Long non-coding RNAs have become the focus of considerable interest over the past few years. Intriguing novel functions have been reported for lincRNAs. Three recent papers identify lincRNAs that work in a more conventional way—encoding protein—in each case a small polypeptide with an interesting biological activity (Magny et al, 2013; Pauli et al, 2014), (Bazzini et al, 2014).

These stories have a precedent, or two. In 2004, the Drosophila polar granule component gene (pgc) was reported to function as a non-coding RNA that acted in the embryonic primordial germ cells to prevent transcription of the zygotic genome (Martinho et al., 2004). pgc RNA localizes to the nascent germ cells in the embryo to transiently block activation of RNAPolII. A few years later, in 2008, the Nakamura and Ladurner laboratories reported that the functional product of the pgc gene was a peptide that blocks RNAPolIII by preventing an activating phosphorylation event mediated by P-Tefb (Hanyu-Nakamura et al., 2008; Timinszky et al., 2008). A second “former lