in late morphogenesis. While complementary roles and mutual regulation underpin paralogue retention, this very functional divergence also resulted in both beetle paralogues now differing from single orthologues in other species.

INTRODUCTION

Over macroevolutionary time scales, changes in transcriptional regulation may result in the acquisition of novel gene functions. The *Hox3/zen* genes of insects represent a case in point. Across the bilaterian animals, Hox genes are conserved in genomic organization, expression, and function, with roles in tissue specification along the anterior-posterior body axis of the developing embryo (Krumlauf 1992). Instead, the *Hox3* genes in winged insects, known as *zen*, are prone to genomic microinversions (Negre and Ruiz 2007; McKenna, et al. 2016; Armisen, et al. 2018) and are required in the novel tissue domain of the extraembryonic membranes (EEMs), epithelial tissues that surround and protect the developing embryo (Falciani, et al. 1996; Hughes and Kaufman 2002; Horn, et al. 2015).

Typically an insect embryo is covered by two distinct membranes, the serosa and the amnion (Panfilio 2008). Precise development of these simple (monolayer) epithelia is essential for embryogenesis. In early development the EEMs surround the embryo, forming a protective barrier from the environment. In particular, the outer serosal tissue is capable of innate immune responses (Chen, et al. 2000; Jacobs, et al. 2014) and it secretes a thick chitin-based cuticle that mechanically reinforces the eggshell and provides desiccation resistance (Rezende, et al. 2008; Jacobs, et al. 2013; Panfilio, et al. 2013; Farnesi, et al. 2015). In later development, active withdrawal of the EEMs is essential for correct closure of the embryo's back, completing the outer form of the body (Panfilio, et al. 2013; Hilbrant, et al. 2016).

Extraembryonic expression and function of *zen* accompanied the evolutionary origin of the insect EEMs as protective covers. The nature of this novel extraembryonic role has been functionally investigated in several species spanning the breadth of the insects, variously identifying roles in early EEM specification or in late EEM withdrawal (reviewed in Horn, et al. 2015). Notably, although the *Hox3* locus is prone to lineage-specific duplications (*e.g.*, Ferguson, et al. 2014) only a single EEM function – specification or morphogenesis (tissue remodeling for withdrawal) – is known per species. This is even true in the derived case of the fruit fly *Drosophila melanogaster*, which has three functionally distinct paralogues: *zen* itself is involved in EEM specification, the duplicate *z2* is not required for embryogenesis, and the dipteran-specific *bicoid* has become a maternal determinant with no extraembryonic role (Pultz, et al. 1988; Rushlow and Levine 1990; Stauber, et al. 1999; McGregor 2005; Rafiqi, et al. 2008). Furthermore, secondary tissue simplification of the EEMs in *Drosophila* obviated the requirement for the late withdrawal function (Horn, et al. 2015). Thus, the ancestral role of *zen* within the extraembryonic domain has been obscured by ongoing evolutionary changes in both the EEMs and in *zen* in extant species.

There is a striking exception to the pattern of a single EEM role of *zen* per species. In the red flour beetle, *Tribolium castaneum*, *zen* has undergone a tandem duplication. *Tc-zen1* was first cloned from cDNA (Falciani, et al. 1996), while *Tc-zen2* was later identified by sequencing the Hox cluster directly (Brown, et al. 2002). The paralogues are striking for their compact, shared gene structure and for their proximity: within the 58-kb region between *Hox2/mxp* and *Hox4/Dfd,* the paralogues occupy a <3-kb interval, with only 216 bp between

the 3′ UTR of *Tc-zen1* and the initiation codon of *Tc-zen2* (Brown, et al. 2002). Nonetheless, subsequent functional diversification has equipped the paralogues with either of the two known EEM functions: early-acting *Tc-zen1* specifies the serosal tissue, while *Tc-zen2* is required for late EEM withdrawal morphogenesis (van der Zee, et al. 2005; Hilbrant, et al. 2016). We thus asked to what extent a detailed molecular characterization of the beetle paralogues could elucidate the evolutionary history of changes between the specification and morphogenesis functions of *zen* orthologues.

Here we investigate differences in the regulation of *Tc-zen1* and *Tc-zen2* as well as their own transcriptional signatures as homeodomain transcription factors. In particular, we conducted expression assays for both transcript and protein, combined with analysis of RNA-seq after RNAi data. Strikingly, the time of peak expression coincides with the time of primary function – detectable morphologically and transcriptionally – for *Tc-zen1* but not for *Tc-zen2*. The RNA-seq data also clarify the extent of regulatory overlap of the paralogues and reveal subtle aspects of temporal variability (heterochrony) after *Tc-zen2* RNAi. Our validation of specific transcriptional targets of the *zen* genes opens new avenues into serosal tissue biology and identifies a novel, paralogue-based regulatory circuit at the developmental transition from specification to maturation of the serosa.

RESULTS

Recent tandem duplication of *zen* in the *Tribolium* lineage

As multiple, evolutionarily independent instances of *Hox3/zen* gene duplication are known from various insect lineages (reviewed in Horn, et al. 2015), we first surveyed other *Tribolium* beetle genomes to assess sequence conservation at the *Hox3* locus. Using the *T. castaneum* paralogues as BLASTn queries, we were able to confirm that the tandem duplication of *zen* is conserved across three closely related congeneric species: *T. freemani*, *T. madens* and *T. confusum* (14-61 million years estimated divergence, Angelini and Jockusch 2008; see Methods). Conservation includes the exons and several discrete non-coding regions (Fig. 1A), supporting the recent nature of this duplication, while phylogenetic analysis of Zen proteins is consistent with a single duplication event at the base of the *Tribolium* lineage (Fig. 1B).

The single known functional domain of Zen proteins is the 60-amino acid homeodomain, encoded by the 180-bp homeobox (Panfilio and Akam 2007). Across *Tribolium* beetle Zen proteins, sequence alignment shows that most sites within the homeodomain are conserved, with up to three individual, non-synonymous substitutions per species across the Zen1 homeodomain. Zen2 orthologues are even more highly conserved, with a single non-synonymous substitution in *T. confusum* compared to the other three species (Figs. 1B, S1).

Next, we investigated levels of coding sequence conservation between the *T. castaneum zen* (hereafter "*Tc-zen*") paralogues. Strongest nucleotide conservation occurs

within the homeobox, where three conservation peaks correspond to the three encoded α-helices (Fig. 1C: >80% identity). In fact, within the coding sequence for the third α-helix there is a 20-bp stretch with 100% nucleotide identity (Fig. 1C), which is roughly the effective length of sequence for achieving systemic knockdown by RNA interference (RNAi; Svobodova, et al. 2016). Indeed, *Tc-zen1*-specific double-stranded RNA (dsRNA) that spans the homeobox is sufficient to effect cross-paralogue knockdown of *Tc-zen2* (Fig. 1C-D; beta regression, z=4.718, $p<0.001$), although a short fragment alone is sufficient to strongly knock down *Tc-zen1* itself (no significant change in knockdown efficiency between the long and short fragments: beta regression, z=0.558, $p=0.577$). For all subsequent paralogue-specific functional testing, we thus designed our dsRNA fragments to exclude the homeobox and thereby avoid off-target effects (Fig. 1C: *Tc-zen1* short fragment: yellow; *Tc-zen2*: green).

Distinct roles of the *Tc-zen* paralogues at different developmental stages

Embryogenesis has been well characterized in the beetle *T. castaneum* (Handel, et al. 2000; Benton, et al. 2013; Panfilio, et al. 2013; Koelzer, et al. 2014), including the *Tc-zen* paralogues' roles in EEM development. The first differentiation event distinguishes the serosa from the germ rudiment (embryo and amnion) within the cellularized blastodermal epithelium (Fig. 2A, at ~10% embryonic development). Early tissue reorganization results in the embryo proper becoming internalized relative to the amnion and serosa (EEM formation, subdivided into the "primitive pit" and "serosal window" stages). Later, this topology is reversed when the EEMs actively rupture and contract ("withdrawal"), coordinated with expansion of the embryo's flanks for dorsal closure of the body (Fig. 2C, at ~75% development). After *Tc-zen1* RNAi, presumptive serosal cells are respecified to anterior germ rudiment fates, leading to an early enlargement of the head and amnion (Fig. 2B; van der Zee, et al. 2005). *Tc-zen2* RNAi impairs or wholly blocks late EEM withdrawal (van der Zee, et al. 2005; Hilbrant, et al. 2016), confining the embryonic flanks such that the epidermis encloses the embryo's own legs instead of closing the back, leading to an everted (inside out) body configuration (Fig. 2D; Truckenbrodt 1979; Hilbrant, et al. 2016).

Here, we were able to fully reproduce the morphological phenotypes after RNAi for each *Tc-zen* paralogue (Fig. 2A′-D′). RNAi is particularly efficient for *Tc-zen1* (98.8% knockdown, Fig. 2E). Specific phenotypes after *Tc-zen2* RNAi (73.8% knockdown) include complete eversion (20.5%, Fig. 2D′) as well as milder defects in EEM withdrawal (53.3%, Figs. 2F, S2). Furthermore, we newly explored how the paralogues' functions relate to their transcript expression profiles across embryogenesis. Consistent with their functions, *Tc-zen1* has early expression while only *Tc-zen2* persists until the membrane rupture stage (Fig. 2G). Surprisingly, late-acting *Tc-zen2* also has strong expression during early development.

The *Tc-zen* paralogues exhibit subtle differences in expression during early development

To gain insight into *Tc-zen* paralogue regulation and to determine the developmental stages of primary transcription factor function for each paralogue, we undertook a fine-scale

spatiotemporal characterization of *Tc-zen1* and *Tc-zen2* expression for both transcript and protein (RT-qPCR, *in situ* hybridization, western blotting, immunohistochemistry).

As both paralogues are strongly expressed in early development (Fig. 2G), we examined these stages in detail for transcript expression. *Tc-zen1* transcript arises during blastoderm formation (4-6 hours after egg lay, hAEL), peaks at the differentiated blastoderm stage (6-10 hAEL), and retracts from the entire presumptive serosa to a narrow region at the tissue's border during EEM formation (10-14 hAEL; Fig. 3A-F). After the EEMs have fully enclosed the early embryo, *Tc-zen1* transcript is no longer detected (Figs. 2G, 3A). Peak *Tc-zen1* transcript expression is followed shortly by detectable protein for Tc-Zen1, although this, too, only persists during early development (Figs. 4A, S3A).

Tc-zen2 expression starts with a slight temporal offset compared to *Tc-zen1*, at the differentiated blastoderm stage (6-8 hAEL), with peak levels occurring during EEM formation (10-14 hAEL; Fig. 3A). We also observed several differences in the paralogues' spatial expression patterns. In line with the RT-qPCR data, we did not observe *Tc-zen2* expression before blastoderm differentiation (Fig. 3G), and its first appearance in an anterior subset of the serosa occurs at the stage when *Tc-zen1* is broadly expressed throughout the tissue (compare Fig. 3C,H). Then, *Tc-zen2* transcript expands throughout the serosa while *Tc-zen1* transcript retracts, concomitant with the expansion of the entire serosal tissue as it encloses the germ rudiment during EEM formation (compare Fig. 3D-F,I-K). Notably, the *Tc-zen* paralogues are expressed consecutively, but not concurrently, at the rim of the serosa. It is only during late EEM formation, at the serosal window stage, that we first observe *Tc-zen2* expression throughout the entire serosal tissue (Fig. 3K). By this time, Tc-Zen2 protein is also strongly expressed and persists (Figs. 4A, S3B, and see below), while *Tc-zen2* transcript wanes gradually (from 14 hAEL; Figs. 2G, 3A).

Transcriptional impact of *Tc-zen1* and *Tc-zen2* during early embryogenesis

Since protein expression follows shortly after peak transcript expression for both paralogues (Figs. 3A, 4A), we used the high sensitivity of our RT-qPCR survey (Fig. 3A) to inform our staging for functional testing by RNAi. To identify transcriptional targets for each *zen* gene, our RNA-seq after RNAi approach assessed differential expression (DE) between wild type and knockdown samples. We focused specifically on the time windows of peak gene expression: 6-10 hAEL for *Tc-zen1* and 10-14 hAEL for *Tc-zen2* (curly brackets in Fig. 3A). These four-hour windows were chosen to maximize the number of identified target genes while restricting detection to prioritize direct targets of Zen transcription factor binding. Note that, throughout, our reporting of "DE genes" refers to analyses across all isoforms (18,536 isoform models) in the *T. castaneum* official gene set OGS3 (see Methods).

The RNA-seq data are consistent with *a priori* expectations based on the stage affected by RNAi for each *zen* gene (Fig. 2A-D). That is, *Tc-zen1* has a clear early role in tissue specification, and its knockdown at these stages has a strong transcriptional impact, wherein principal component analysis (PCA) clearly distinguishes experimental treatments

(Fig. 5A: blue and yellow samples). In contrast, *Tc-zen2* has an early expression peak but its manifest role in late EEM withdrawal occurs nearly two days later (56% development later). Not surprisingly, we therefore find a negligible effect on the early egg's total transcriptome after *Tc-zen2* RNAi (Fig. 5A: grey and green samples), despite verification of efficient knockdown (Fig. 2F). Overall, we obtained 338 DE genes after *Tc-zen1* RNAi compared to only 26 DE genes after *Tc-zen2* RNAi, while global transcriptional changes during early embryogenesis affect nearly 12% of the OGS (2221 DE genes: Tables 1A,C,D, S1A-C).

Given the recent nature of the *Tribolium zen* gene duplication and the similarity of the *Tc-zen* paralogues' early expression profiles, we asked whether there is a weak legacy of shared function. If this is the case, *Tc-zen2*, which does not have a strong early role, might exhibit a regulatory profile similar to *Tc-zen1* under relaxed thresholds for differential expression. With rather low stringency cut-off criteria ($P_{adj} \leq 0.05$, $|FC|>1$), we identified 1339 and 477 DE genes after *Tc-zen1* RNAi and *Tc-zen2* RNAi, respectively, of which 120 are shared by both paralogues (Fig. 5B, Table S2A-B). However, only 42 shared transcriptional targets are regulated in the same way by both *Tc-zen* paralogues (Fig. 5C, Table S2A), comprising a mere 3.1% of *Tc-zen1*'s low-stringency targets. For comparison, nearly twice as many target genes were strongly downregulated after knockdown of *Tc-zen1* and showed minor upregulation after *Tc-zen2* RNAi (Fig. 5C: third panel, Table S2B). Thus, we conclude that *Tc-zen2* has a minimal effect on early development, and that this does not constitute a transcriptional "echo" of co-regulation with *Tc-zen1* due to common ancestry. Why, then, is *Tc-zen2* strongly expressed during early development?

The *Tc-zen* paralogues are mutual regulatory targets in the serosa

We next considered the *Tc-zen* paralogues as factors necessary for defining the serosal tissue, and sought to elucidate how this is represented by some of their specific transcriptional targets. *Tc-zen1* is strictly required to confer serosal tissue identity (van der Zee, et al. 2005). Differentiation of the serosa from the germ rudiment involves an early switch from mitosis to the endocycle (Handel, et al. 2000; Benton, et al. 2013), resulting in characteristic polyploidy of the serosa (Panfilio, et al. 2013). Consistent with this, we identified a homologue of the essential endocycle factor *fizzy-related* among DE genes upregulated by Tc-Zen1 (Table S1A; Schaeffer, et al. 2004; Cohen, et al. 2018). We also hypothesized that the slight offset whereby *Tc-zen1* expression precedes *Tc-zen2* is consistent with *Tc-zen1* activating *Tc-zen2*. We confirmed this regulatory interaction both by RNA-seq and RT-qPCR after *Tc-zen1* RNAi (Fig. 6A-B). Thus, Tc-Zen1 as a serosal specifier targets factors for definitive tissue differentiation, including *Tc-zen2* as a candidate.

Are there Tc-Zen2 transcriptional targets that could support an early role in the serosa? Among the handful of genes with strong differential expression (Table 1D, $P_{adj} \leq 0.01$, $|FC| \geq 2$), we could validate several as likely targets for transcriptional activation. Specifically, these candidate genes have expression in the early serosa and/or their transcript levels are first strongly upregulated within the time window of peak *Tc-zen2* expression (12-14 hAEL; *e.g.*, Fig. S4). Their putative functions as structural components of chitin-based cuticle or as

signaling molecules are consistent with a role for *Tc-zen2* in the physiological maturation of the serosa: one of its first tasks upon enclosing the embryo is to secrete the protective cuticle (Jacobs, et al. 2013; Martins Vargas, et al. 2014).

In performing reciprocal validation assays, we then uncovered an unexpected early function of Tc-Zen2 in the repression of its own paralogous activator. After *Tc-zen2* RNAi, we detect an upregulation of *Tc-zen1* that was only weakly suggested by the RNA-seq data but then strongly supported in subsequent RT-qPCR assays (Fig. 6A-B). We confirmed this observation by *in situ* hybridization. After *Tc-zen2* RNAi, *Tc-zen1* transcript is expressed at higher levels than in wild type (compare Fig. 6C-D,F-G). *Tc-zen1* also remains strongly expressed throughout the serosa at stages when wild type expression is restricted to low levels at the tissue rim (compare Fig. 6E,H). In fact, the abrupt reduction in *Tc-zen1* transcript levels in wild type correlates with increasing *Tc-zen2* expression (Fig. 3A: 10-16 hAEL). This is also reflected at the protein level, as Tc-Zen1 is only detected for the short time before Tc-Zen2 appears (Fig. 4A). Together, these results suggest that early *Tc-zen2* expression may be required for repression of *Tc-zen1* in the maturing serosa (see Discussion). Thus, the *Tribolium* paralogues likely function as mutual regulatory targets (Fig. 6I).

Tc-Zen2 is exclusively serosal, with persistent nuclear localization

To complete our analysis of *Tc-zen2*, we continued to examine its expression and function at later stages. After the early peak in transcript levels, we could detect both transcript (weakly, Figs. 2G, 3A) and protein (particularly strongly in mid-embryogenesis, Fig. 4A) continuously until the stage of EEM withdrawal, spanning 14-75% of development (10-54 hAEL, assayed in two-hour intervals; see also Fig. S3B). Moreover, we find that Tc-Zen2 is persistently localized to the nucleus, demonstrated by fluorescent immunohistochemistry on cryosectioned material of selected stages (Fig. 4B-E,G,H). This suggests that Tc-Zen2 may be active throughout much of embryogenesis, whereas some species' orthologues are regulated by exclusion from the nucleus at certain stages (Dearden, et al. 2000). We could also refine the spatial scope of *Tc-zen2* activity: in contrast to earlier reports (van der Zee, et al. 2005), we found no evidence for *Tc-zen2* transcript or protein in the amniotic tissue (Fig. 4D-H: particularly note arrows in 4F,H′,H′′), indicating that this factor is strictly serosal.

Late transcriptional dynamics are largely serosa-specific and *Tc-zen2*-dependent

Complementing the early RNA-seq after RNAi experiment at the time of peak *Tc-zen2* expression (Figs. 5,6), we used the same approach to examine the stage of known *Tc-zen2* function in late EEM withdrawal. Withdrawal begins with rupture of the EEMs, at 52.1 ± 2.3 hAEL as determined by live imaging (Koelzer, et al. 2014). Here, we assayed the four-hour intervals just before (48-52 hAEL) and after (52-56 hAEL) rupture. These consecutive developmental stages allow us to assess *Tc-zen2* transcriptional regulation that precedes and then accompanies EEM withdrawal. Consistent with *Tc-zen2*'s known late role, we detect >16× more DE genes after *Tc-zen2* RNAi in late development (>430 DE genes, compare

Table 1E-F with 1D, see also Table S3A-B). PCA also clearly separates knockdown and wild type samples at each stage (Fig. 7A).

Our staging helps to contextualize *Tc-zen2* and EEM-specific processes relative to concurrent embryonic development. We thus evaluated differential expression in pairwise comparisons not only between wild type and RNAi samples, but also over time in both backgrounds (Fig. 7B, Tables 1B,E-G, S3C-D). Comparisons across consecutive developmental stages (early: 6-10 hAEL *vs*. 10-14 hAEL; late: 48-52 hAEL *vs.* 52-56 hAEL) reveal two general changes in the wild type transcriptional landscape. There is far less dynamic change in gene expression in late development (5.8× fewer DE genes), consistent with steady state and ongoing processes in later embryogenesis compared to the rapid changes of early development. Also, whereas early development shows a fairly even balance between activation (48%) and repression (52%), late development is predominantly characterized by activation, with increasing expression levels over time (79%, Table 1A,B).

Against this backdrop, the transcriptional impact of *Tc-zen2* is even more pronounced. In the wild type background, most genes with changing expression over time are also affected by *Tc-zen2* RNAi (Fig. 7B: 77%, 293/383 DE genes from green Venn diagram set). We detect this strong effect even though *Tc-zen2* is restricted to the serosa (Fig. 4), a tissue that ceased mitosis in early development (see above) and therefore represents only a small cell population within our whole-egg samples. This suggests that most dynamic transcription at these stages pertains to EEM morphogenesis, with the global transcriptional impact of *Tc-zen2* at these stages even greater than for *Tc-zen1* in early development (Table 1E-F, *cf.* 1C). Many late transcriptional targets of *Tc-zen2* exhibit consistent, ongoing regulation (*i.e.*, either activated or repressed at both stages; 26%, Table 1E-F, Fig. 7C). A handful of genes shows stage-specific regulation, with activation before and repression during withdrawal (Fig. 7C). Meanwhile, most candidate *Tc-zen2* targets are only differentially expressed at a single stage (72%, Fig. 7C). Altogether, these patterns imply that persistent nuclear localization of Tc-Zen2 (Fig. 4) reflects active and dynamic transcriptional control, not merely localization to the nucleus or DNA binding in a paused, non-functional state (Banks, et al. 2016).

Evidence of variable developmental delay after *Tc-zen2* RNAi

Global assessment of the *Tc-zen2*^{RNAi} molecular phenotype also provides new insight into the physical phenotype of defective EEM withdrawal, suggesting that a variable developmental delay in the serosa is the underlying cause. Support for this view derives from both our PCA and pairwise DE assessments, which involve qualitatively different analyses of the RNA-seq data and thus represent congruent rather than redundant approaches (see Methods).

Several observations are consistent with a delay. As noted above, all late RNA-seq biological replicates cluster by treatment by PCA. Interestingly, the older *Tc-zen2*^{RNAi} samples (52-56 hAEL) have intermediate component scores compared to the clusters for the younger *Tc-zen2*^{RNAi} and younger wild type samples (48-52 hAEL, Fig. 7A). Similarly, DE comparisons identify noticeably fewer DE genes between the older *Tc-zen2*^{RNAi} sample and

either of the younger samples (Tables 1G-H, S3D-E). In fact, there is virtually no difference in the transcriptional profile of the older *Tc-zen2*^{RNAi} sample compared to the younger wild type sample (Table 1H: "# DE genes" column). At the same time, nearly all genes that change in expression over time in the *Tc-zen2*^{RNAi} background (Fig. 7B: yellow Venn diagram set) are also candidate targets of *Tc-zen2* at the pre-rupture stage (95%, 165 of 174 DE genes, Fig. 7B: intersection of blue and yellow sets). In other words, *Tc-zen2*^{RNAi} eggs generally require an additional four hours (5.6% development) to attain a transcriptional profile comparable to the wild type pre-rupture stage, and this is achieved by belated activation of Tc-Zen2 target genes. However, only a subset of the candidate target genes exhibit a delayed recovery after *Tc-zen2* RNAi; the majority do not (66%, Fig. 7B: DE genes of the blue set that are not in the yellow set). Thus, the target genes that exhibit transcriptional recovery may be independently activated by other factors, in addition to activation by Tc-Zen2.

Our RNA-seq data also indicate increased variability after *Tc-zen2* RNAi. The pre-rupture *Tc-zen2*^{RNAi} biological replicates show comparably tight clustering to their age-matched wild type counterparts (Fig. 7A). In contrast, the older *Tc-zen2*^{RNAi} samples have a noticeably greater spread along the vectors of the first two principal components (Fig. 7A). This variability may in itself provide explanatory power for the spectrum of end-stage *Tc-zen2*^{RNAi} phenotypes (Fig. 2F, discussed below). In sum, our RNA-seq profiling after *Tc-zen2* RNAi suggests that a partial developmental delay – rather than outright absence – of preparatory transcriptional changes in the serosa impairs or wholly prevents EEM withdrawal.

Functional profiling of candidate late developmental targets of Tc-Zen2

To complete our molecular analyses of how *Tc-zen2* regulates EEM withdrawal morphogenesis, we used gene ontology (GO) functional annotations to characterize candidate transcriptional targets. Initial enrichment tests confirmed that ongoing regulation of serosal cuticle structure is a primary role of *Tc-zen2* (GO terms pertaining to cuticle and extracellular components, Fig. S5, Table S4A-B). Complementing an early role in serosal cuticle synthesis (discussed above, Table 1D), an essential prerequisite for tissue reorganization during EEM withdrawal is that the serosal epithelium detaches from its cuticle (a process known as apolysis; Lamer and Dorn 2001; Panfilio 2009). Consistent with this, enriched GO terms encompass not only structural constituents of chitin-based cuticle but also enzymes involved in cuticle remodeling, degradation, and catabolism (Table S4A-B).

We next devised GO categories of interest that reflect other known biological processes in the late serosa (Fig. 8A, Table S5). In addition to requirements for cytoskeletal, epithelial, and extracellular remodeling for EEM withdrawal, we considered GO terms for *Drosophila* imaginal disc morphogenesis, which has noted similarities to EEM withdrawal (Hilbrant, et al. 2016). We included transcriptional regulation, insofar as *Tc-zen2* may act upstream of several gene regulatory networks for specific biological processes. Lastly, we looked for potential stress response genes, in light of the unnatural constraint of the embryo in the absence of EEM withdrawal (Panfilio 2009). We acknowledge the limitations of using GO annotations to infer function for homologous genes in novel tissue contexts. (*Drosophila*

does not have discrete serosal and amniotic tissues or a direct equivalent of EEM withdrawal morphogenesis.) Nonetheless, our *a priori* categories account for over half of the biological process GO terms for late DE genes (Fig. 8A: left chart). An additional prominent category that we had not deliberately selected was for transmembrane transport proteins (Fig. 8A: left and middle charts). Transporters may serve physiological roles of the serosa as the outermost tissue layer, a barrier epithelium with the potential to mediate exchange between the egg and the outer environment and for yolk catabolism (Dorn 1976; Lamer and Dorn 2001; Panfilio 2008). Nonetheless, there still remain numerous potential *Tc-zen2* targets with other GO designations or that did not receive functional annotations in these analyses.

We took these diverse gene categories into account in selecting a dozen candidate targets for validation. RNA-seq DE predictions for activating or repressive regulation by *Tc-zen2* were confirmed by RT-qPCR for all tested candidates (Fig. 8B). Of particular interest, this included two genes with changing directions of *Tc-zen2* regulation over time (activation followed by repression, Fig. 7C), where both genes encode proteins that have conserved yet uncharacterized domains of unknown function (Fig. 8B: red text; Table S6). Altogether, late DE genes represent a large and unbiased sample of candidate effector genes for EEM withdrawal and lay the foundation for future investigation of the wider roles that *Tc-zen2* plays in extraembryonic tissue biology.

DISCUSSION

The changes that led to a role for *Hox3* in the evolutionarily novel domain of the insect extraembryonic membranes involved multiple events (Horn, et al. 2015). Here, we focused our investigations on the holometabolous beetle *T. castaneum*. Tandem duplication and subsequent functional divergence partitioned the two known roles of insect *zen* genes between the paralogues: *Tc-zen1* with the early serosal specification function and *Tc-zen2* with the late withdrawal morphogenesis function. Our detailed characterization of gene regulation both upstream and downstream of the *Tc-zen* genes reveals several unexpected features surrounding the selective pressures and biological value of these unusual paralogues.

Coding and non-coding sequence conservation belie the extent of *zen* paralogue functional divergence

The high sequence similarity between Tc-Zen1 and Tc-Zen2 was previously known from comparisons with other insect Zen proteins (Panfilio and Akam 2007). Our genomic analyses of additional *Tribolium* species confirm conservation within this genus of the entire *Hox3* locus spanning both paralogues, consistent with a recent, lineage-specific origin of this duplication (Fig. 1A-B).

The *T. castaneum* paralogues have scarcely diverged, with the homeobox presenting a target for cross-paralogue knockdown, particularly within the region encoding the third α-helix (Fig. 1C-D). This α-helix largely confers DNA-binding specificity (Passner, et al. 1999), demonstrated by amino acid substitutions within the functionally derived class 3 Hox

gene *bicoid* in the Diptera (McGregor 2005). The similarity of the *Tc-zen* paralogues led us to speculate that they may retain a degree of overlap in their transcriptional targets. Instead, our RNA-seq analyses demonstrate that even with relaxed statistical thresholds there is little evidence of their shared ancestry or redundant activity (Fig. 5). In light of these findings, high nucleotide conservation, particularly of the *zen2* homeobox (Fig. S1), may reflect both limited divergence and positive, stabilizing selection. Where, then, does the specificity of the *Tc-zen* paralogues lie? In canonical Hox3 proteins, DNA-binding specificity can be enhanced by the common Hox cofactor Extradenticle (Passner, et al. 1999). In contrast, insect *zen* genes have lost the hexapeptide motif required for this interaction, and no other co-factor binding motifs are known (Panfilio and Akam 2007), deepening the long recognized "Hox specificity paradox" (Crocker, et al. 2015) in the case of the beetle *zen* paralogues.

The question of specificity also arises for upstream regulation of the *T. castaneum* paralogues. Given the extreme proximity of the paralogues' tandem gene loci, fine-tuned transcriptional regulation, or a restriction of regulatory crosstalk, should be required. Conserved non-coding regions may thus contribute to the nuanced transcriptional regulation of the *Tc-zen* paralogues. The >1-kb region upstream of *zen1* has particularly high sequence similarity across species, with >100-bp stretches of 100% nucleotide identity. This region, including the 5′ UTR of *Tc-zen1*, was recently cloned into a reporter vector and tested *in vivo* (Fig. 1A: dashed line; Strobl, et al. 2018). This construct could recapitulate late expression around the rim of the closing serosal window, a feature common to both *Tc-zen* paralogues (as in Figs. 3F,K, 6E). However, early blastoderm *Tc-zen1* expression is absent (*cf*., Fig. 3B-C), while increasing embryonic/amniotic expression in this reporter represents a wholly ectopic domain. Thus, early *Tc-zen1* expression requires additional regulatory inputs.

Tight mutual regulation suggests that the *Tc-zen* paralogues represent a novel genetic kernel for early serosal development

Although the *Tc-zen* paralogues exhibit similar serosal expression (Falciani, et al. 1996; van der Zee, et al. 2005), our high-resolution spatiotemporal analyses reveal distinct profiles (Figs. 3-4) that can be explained by their mutual regulatory interaction (Fig. 6). The paralogues' dynamic expression across the serosa is largely complementary, if not outright mutually exclusive, with the posteriorward expansion of *Tc-zen2* coinciding with retraction of *Tc-zen1* to the tissue rim. That is, Tc-Zen1 upregulates *Tc-zen2* in its wake, and Tc-Zen2 in turn represses *Tc-zen1*. This newly discovered negative feedback loop constitutes a tight linkage between the paralogues, with the implication that there is a strong developmental requirement to repress *Tc-zen1* even before the serosa has fully enclosed the embryo.

As the expression of both *Tc-zen* genes is restricted to the serosa, it is unclear why *Tc-zen1* as the essential specifier of this domain is repressed while *Tc-zen2* persists. Studies in the fly species *Drosophila melanogaster* and *Megaselia abdita* have identified several subtle features of the temporal control of *zen* orthologues (Rafiqi, et al. 2010; Schmidt-Ott, et al. 2010; Gavin-Smyth, et al. 2013). *Dm-zen* is short-lived, downregulating in the amnioserosa during gastrulation, while *Ma-zen* expression persists in the serosa (Schmidt-Ott, et al. 2010).

Ubiquitous, ectopic overexpression of *zen* impairs amnion specification, germband retraction, and/or dorsal closure in these species (Rafiqi, et al. 2010). Yet the relevance of these findings to the beetle is unclear. Amnioserosal-dependent germband retraction is a *Drosophila*-specific morphogenetic feature (Panfilio 2008), and induced ubiquitous expression contrasts with the endogenous situation of serosa-specific expression of the *Tc-zen* genes (Figs. 3-4). Nonetheless, one intriguing observation from the *Drosophila* overexpression studies was a noticeable increase in amnioserosal nuclear and cell size (Rafiqi, et al. 2010). The insect extraembryonic epithelia are known to be polyploid to characteristic levels (Reim, et al. 2003; Panfilio and Roth 2010; Panfilio, et al. 2013), and excessive ploidy could conceivably interfere with the tissues' final structure and function as barrier epithelia with precise morphogenetic requirements (Orr-Weaver 2015), including in *Tribolium*.

It would thus be intriguing to investigate overexpression of *Tc-zen1* in a tissue-specific manner. Herein lies a genetic challenge due to the mutual regulation of the paralogues. To what extent could overexpression of *Tc-zen1* overcome strong upregulation of *Tc-zen2* as its target, resulting in turn in strong repression of *Tc-zen1* and thus cancelling out the manipulation? In fact *Tc-zen2* RNAi does confer serosa-specific overexpression of *Tc-zen1* and reduction in *Tc-zen2* levels (Fig. 6, Table S1B). Phenotypically, there are no ostensible consequences of this manipulation until EEM withdrawal (Fig. 6H; Hilbrant, et al. 2016), but it does ultimately impair embryogenesis (Fig. 2F). The fact that *Tc-zen2* RNAi knockdown efficiency is consistently lower than for *Tc-zen1* RNAi (Fig. 2; van der Zee, et al. 2005, and see Methods) could in fact reflect a dose-limiting lack of regulatory disentanglement. Furthermore, it was previously shown that *Tc-zen2* has an unusual early role in translational repression of the posterior embryonic factor *caudal* (Schoppmeier, et al. 2009). Conceivably, Tc-Zen2 repression of *Tc-zen1* could act in a composite fashion at both the transcriptional and translational levels (Alon 2007). While embryonic injection of Tc-Zen1 protein could bypass this repression, spatial restriction within the blastoderm, efficient protein uptake after cellularization, and protein perdurance would be technical challenges.

The *Tc-zen* feedback loop represents a regulatory unit that is difficult to disentangle and that appears to be essential for serosal development. Together *Tc-zen1* and *Tc-zen2* satisfy the criteria to be viewed as a minimal gene regulatory network (GRN) kernel (Davidson and Erwin 2006), including "recursive wiring" and the experimental challenges this entails. Alternatively, the *Tc-zen* paralogues could be viewed as a single unit in a serosal GRN and thus qualify as a "paradoxical component" (Hart and Alon 2013). Under this conceptual framework, *Tc-zen1+2*, collectively, is paradoxical in that it both activates and then inhibits (Fig. 6I). Consistent with theoretical expectations, delayed inhibition produces a discrete pulse of *Tc-zen1* (Figs. 3A, 4A). Since the pulse occurs once and is not oscillatory, this could further imply that *Tc-zen2* is a positive autoregulator (Hart and Alon 2013). Thus, the beetle *zen* paralogues have functionally diverged such that both are essential, with mutual regulation resulting in a subtle genetic separation of serosal specification and maturation functions.

Tc-Zen2 has multiple roles throughout embryogenesis

We detect *Tc-zen2* throughout most of the lifetime of the EEMs, spanning ~60% of embryogenesis (Figs. 2G,3,4), and we have now uncovered manifold roles at different stages.

Although few in number (Fig. 5A, Table 1D), we confirmed several Tc-Zen2 targets in early development (Fig. S4). Repression of *Tc-zen1* (Fig. 6) and *Tc-caudal* (Schoppmeier, et al. 2009) also highlights two unusual features of Tc-Zen2 function. First, a predominantly repressive role contrasts with Hox genes typically serving as transcriptional activators, as do both *Tc-zen* paralogues at the stages of their primary transcriptional impact (Table 1C,E). Secondly, the number of targets and precise mechanism of Tc-Zen2 translational repression remain open questions. Translational repression is a hallmark of *Drosophila* Bicoid (Stauber, et al. 1999; McGregor 2005). Further work on this aspect will clarify the extent to which such a function is ancestral or arose independently in these *Hox3/zen* derivatives.

Persistent nuclear localization of Tc-Zen2 (Fig. 4) during the lengthy interval between EEM formation and withdrawal suggests an active role, such as in serosa-specific maturation. For example, we had previously speculated that ongoing physiological processes in the serosa are reflected in the KT650 enhancer trap line, which exhibits a gradual increase in serosal EGFP (Koelzer, et al. 2014). In fact, the two genes that flank the KT650 insertion show Tc-Zen2-dependent upregulation in late development (Table S3A), although their molecular functions and homologues remain unknown. These genes, along with our validated early Tc-Zen2 targets (Fig. S4), await functional characterization, focusing on mid-embryogenesis.

Complementing this unbiased, expression-based approach, we also evaluated Tc-Zen2 regulation of serosal immune genes (Jacobs, et al. 2014), as constant Tc-Zen2 expression might reflect basal immune competence. Our pre-rupture RNA-seq data represent the oldest stage with a fully closed and protective serosa. Although our samples were not pathogen challenged, we could detect expression of 83% of serosal immune genes (n= 107 genes), with 20% showing differential expression after *Tc-zen2* RNAi (Table S3A), including upregulation of all lectins in this immune gene set. Thus, while Tc-Zen2 is not a global effector, it may regulate subsets of immune genes. Notably, most of the expressed immune genes also maintain transcript expression even after the serosa opens during withdrawal (87 of 89 genes), and with continued Tc-Zen2-dependent DE for 8% of all immune genes (Table S3B), supporting these expression features as inherent characteristics of the serosa.

Finally, our RNA-seq after RNAi analyses identified late candidate *Tc-zen2* transcriptional targets (Fig. 7, Table 1). We find that *Tc-zen2*-dependent EEM withdrawal is the major transcriptionally regulated event at these stages (Fig. 7B). Moreover, temporal and molecular variability after *Tc-zen2* RNAi may underpin observed phenotypic variability in terms of how severely EEM tissue structure, integrity, and morphogenetic competence are impaired. This ranges from mild defects in dorsal closure after transient EEM obstruction to persistently closed EEMs that cause complete eversion of the embryo (Figs. 2, S2). While this phenotypic spectrum is broad in end-stage manifestation, the unifying feature is a

heterochronic shift of extraembryonic compared to embryonic developmental processes (delayed EEM withdrawal compared to epidermal outgrowth for dorsal closure).

For the specific genes involved in withdrawal, as expected a key molecular category pertains to remodeling of the serosal cuticle to enable tissue sliding when the serosa contracts. Although our chosen GO category of stress was not strongly represented (Fig. 8), the stages we analyzed may be too young to capture the full extent of a stress response after the time window for normal withdrawal has passed (Panfilio 2009). Meanwhile, our rich set of DE genes, which includes diverse GO functional annotations (Fig. 8: "other") and novel genes without GO terms, will help reveal the full picture of EEM withdrawal. The sole *zen* orthologue in the milkweed bug *Oncopeltus fasciatus* has a similarly persistent expression profile and specific role in withdrawal morphogenesis, termed "katatrepsis" in this and other hemimetabolous insects (Panfilio, et al. 2006). We previously observed a number of *Of-zen*-dependent, long-term morphological changes prior to the rupture stage (Panfilio 2009). In contrast, the tight PCA clustering of the pre-rupture *Tc-zen2*^{RNAi} samples suggests that this is the stage of primary Tc-Zen2 function, without cumulative transcriptional variability. Taking the work forward, it will be interesting to compare *Tc-zen2* and *Of-zen* transcriptional targets as a way to determine conserved regulatory features of EEM withdrawal across the breadth of the insects and the dynamics of insect Zen function throughout development.

Concluding remarks

This study elucidates the precise nature of diversification in both expression and function after a tandem duplication event gave rise to two copies of *zen* in the *Tribolium* beetle lineage. Despite high sequence conservation and the proximity of the gene loci, spatiotemporal differences in the paralogues' expression dynamics reflect their mutual regulation in a negative feedback loop. Developmentally, it will be intriguing to determine what governs the manner in which gene expression waves pass through the serosal tissue, and how this relates to final tissue specification. From a molecular evolutionary perspective, the genetic precision for paralogue-specific regulation awaits still further functional and taxonomic investigation. Lastly, our analyses of *Tc-zen2* at later developmental stages have substantially expanded our understanding of the physiological underpinnings of the serosa as a novel tissue.

METHODS

Tribolium castaneum stock husbandry

All experiments were conducted with the San Bernardino wild type strain, maintained under standard culturing conditions at 30 °C and 40-60% relative humidity (Brown, et al. 2009).

In silico analyses

Draft genome assemblies for *T. freemani*, *T. madens*, and *T. confusum* were obtained as assembled scaffolds in FASTA-format (version 26 March 2013 for each species), accessed from the BeetleBase.org FTP site at Kansas State University (ftp://ftp.bioinformatics.ksu.edu/pub/BeetleBase/). Transcripts for *Tc-zen1* (TC000921-RA) and *Tc-zen2* (TC000922-RA) were obtained from the *T. castaneum* official gene set 3 (OGS3, http://bioinf.uni-greifswald.de/tcas/genes/tcas5_annotation/). These sequences were used as queries for BLASTn searches in the other species' genomes (BLAST+ 2.2.30, (Altschul, et al. 1997; Camacho, et al. 2009)). Sequences were extracted to comprise the *Hox3/zen* genomic loci, spanning the interval from 5 kb upstream of the BLASTn hit for the 5´ UTR of *Tc-zen1* to 5 kb downstream of the BLASTn hit for the 3´ UTR of *Tc-zen2*. These genomic loci were then aligned with the mVista tool (Mayor, et al. 2000; Frazer, et al. 2004) using default parameters. Nucleotide identities were calculated for a sliding window of 100 bp.

The maximum likelihood phylogenetic tree (Fig. 1B) was constructed based on an alignment of full-length Zen proteins, with gaps permitted, using the Phylogeny.fr default pipeline settings (Dereeper, et al. 2008).

Coding sequence for the *Tc-zen* paralogues was aligned with ClustalW (Larkin, et al. 2007), with manually curation to ensure a gap-free alignment of the homeobox. Nucleotide identities were calculated for a sliding window of 20 bp, using Simple Plot (Stothard 2000).

RT-qPCR

RNA was extracted using TRIzol Reagent (Ambion) according to the manufacturer's protocol. RNA quality was assessed by spectrophotometry (NanoDrop 2000, Thermo Fisher Scientific). cDNA was synthesized using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). RT-qPCR was performed as described (Horn and Panfilio 2016), using SYBR Green Master Mix (Life Technologies) and GoTaq qPCR Master Mix (Promega), with *Tc-RpS3* as the reference gene. Note that for *Tc-zen2* more consistent results were obtained using SYBR Green Master Mix. Samples were measured for the *Tc-zen* paralogues' wild type expression profiles (four biological replicates: Figs. 2G,3A) and evaluation of knockdown strength (three biological replicates: Figs. 1D,6B,8B). Intron-spanning primers were used for each *Tc-zen* paralogue and the selected candidate target genes (Table S7).

Parental RNAi and knockdown assessments

Parental RNAi was performed as described (van der Zee, et al. 2005), with dsRNA synthesized with specific primers (Table S7) and resuspended in double-distilled water (ddH_2O). Generally, 0.3-0.4 µg of dsRNA was used to inject one pupa.

Analysis of knockdown efficiency with different *Tc-zen1* dsRNA fragments involved statistical tests on RT-qPCR data. The strength of the *Tc-zen* paralogues' knockdown using

short and long *Tc-zen1* dsRNA fragments (Fig. 1C-D) was tested with a beta regression analysis in R v3.3.2 (R Core Team 2016) using the package betareg v3.1-0 (Cribari-Neto and Zeileis 2010). Relative expression of the *Tc-zen* paralogues in knockdown samples relative to wild type was used as the response variable and dsRNA fragment length as the explanatory variable.

For *Tc-zen1*^{RNAi} phenotypic scoring (Fig. 2E), serosal cuticle presence/absence was determined by piercing the fixed, dechorionated egg with a disposable needle (Braun Sterican 23G, 0.60 x 25 mm): mechanically resistant eggs were scored for presence of the serosal cuticle while soft eggs that collapsed lacked serosal cuticle.

For *Tc-zen2*^{RNAi} phenotypic scoring, larval cuticle preparations (Figs. 2C′,D′,F, S2) were produced as previously described (van der Zee, et al. 2005).

Histology: *in situ* hybridization, cryosectioning, immunohistochemistry

Whole mount *in situ* hybridization was performed as described (Koelzer, et al. 2014), with probes synthesized from gene specific primers (Table S7) and colorimetric detection with NBT/BCIP. Specimens were imaged in Vectashield mountant with DAPI (Vector Laboratories) for nuclear counterstaining. Images were acquired on an Axio Plan 2 microscope (Zeiss). Image projections were generated with AxioVision (Zeiss) and HeliconFocus 6.7.1 (Helicon Soft).

For cryosectioning, embryos were embedded in liquid sucrose-agarose embedding medium (15% sucrose, 2% agarose, [my-Budget Universal Agarose, Bio-Budget], PBS). Solid blocks of embedding medium containing embryos were stored overnight in 30% sucrose solution in PBS at 4 °C. The blocks were then embedded in Tissue Freezing Medium (Leica Biosystems) and flash-frozen in ice-cold isopentene (2-methylbutane). Samples were serially sectioned (20 µm, longitudinal; 30 µm, transverse) with a CM1850 cryostat (Leica Biosystems).

Protein was detected for both Tc-Zen1 and Tc-Zen2 with specific peptide antibodies (gift from the laboratory of Michael Schoppmeier, (Mackrodt 2016)). Immunohistochemistry on whole mounts and on sectioned material was performed by washing the samples six times for 10 min. in blocking solution (2% BSA, 1% NGS, 0.1% Tween-20, PBS) followed by overnight incubation with the first antibody (rabbit anti-Tc-Zen1 and anti-Tc-Zen2, 1:1,000) at 4 °C. Next, the samples were washed six times for 10 min. in the blocking solution, followed by incubation with the secondary antibody (anti-rabbit Alexa Fluor 488 conjugate, 1:400, Invitrogen) for 3 h at room temperature (RT). Last, the samples were washed six times for 10 min. in the blocking solution. Samples were then mounted in Vectashield mountant with DAPI. Low magnification images were acquired with an Axio Imager 2 equipped with an ApoTome 2 (Zeiss) structured illumination module, and maximum intensity projections were generated with ZEN blue software (Zeiss). High magnification images were acquired with an LSM 700 confocal microscope (Zeiss) and the projections were generated with ZEN 2 black software (Zeiss).

Western blots

For each two-hour developmental interval, 50 µg of protein extract was separated by SDS-PAGE. Separated proteins were transferred onto nitrocellulose membrane (Thermo Fisher

Scientific), which was blocked for 1 h in the blocking solution (100 mM Tris, 150 mM NaCl, pH 7.5, 0.1% Tween-20, 3% milk powder [Bebivita, Anfangsmilch]). Next, the membrane was incubated overnight at 4 °C with the first antibody (rabbit anti-Tc-Zen1 and anti-Tc-Zen2, 1:1,000; mouse anti-Tubulin [Sigma-Aldrich #T7451: Monoclonal anti-acetylated tubulin], 1:10,000). Afterwards, the membrane was washed three times for 10 min. with the blocking solution at RT. The membrane was then incubated with the secondary antibodies (anti-rabbit and anti-mouse, HRP, 1:10,000, Novex) for 1 h at RT. Last, after the membrane was washed three times for 10 min. with the blocking solution at RT, the membrane was incubated with ECL substrate according to the manufacturer´s protocol (WesternSure ECL Substrate, LI-COR) and digital detection was performed on a western blot developing machine (C-DIGIT, LI-COR) with the high sensitivity settings.

RNA-sequencing after RNAi

For transcriptomic profiling, a total of six *Tc-zen1*^{RNAi} experiments were conducted: three performed with the short and three with the long dsRNA fragment (Fig. 1D). A total of seven *Tc-zen2*^{RNAi} experiments were conducted: one for each biological replicate at each developmental stage. Samples chosen for sequencing were assessed by RT-qPCR for level of knockdown in RNAi samples, with *Tc-zen1* reduced to ~10% of wild type levels and *Tc-zen2* to ~24% across biological replicates. For early development (6-14 hAEL), three biological replicates were sequenced for each experimental treatment, with 100-bp paired end reads on an Illumina HiSeq2000 machine. For late development (48-56 hAEL), four biological replicates were sequenced with 75-bp paired end reads on a HiSeq4000 machine. All sequencing was performed at the Cologne Center for Genomics (CCG), with six (HiSeq2000) or eight (HiSeq4000) multiplexed samples per lane yielding ≥6.6 Gbp per sample.

The quality of raw Illumina reads was examined with FastQC (Andrews 2010). The adaptor sequences and low quality bases were removed with Trimmomatic v0.36 (Bolger, et al. 2014). Trimmomatic was also used to shorten 100-bp reads from the 3´ end to 75-bp reads to increase mapping efficiency (Table S8, (Li, et al. 2010)). The overrepresented sequences of mitochondrial and ribosomal RNA were filtered out by mapping to a database of 1266 *T. castaneum* mitochondrial and ribosomal sequences extracted from the NCBI nucleotide database (accessed 21 October 2016, search query "'tribolium [organism] AND (ribosomal OR mitochondrial OR mitochondrion) NOT (whole genome shotgun) NOT (Karroochloa purpurea)") with Bowtie2 v2.2.9 (Langmead and Salzberg 2012). Trimmed and filtered reads were mapped to the *T. castaneum* OGS3 (see above, file name: Tcas5.2_GenBank.corrected_v5.renamed.mrna.fa) with RSEM (Li and Dewey 2011). The raw read count output from RSEM was compiled into count tables.

Both principal component and differential expression analyses were performed in R using the package DESeq2 v1.14.1 (Love, et al. 2014) with default parameters. For PCA, raw (unfiltered) read counts were used. For DE analyses, to eliminate noise all genes with very low read counts were filtered out by sorting in Microsoft Excel (following recommendations in (Busby, et al. 2013)): specifically, genes were excluded from DE analysis if read counts ≤10 in ≥1 biological replicates for both the knockdown and wild type samples.

Gene ontology (GO) analyses

GO enrichment analysis was performed by Blast2GO (Conesa, et al. 2005) using two-tailed Fisher´s exact test with a threshold false discovery rate (FDR) of 0.05.

GO term analysis was performed by Blast2GO against the *Drosophila* database (accessed 9 June 2017). Only GO terms from the level 5 were considered. Next, GO terms were grouped into categories of interest based on similarity in function (Table S5). Afterwards a unique count of *T. castaneum* gene sequences was calculated for each category of interest and the percentage was compared to the rest of the GO terms in the level 5 for each GO domain.

ACKNOWLEDGMENTS

We thank Denise Mackrodt and Michael Schoppmeier for the kind gift of the Tc-Zen1 and Tc-Zen2 peptide antibodies, Viera Kovacova for bioinformatic program recommendations, Luigi Pontieri for assistance with statistical analyses, and Thorsten Horn for sharing unpublished data on cuticle gene expression. We also thank Miltos Tsiantis and Siegfried Roth for helpful discussions and recommendations throughout the course of this research project. Siegfried Roth, Peter Heger, and Matthias Pechmann provided helpful feedback on the manuscript.

FUNDING

This work was supported by funding from the German Research Foundation (Deutsche Forschungsgemeinschaft) through SFB 680 project A12 and Emmy Noether Program grant PA 2044/1-1 to KAP.

AUTHOR CONTRIBUTIONS

DG designed experiments, collected and analyzed data, established the bioinformatic pipeline for the RNA-seq data, wrote the paper.

IMVJ analyzed data, established the bioinformatic pipeline for the RNA-seq data, edited the manuscript.

KAP conceived the project, designed experiments, analyzed data, established the bioinformatic pipeline for the RNA-seq data, wrote the paper.

Figure 1. High conservation of *Tribolium zen* orthologues and paralogues.
(A) Sequence conservation at the *Hox3/zen* locus of four *Tribolium* congenerics (represented by their four-letter species abbreviations), using the *T. castaneum* locus (10 kb) as the reference sequence, assayed with a 100-bp sliding window. High nucleotide identity occurred in the homeobox (black rectangles), while conserved non-coding regions (pink) could comprise important regulatory regions. For example, the dashed line region was tested in a recent attempt to produce a *Tc-zen1* reporter (see Discussion, (Strobl, et al. 2018)).
(B) (Left) Maximum likelihood phylogenetic analysis supports the lineage-specific duplication of the *Tribolium zen* paralogues. Single-copy orthologue outgroups represent either the specification or morphogenesis function (*Megaselia abdita* and *Oncopeltus fasciatus*, respectively, (Panfilio, et al. 2006; Rafiqi, et al. 2008)). Node support values (%) are indicated; branch length unit is substitutions per site. (Right) Although the *Tribolium zen* paralogues have comparable levels of nucleotide conservation within the homeobox (Nt.), Zen1 has more amino acid substitutions than Zen2 (AA; see also Fig. S1).
(C) *T. castaneum zen* paralogue coding sequence comparison. Nucleotide identity is shown based on a 20-bp sliding window analysis. The final alignment of 887 positions includes 399 identities and 154 unaligned positions (gaps). Colored bars represent the homeobox and dsRNA fragments used for RNAi (see legend).
(D) *Tc-zen1* RNAi with the long dsRNA fragment causes cross-paralogue knockdown, with significantly stronger reduction in *Tc-zen2* levels (beta regression tests: see main text). Mean expression levels (RT-qPCR) are shown from three biological replicates after RNAi with the indicated dsRNA fragments (mapped in panel C, same color code); error bars represent the upper and lower quantiles (lower=0.025, upper=0.975).

Figure 2. *Tc-zen* paralogue roles in early specification (*Tc-zen1*) or late EEM morphogenesis (*Tc-zen2*).
(A-D′) Phenotypic consequences of RNAi for the *Tc-zen* paralogues are depicted in comparison with their wild type counterparts: schematically at the time of primary defect (A-D, upper row) and with micrographs for the resulting phenotypes (A′-D′, lower row). Briefly, knockdown results in an early fate map shift, resulting in an enlarged head region (*Tc-zen1* RNAi), or in impairment of withdrawal morphogenesis, leading to everted embryos (*Tc-zen2* RNAi; see main text and Fig. S2). The dashed lines in the schematics in A,B denote the anterior embryonic border. Anatomical abbreviations: a, antenna; h, head; l, leg; t, telson; t3, third thoracic segment). Scale bars are 100 µm.
(E-F) Phenotypic penetrance after *Tc-zen1*^{RNAi} and *Tc-zen2*^{RNAi}. Sample sizes are given in parentheses (# embryos).
(G) Expression profiles of *Tc-zen1* and *Tc-zen2* during early and late development (RT-qPCR). Mean expression levels are shown from four biological replicates; error bars represent one standard deviation. "Relative abundance" was calculated for each sample as the ratio relative to a pooled template control with cDNA from all depicted samples (method as in (Horn and Panfilio 2016)). Staging abbreviations: BF, blastoderm formation; DB, differentiated blastoderm; PP, primitive pit; SW, serosal window; pre-R, pre-rupture; MR, extraembryonic membrane rupture. Time is given in hours after egg lay (hAEL).

Figure 3. Transcript expression dynamics of the *Tc-zen* paralogues during early embryogenesis.
(A) Fine-scale quantification of transcript levels in two-hour intervals (RT-qPCR, as in Fig. 2G, additional staging abbreviation: GBE, germband extension).
(B-K´) Whole mount *in situ hybridization* for *Tc-zen1* (B-F) and *Tc-zen2* (G-K), with nuclear counterstains for morphological staging (B´-K´, arrows label the expanding serosal border). Initial *Tc-zen1* expression forms a gradient in the anterior of the blastoderm (B) before expanding to ubiquitous expression throughout the serosa (C). During the primitive pit stage (D), *Tc-zen1* expression becomes patchy and starts to retract to the border of the serosa, the final region with detectable transcript (E,F). After blastoderm formation (G), *Tc-zen2* expression starts with evenly distributed expression in the anterior serosa (H; dashed line: serosal tissue border). This expression expands posteriorly (I,J) and first encompasses the serosal border and the entire tissue domain as the serosal window closes (K). All micrographs are oriented with anterior left and shown in lateral aspect with dorsal up (except in B and G, which depict stages before this can be determined). Scale bars in B and G are 100 µm and apply to B-F´ and G-K´, respectively.

Figure 4. Paralogue-specific Tc-Zen protein expression time courses and spatial restriction of Tc-Zen2.
(A) Tc-Zen1 and Tc-Zen2 expression profiles, spanning 4-54 hAEL (western blots, with anti-Tubulin as an internal loading control at 59.0 kDa). Each blot shows a chronological assay in 2-hour intervals (labels indicate the minimum age from each interval). Tc-Zen1 protein (upper blots, arrow at 28.4 kDa) is detectable for a short time window, from blastoderm differentiation until early serosal window closure. Tc-Zen2 (lower blots, arrows at 33.7 kDa) arises at the primitive pit stage and persists until membrane rupture (see also Fig. S3 for additional blots). Staging abbreviations as in Figs. 2-3, and: GBR, germband retraction.
(B-H) Staining in whole mount (B,C,F) and cryosectioned (D,E,G,H) preparations shows that *Tc-zen2* transcript (F) and protein (green: B-E,G,H) are specific to the serosal tissue and localized to the nucleus throughout development (note staging labels in lower left corner of micrographs). We found no evidence for *Tc-zen2* expression in the second extraembryonic tissue, the amnion (note Tc-Zen2-negative tissue and nuclei of the amnion labeled with arrows in F,F´,H´,H´´ and the dashed outline in G´). Furthermore, the specimen shown in panel F derives from a staining experiment spanning all early developmental stages, and where the serosa was specifically stained, as seen in Fig. 3H-K. Images labeled with the same letter are of a single embryo. Nuclear counterstains are shown in magenta. Scale bars are as indicated: 100 µm (whole mounts and overviews of longitudinal sections), 50 µm (transverse sections and G´), 10 µm (insets and H´,H´´ focusing on nuclear localization).

Figure 5. Early transcriptional control: relative impact and assessment of shared targets between the *Tc-zen* paralogues.
(A) PCA score plot of RNA-seq after RNAi replicates for *Tc-zen1* and *Tc-zen2* and their respective age-matched wild type samples at the stage of peak expression (see Fig. 3A). The percentage of variance explained by each principal component is indicated parenthetically. These components clearly distinguish between *Tc-zen1*^{RNAi} (blue) and corresponding wild type samples (yellow), while *Tc-zen2*^{RNAi} (gray) and its wild type control (green) remain clustered.
(B) Venn diagram of candidate *Tc-zen* paralogues´ targets with relaxed DE criteria ($P_{adj} \leq 0.05$, $|FC|>1$) identified 120 possible regulatory targets shared by both *Tc-zen* paralogues.
(C) For these 120 candidates, box plots show that most are indeed more strongly regulated by *Tc-zen1* than by *Tc-zen2* (greater absolute fold change in expression after RNAi). Furthermore, most of these (78 of 120) are regulated in opposite directions by the paralogues (see also Table S2A-B).

Figure 6. Activation of *Tc-zen2* by Tc-zen1 leads to *Tc-zen1's* own repression.
(A-B) *Tc-zen1* and *Tc-zen2* as candidate transcriptional targets of their respective paralogues, as determined in comparisons of wild type and RNAi samples at the stages of peak paralogue expression (see Fig. 3A). Regulatory changes are corroborated by three biological replicates each in assays from RNA-seq (A: mean fold change, P_{adj}) and RT-qPCR (B: mean relative expression, error bars represent one standard deviation).
(C-H´) Whole mount *in situ* hybridization for *Tc-zen1* in wild type (C-D) and after *Tc-zen2* RNAi (F-H), with nuclear counterstains for morphological staging (C´-H´). *Tc-zen1* expression is upregulated in the *Tc-zen2*^{RNAi} serosa. Wild type and *Tc-zen2*^{RNAi} eggs were stained for the same duration. All micrographs are oriented with anterior left and shown in lateral aspect with dorsal up, unless stated otherwise. Scale bar in C is 100 µm and applies to all micrographs.
(I) The *Tc-zen* paralogues comprise a potential genetic kernel for early serosal development: *Tc-zen1* activates *Tc-zen2*, which in return represses *Tc-zen1* (see Discussion).

Figure 7. Tc-Zen2 has a strong, stage-specific transcriptional effect during late development.
(A) PCA score plot of *Tc-zen2*^{RNAi} and wild type RNA-seq samples at the stages before and during EEM withdrawal (four biological replicates). The percentage of variance is indicated parenthetically for each principal component.
(B) Venn diagram of numbers of DE genes across four pairwise comparisons (see legend, with total DE gene counts per comparison listed parenthetically; see also Table 1B,E,F,G).
(C) Detailed comparisons of DE genes between the pre- (48-52 hAEL) and post-rupture stages (52-56 hAEL), categorized both by the stage at which a given target gene is regulated by *Tc-zen2* (either stage alone or both) and by the direction of regulation (activation or inhibition; see legend). Values are the number of DE genes per category.

Figure 8. Functional annotation and validation of candidate late developmental targets of Tc-Zen2.
(A) Gene ontology profiles of the late Tc-Zen2 candidate targets were retrieved for approximately half of all DE genes and grouped into chosen categories indicated in the legend (see main text, Methods, Table S5). These categories account for ~60%, ~35%, and ~50% of all sequences for the respective GO domains biological process, molecular function, and cellular component. Number of GO domains is given in parentheses per pairwise DE comparison (same four comparisons as in Fig. 7B). Note that a given gene may be assigned to multiple GO categories, depending on available information in public databases.
(B) Evaluation of selected Tc-Zen2 late candidate target genes. Candidates were chosen from the GO categories of interest (A, note color-coding, where white indicates an unknown/ novel gene, see also Table S6) and based on the predicted nature of Tc-Zen2 regulation. Seven candidates were chosen from each stage (48-52 hAEL and 52-56 hAEL), where two of these show changing direction of regulation across stages (gene names in red; Fig. 7C). The RNA-seq DE predictions (fold change values are given above the graphs for each gene) were validated for all genes by RT-qPCR (mean relative expression from three biological replicates, error bars represent one standard deviation).

Table 1. Pairwise differential expression (DE) comparisons of RNA-seq datasets.
Summary metrics indicate the number of DE genes per comparison, and the proportion of DE genes that are activated or inhibited by a given *Tc-zen* paralogue or that increase or decrease in expression over time in a given background (WT or RNAi). DE cut-off values are $P_{adj} \leq 0.01$, $|FC| \geq 2$. For complete gene lists and statistics, see Tables S1A-C and S3A-E.

ID	Comparison	# DE genes	% of OGS3 (all isoforms)	Activation or increasing expression over time	Repression or decreasing expression over time	% Act./ Incr.	% Repr./ Decr.
A	WT, early (6-10 vs. 10-14 hAEL)	2221	11.98%	1068	1153	48.1%	51.9%
B	WT, late (48-52 vs. 52-56 hAEL)	383	2.07%	304	79	79.4%	20.6%
C	*zen1* vs. WT, early (6-10 hAEL)	338	1.82%	331	7	97.9%	2.1%
D	*zen2* vs. WT, early (10-14 hAEL)	26	0.14%	8	18	30.8%	69.2%
E	*zen2* vs. WT, late 1 (48-52 hAEL)	481	2.59%	400	81	83.2%	16.8%
F	*zen2* vs. WT, late 2 (52-56 hAEL)	431	2.33%	306	125	71.0%	29.0%
G	*zen2*, late (48-52 vs. 52-56 hAEL)	174	0.94%	170	4	97.7%	2.3%
H	*zen2* late 2 (52-56 hAEL) vs. WT late 1 (48-52 hAEL)	32	0.17%	11	21	34.4%	65.6%

SUPPLEMENTARY FILES

PDF:

Figure S1. Alignments of zen1 and zen2 homeoboxes/domains.
Figure S2. Phenotypic consequences of Tc-zen2RNAi - milder defects.
Figure S3. Tc-Zen1 and Tc-Zen2 expression detected by western blotting.
Figure S4. Evaluation of Tc-Zen2 candidate targets.
Figure S5. Gene ontology (GO) enrichment analysis after Tc-zen2RNAi.

Table S5. List of gene ontology (GO) terms grouped to each category of interest in each GO domain.
Table S6. Tc-Zen2 late candidate targets evaluated by RT-qPCR.
Table S7. *Tribolium castaneum* (*TC*) primer sequences for in situ hybridization, RNAi, and RT-qPCR.
Table S8. Comparison of alignment statistics by read length.

Excel workbook with large datasets and gene lists:

Table			Corresponding main text element
S1A	List of differentially expressed genes during early development after *Tc-zen1* RNAi (P_{adj}≤0.01)		Table 1C
S1B	List of differentially expressed genes during early development after *Tc-zen2* RNAi (P_{adj}≤0.01)		Table 1D
S1C	List of differentially expressed genes during early development (WT *Tc-zen1* peak, 6-10 hAEL vs. WT *Tc-zen2* peak, 10-14 hAEL) (P_{adj}≤0.01)		Table 1A
S2A	Regulatory targets shared by both *Tc-zen* paralogues with relaxed thresholds (P_{adj}≤0.05): same direction of regulation, N=42		Figure 5C
S2B	Regulatory targets shared by both *Tc-zen* paralogues with relaxed thresholds (P_{adj}≤0.05): opposite direction of regulation, N=78		Figure 5C
S3A	List of differentially expressed genes during late development (pre-rupture, 48-52 hAEL) after *Tc-zen2* RNAi (P_{adj}≤0.01)		Table 1E
S3B	List of differentially expressed genes during late development (rupture/post-rupture, 52-56 hAEL) after *Tc-zen2* RNAi (P_{adj}≤0.01)		Table 1F
S3C	List of differentially expressed genes during late development (WT pre-rupture, 48-52 hAEL vs. WT rupture/post-rupture, 52-56 hAEL) (P_{adj}≤0.01)		Table 1B
S3D	List of differentially expressed genes during late development (*Tc-zen2* RNAi pre-rupture, 48-52 hAEL vs. *Tc-zen2* RNAi rupture/post-rupture, 52-56 hAEL) (P_{adj}≤0.01)		Table 1G
S3E	List of differentially expressed genes during late development (WT pre-rupture, 48-52 hAEL vs. *Tc-zen2* RNAi rupture/post-rupture, 52-56 hAEL) (P_{adj}≤0.01)		Table 1H
S4A	Enriched GO terms in the dataset "Differentially expressed genes during pre-rupture (48-52 hAEL stage) after *Tc-zen2* RNAi"		Figure 8A
S4B	Enriched GO terms in the dataset "Differentially expressed genes during rupture/post-rupture (52-56 hAEL stage) after *Tc-zen2* RNAi"		Figure 8A

REFERENCES

Alon U. 2007. Network motifs: theory and experimental approaches. Nat Rev Genet 8:450-461.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nuc. Acids Res. 25:3389-3402.

Andrews S. 2010. FASTQC. A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ version 0.11.5, released 8 March 2016.

Angelini DR, Jockusch EL. 2008. Relationships among pest flour beetles of the genus *Tribolium* (Tenebrionidae) inferred from multiple molecular markers. Mol. Phylogenet. Evol. 46:127-141.

Armisen D, Rajakumar R, Friedrich M, Benoit JB, Robertson HM, Panfilio KA, Ahn S-J, Poelchau MF, Chao H, Dinh H, et al. 2018. The genome of the water strider *Gerris buenoi* reveals expansions of gene repertoires associated with adaptations to life on the water. bioRxiv 242230:doi: 10.1101/242230.

Banks CJ, Joshi A, Michoel T. 2016. Functional transcription factor target discovery via compendia of binding and expression profiles. Sci Rep 6:20649.

Benton MA, Akam M, Pavlopoulos A. 2013. Cell and tissue dynamics during *Tribolium castaneum* embryogenesis revealed by versatile fluorescence labeling approaches. Development 140:3210-3220.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

Brown SJ, Fellers JP, Shippy TD, Richardson EA, Maxwell M, Stuart JJ, Denell RE. 2002. Sequence of the *Tribolium castaneum* homeotic complex: The region corresponding to the *Drosophila melanogaster* Antennapedia Complex. Genetics 160:1067-1074.

Brown SJ, Shippy TD, Miller S, Bolognesi R, Beeman RW, Lorenzen MD, Bucher G, Wimmer EA, 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.

Busby MA, Stewart C, Miller CA, Grzeda KR, Marth GT. 2013. Scotty: a web tool for designing RNA-Seq experiments to measure differential gene expression. Bioinformatics 29:656-657.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

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.

Cohen E, Allen SR, Sawyer JK, Fox DT. 2018. Fizzy-related dictates a cell cycle switch during organ repair and tissue growth responses in the Drosophila hindgut. eLife 7.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.

Cribari-Neto F, Zeileis A. 2010. Beta Regression in R. Journal of Statistical Software 34:1-24.

Crocker J, Abe N, Rinaldi L, McGregor AP, Frankel N, Wang S, Alsawadi A, Valenti P, Plaza S, Payre F, et al. 2015. Low affinity binding site clusters confer Hox specificity and regulatory robustness. Cell 160:191-203.

Davidson EH, Erwin DH. 2006. Gene regulatory networks and the evolution of animal body plans. Science 311:796-800.

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*. Evol. Dev. 2:261-270.

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist Nucleic Acids Res. 36(Web Server issue):W465-W469.

Dorn A. 1976. Ultrastructure of embryonic envelopes and integument of *Oncopeltus fasciatus* Dallas (Insecta, Heteroptera) I. Chorion, amnion, serosa, integument. Zoomorphologie 85:111-131.

Falciani F, Hausdorf B, Schröder R, Akam M, Tautz D, Denell R, Brown S. 1996. Class 3 Hox genes in insects and the origin of *zen*. Proc. Natl. Acad. Sci. USA 93:8479-8484.

Farnesi LC, Menna-Barreto RFS, Martins AJ, Valle D, Rezende GL. 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, Marlétaz F, Carter J-M, Taylor WR, Gibbs M, Breuker CJ, Holland PWH. 2014. Ancient expansion of the Hox cluster in Lepidoptera generated four homeobox genes implicated in extra-embryonic tissue formation. PLoS Genet. 10:e1004698.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res 32:W273-279.

Gavin-Smyth J, Wang YC, Butler I, Ferguson EL. 2013. A genetic network conferring canalization to a bistable patterning system in Drosophila. Curr Biol 23:2296-2302.

Handel K, Grünfelder CG, Roth S, Sander K. 2000. *Tribolium* embryogenesis: a SEM study of cell shapes and movements from blastoderm to serosal closure. Dev. Genes Evol. 210:167-179.

Hart Y, Alon U. 2013. The utility of paradoxical components in biological circuits. Mol Cell 49:213-221.

Hilbrant M, Horn T, Koelzer S, Panfilio KA. 2016. The beetle amnion and serosa functionally interact as apposed epithelia. eLife 5:e13834.

Horn T, Hilbrant M, Panfilio KA. 2015. Evolution of epithelial morphogenesis: phenotypic integration across multiple levels of biological organization. Front. Genet. 6:303.

Horn T, Panfilio KA. 2016. Novel functions for *Dorsocross* in epithelial morphogenesis in the beetle *Tribolium castaneum*. Development 143:3002-3011.

Hughes CL, Kaufman TC. 2002. Hox genes and the evolution of the arthropod body plan. Evol. Dev. 4:459-499.

Jacobs CGC, Rezende GL, Lamers GEM, van der Zee M. 2013. The extraembryonic serosa protects the insect egg against desiccation. Proc. R. Soc. B 280:20131082.

Jacobs CGC, Spaink HP, 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:e04111.

Koelzer S, Kölsch Y, Panfilio KA. 2014. Visualizing late insect embryogenesis: Extraembryonic and mesodermal enhancer trap expression in the beetle *Tribolium castaneum*. PLoS One 9:e103967.

Krumlauf R. 1992. Evolution of the vertebrate Hox homeobox genes. Bioessays 14:245-252.

Lamer A, Dorn A. 2001. The serosa of *Manduca sexta* (Insecta, Lepidoptera): ontogeny, secretory activity, structural changes, and functional considerations. Tissue Cell 33:580-595.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. 2010. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 26:493-500.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

Mackrodt D. 2016. Etablierung und Funktion maternaler Proteingradienten im *Tribolium* Blastoderm. [Ph.D. (Dr. rer. nat.)]. [Erlangen]: Friedrich-Alexander-Universität Erlangen-Nürnberg.

Martins Vargas HC, Farnesi LC, Martins AJ, Valle D, Rezende GL. 2014. Serosal cuticle formation and distinct degrees of desiccation resistance in embryos of the mosquito vectors *Aedes aegypti, Anopheles aquasalis* and *Culex quinquefasciatus*. J. Insect Physiol. 62:54-60.

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I. 2000. VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16:1046-1047.

McGregor AP. 2005. How to get ahead: the origin, evolution and function of bicoid. Bioessays 27:904-913.

McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, Mitchell RF, Waterhouse RM, Ahn S-J, Arsala D, et al. 2016. Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle–plant interface. Genome Biol. 17:227.

Negre B, Ruiz A. 2007. HOM-C evolution in *Drosophila*: is there a need for *Hox* gene clustering. Trends Genet. 23:55-59.

Orr-Weaver TL. 2015. When bigger is better: the role of polyploidy in organogenesis. Trends Genet. 31:307-315.

Panfilio KA. 2008. Extraembryonic development in insects and the acrobatics of blastokinesis. Dev. Biol. 313:471-491.

Panfilio KA. 2009. Late extraembryonic development and its *zen-RNAi*-induced failure in the milkweed bug *Oncopeltus fasciatus*. Dev. Biol. 333:297-311.

Panfilio KA, Akam M. 2007. A comparison of Hox3 and Zen protein coding sequences in taxa that span the *Hox3/zen* divergence. Dev. Genes Evol. 217:323-329.

Panfilio KA, Liu PZ, Akam M, Kaufman TC. 2006. *Oncopeltus fasciatus zen* is essential for serosal tissue function in katatrepsis. Dev. Biol. 292:226-243.

Panfilio KA, Oberhofer G, Roth S. 2013. High plasticity in epithelial morphogenesis during insect dorsal closure. Biol. Open 2:1108-1118.

Panfilio KA, Roth S. 2010. Epithelial reorganization events during late extraembryonic development in a hemimetabolous insect. Dev. Biol. 340:100-115.

Passner JM, Ryoo HD, Shen L, Mann RS, Aggarwal AK. 1999. Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex. Nature 397:714-719.

Pultz MA, Diederich RJ, Cribbs DL, Kaufman TC. 1988. The proboscipedia locus of the Antennapedia complex: a molecular genetic analysis. Genes Dev. 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 AM, 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*. Proc. Natl Acad. Sci. USA 105:234-239.

Rafiqi AM, Lemke S, Schmidt-Ott U. 2010. Postgastrular *zen* expression is required to develop distinct amniotic and serosal epithelia in the scuttle fly *Megaselia*. Dev. Biol. 341:282-290.

Reim I, Lee H-H, Frasch M. 2003. The T-box encoding Dorsocross genes function in amnioserosa development and the patterning of the dorsolateral germ band downstream of Dpp. Development 130:3187-3204.

Rezende GL, Martins AJ, Gentile C, Farnesi LC, Pelajo-Machado M, Peixoto AA, Valle D. 2008. Embryonic desiccation resistance in *Aedes aegypti*: presumptive role of the chitinized serosal cuticle. BMC Dev. Biol. 8:82.

Rushlow C, Levine M. 1990. Role of the *zerknüllt* gene in dorsal-ventral pattern formation in *Drosophila*. In: Wright TRF, editor. Advances in Genetics: Genetic Regulatory Hierarchies in Development. San Diego: Academic Press.

Schaeffer V, Althauser C, Shcherbata HR, Deng WM, Ruohola-Baker H. 2004. Notch-dependent Fizzy-related/Hec1/Cdh1 expression is required for the mitotic-to-endocycle transition in Drosophila follicle cells. Curr Biol 14:630-636.

Schmidt-Ott U, Rafiqi AM, Lemke S. 2010. *Hox3/zen* and the evolution of extraembryonic epithelia in insects. In: Deutsch JS, editor. Hox Genes: Studies from the 20th to the 21st Century. Austin: Landes Bioscience. p. 133-144.

Schoppmeier M, Fischer S, Schmitt-Engel C, Löhr U, Klingler M. 2009. An ancient anterior patterning system promotes *caudal* repression and head formation in Ecdysozoa. Curr. Biol. 19:1811-1815.

Stauber M, Jäckle H, Schmidt-Ott U. 1999. The anterior determinant *bicoid* of *Drosophila* is a derived class 3 Hox gene. Proc. Natl. Acad. Sci. USA 96:3786-3789.

Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102, 1104.

Strobl F, Anderl A, Stelzer EH. 2018. A universal vector concept for a direct genotyping of transgenic organisms and a systematic creation of homozygous lines. eLife 7.

Svobodova E, Kubikova J, Svoboda P. 2016. Production of small RNAs by mammalian Dicer. Pflugers Arch 468:1089-1102.

Truckenbrodt W. 1979. The embryonic covers during blastokinesis and dorsal closure of the normal and of the actinomycin D treated egg of *Odontotermes badius* (Hav.) (Insecta, Isoptera). Zool. Jb. Anat. Ont. 101:7-18.

van der Zee M, Berns N, Roth S. 2005. Distinct functions of the *Tribolium zerknüllt* genes in serosa specification and dorsal closure. Curr. Biol. 15:624-636.

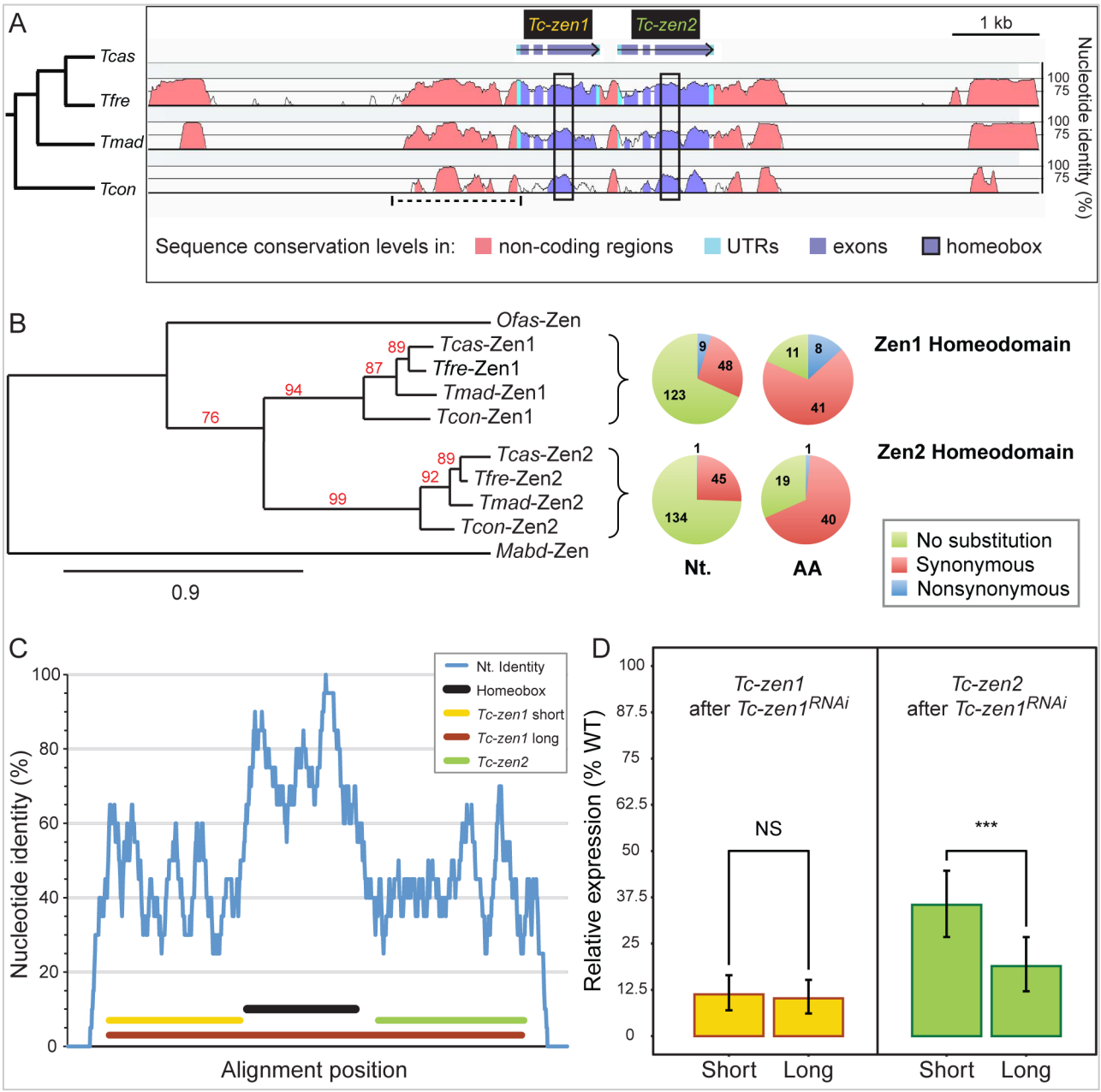

Fig 1. High conservation of *Tribolium zen* orthologues and paralogues.

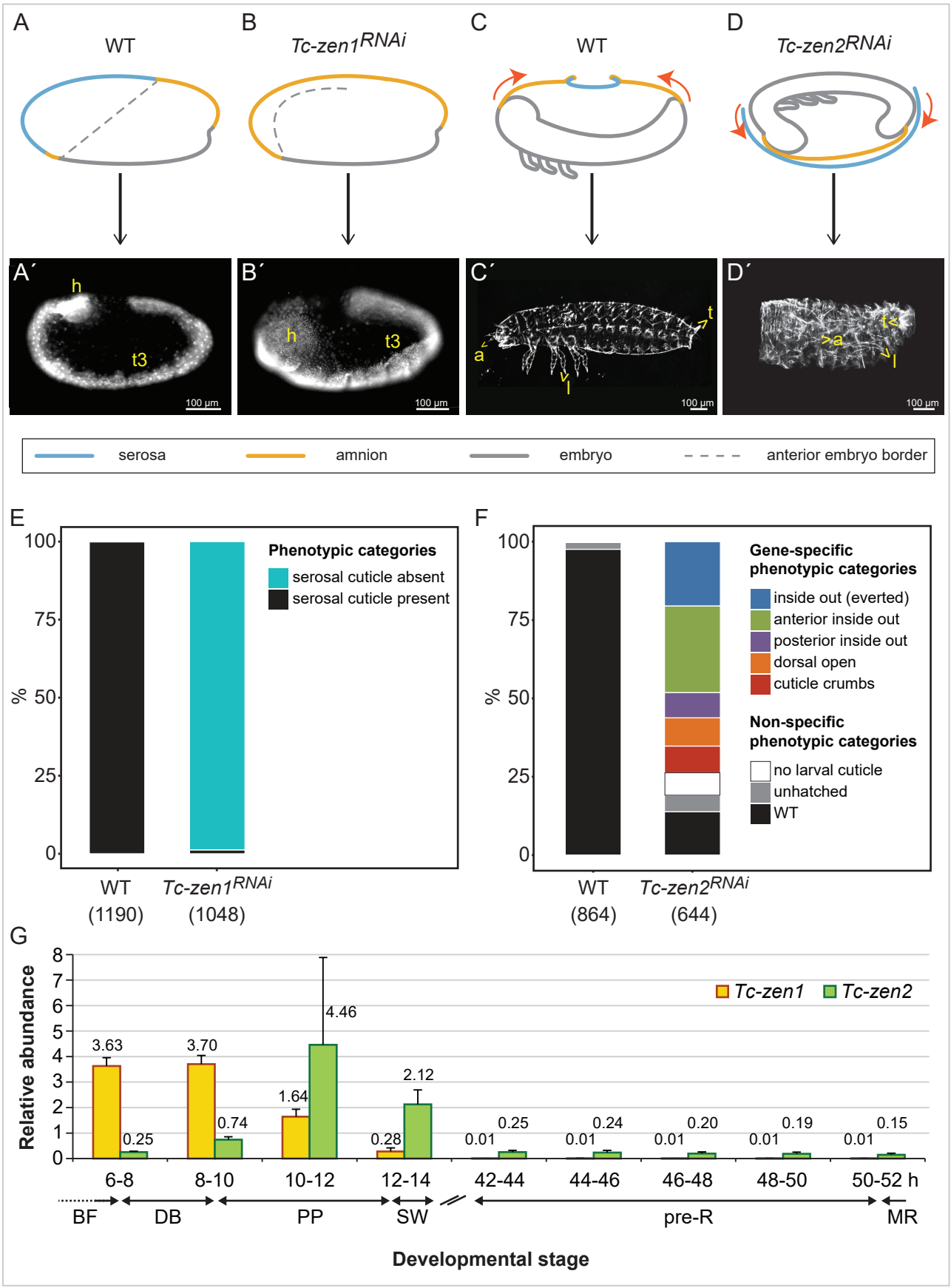

Fig 2. *Tc-zen* paralogue roles in early specification (*Tc-zen1*) or late EEM morphogenesis (*Tc-zen2*).

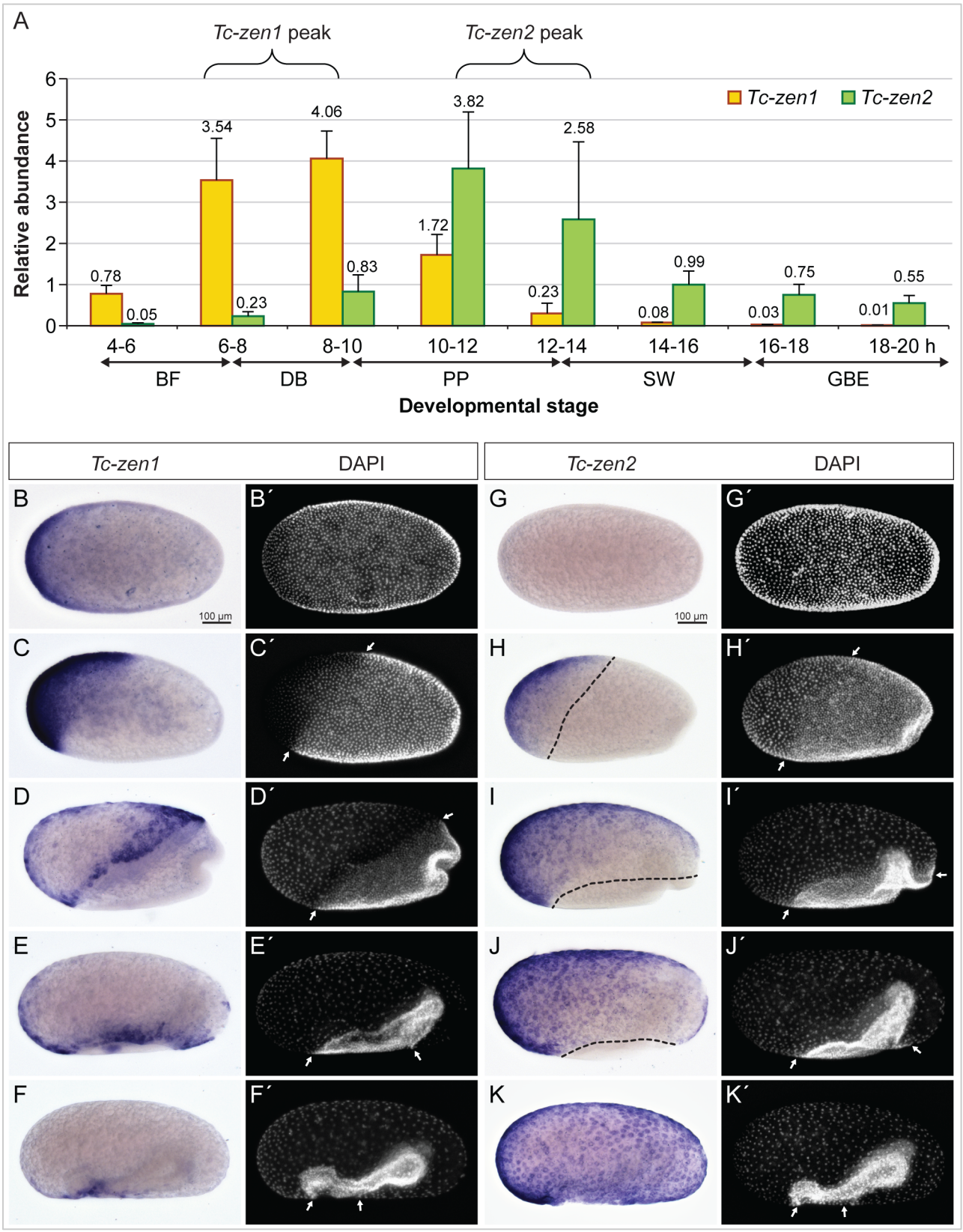

Fig 3. Transcript expression dynamics of the *Tc-zen* paralogues during early embryogenesis.

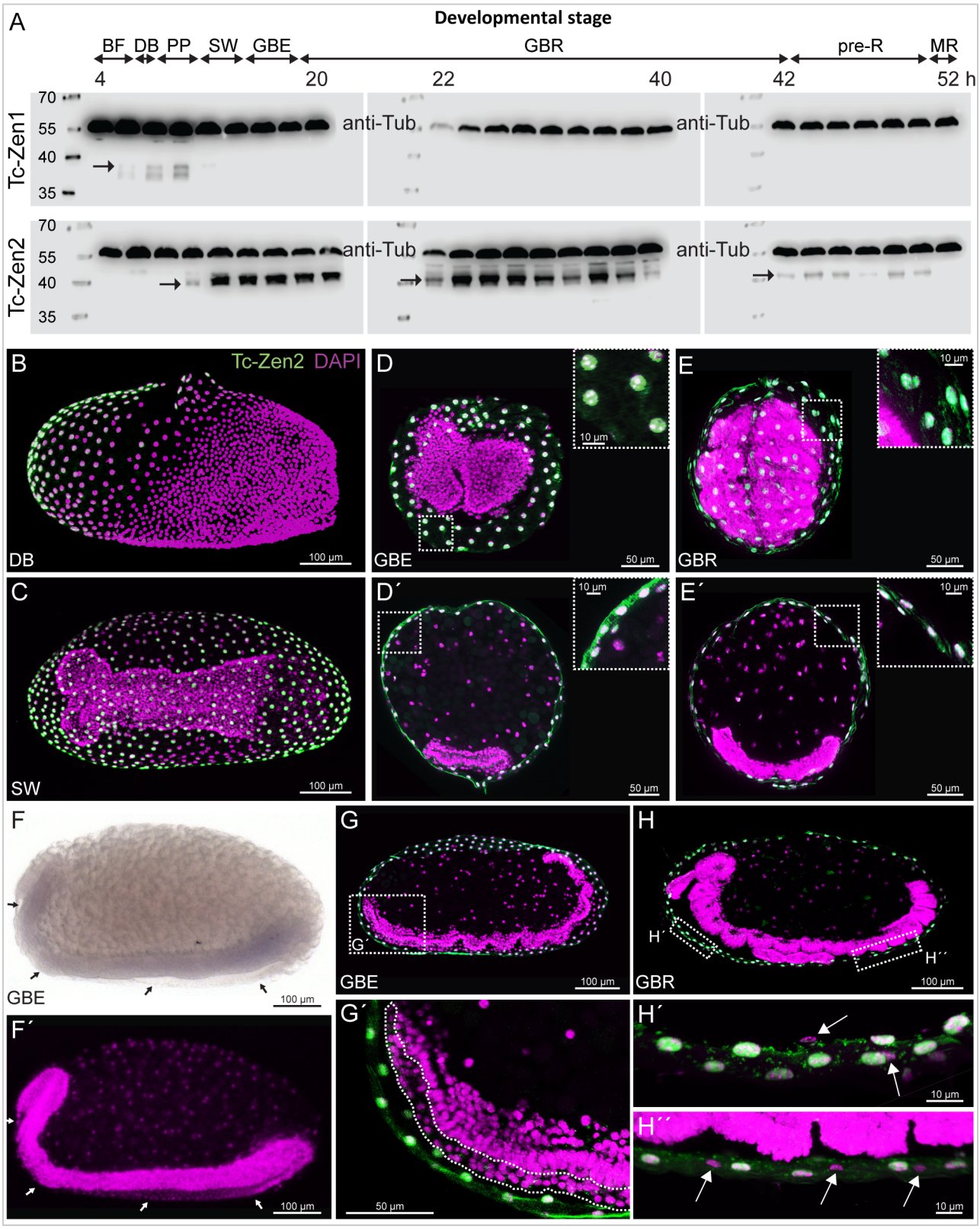

Fig 4. Paralogue-specific Tc-Zen protein expression time courses and spatial restriction of Tc-Zen2.

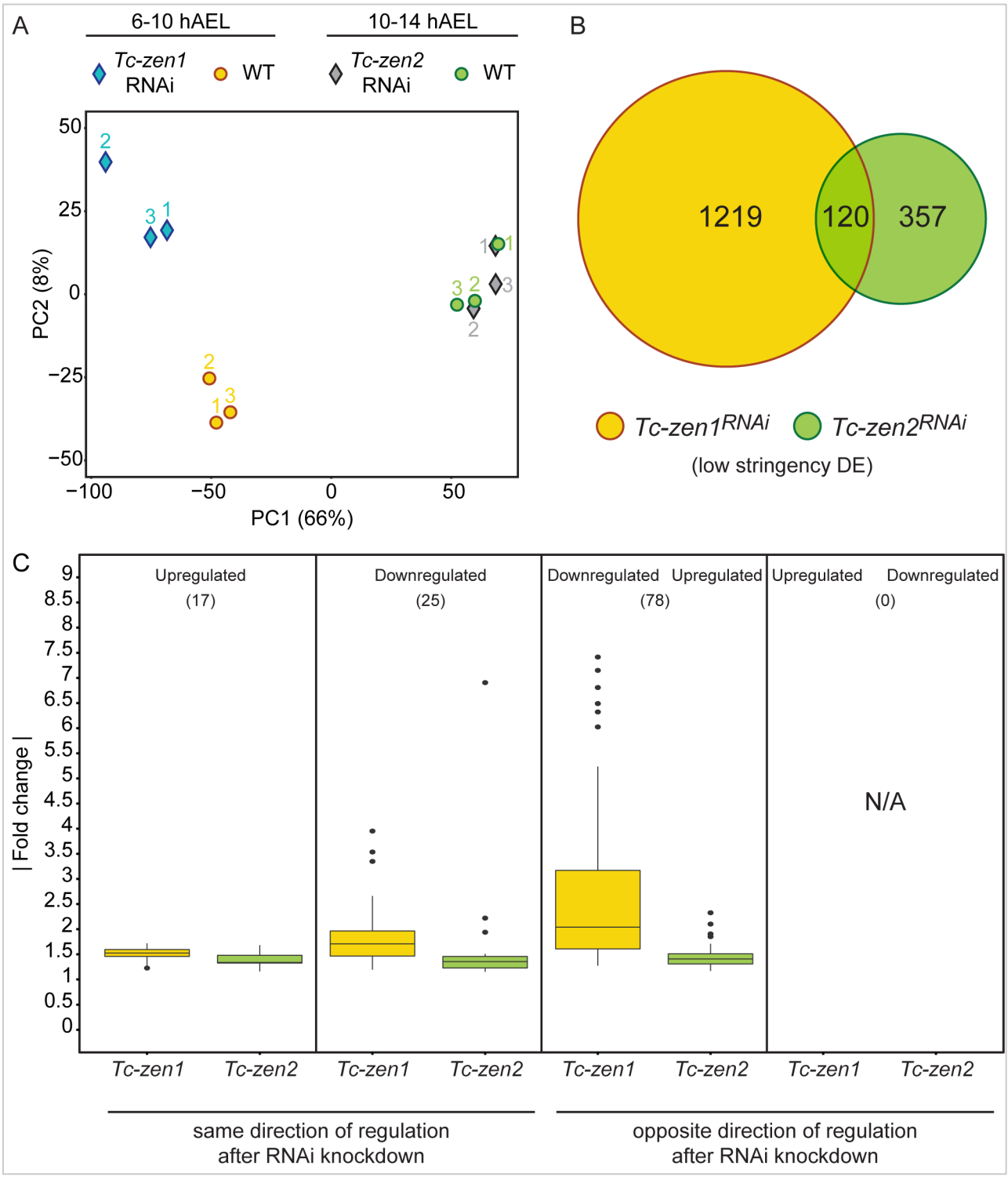

Fig 5. Early transcriptional control: relative impact and assessment of shared targets between the *Tc-zen* paralogues.

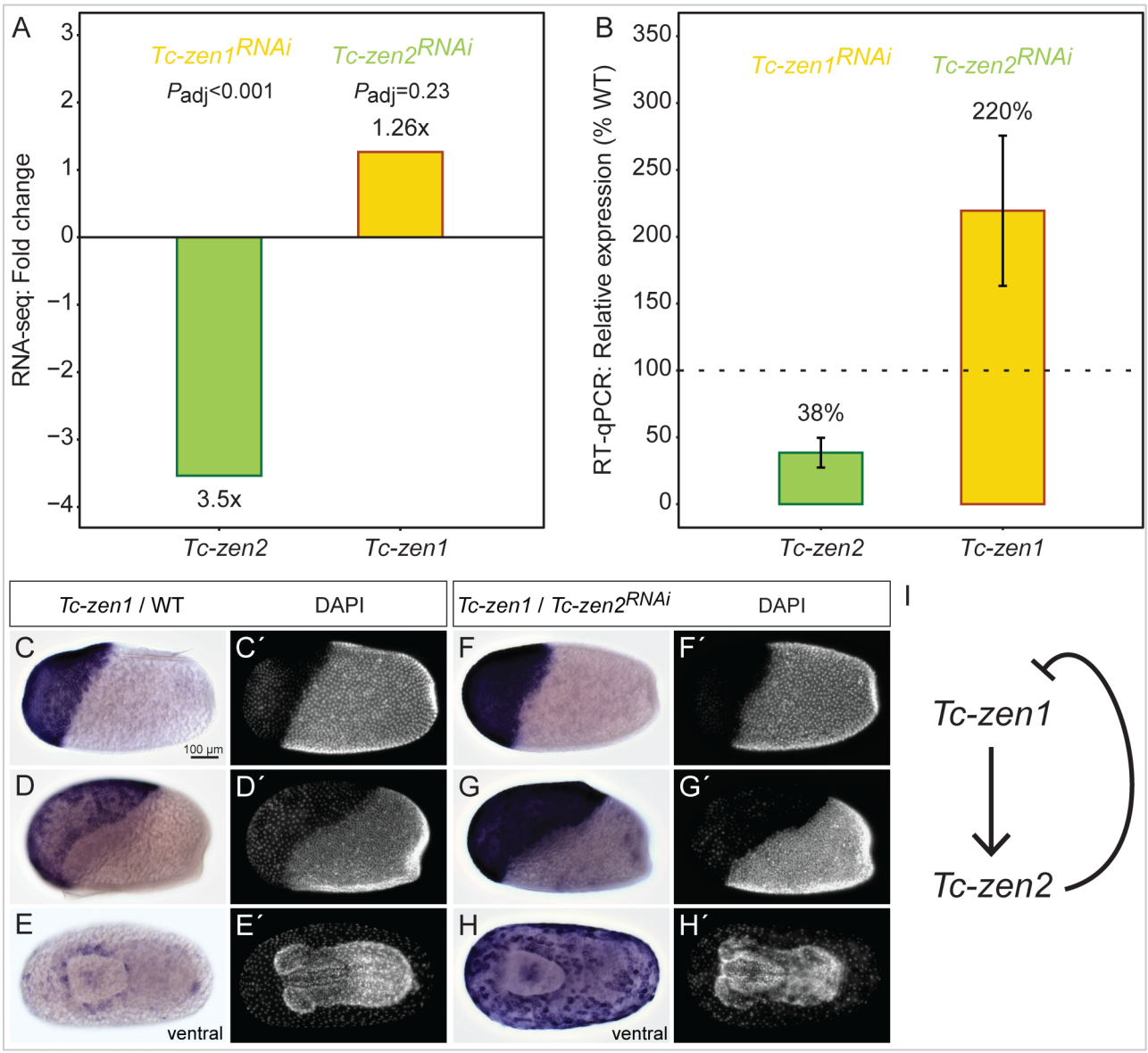

Fig 6. Activation of *Tc-zen2* by Tc-zen1 leads to *Tc-zen1*'s own repression.

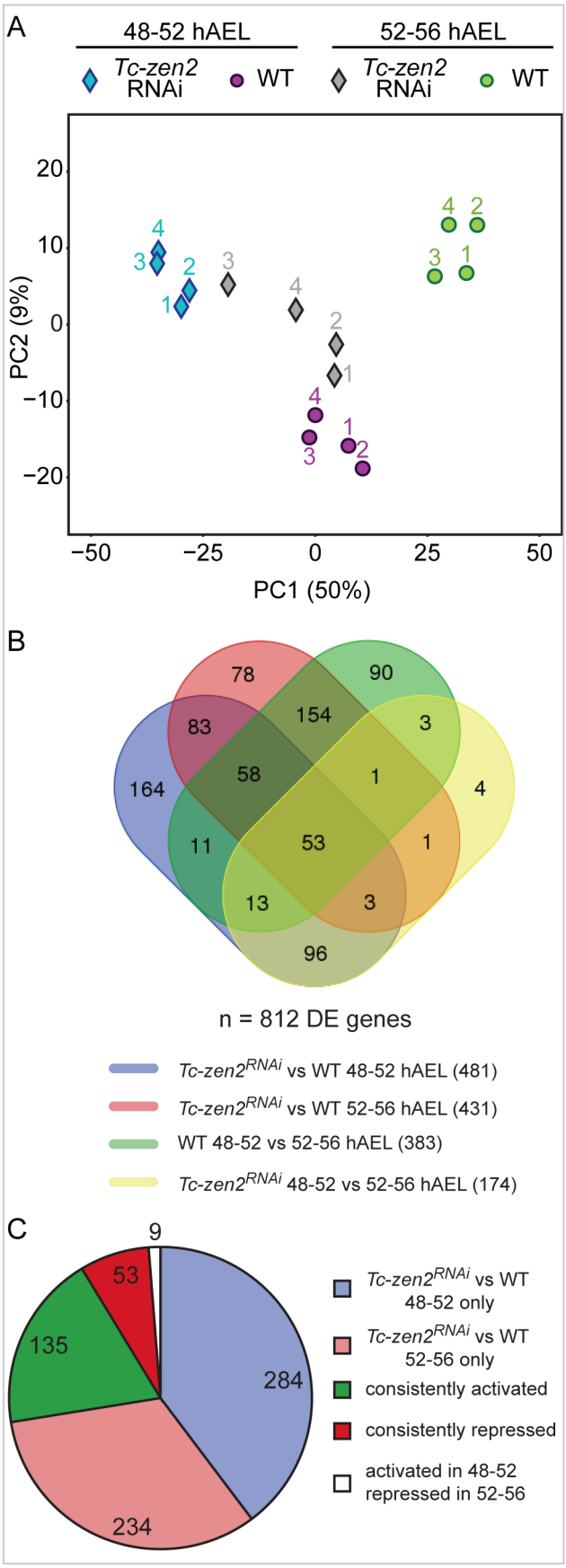

Fig 7. Tc-Zen2 has a strong, stage-specific transcriptional effect during late development.

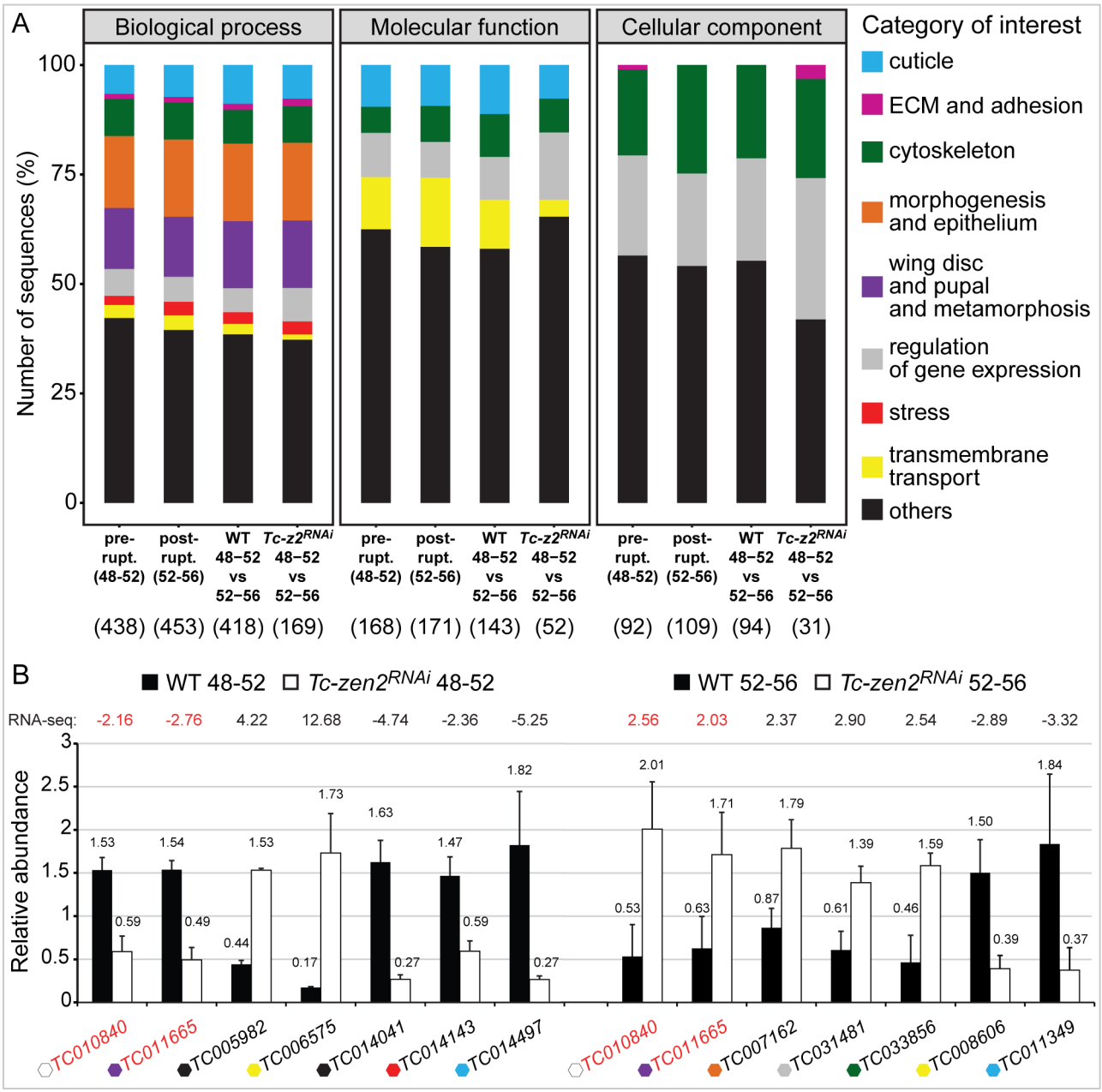

Fig 8. Functional annotation and validation of candidate late developmental targets of Tc-Zen2.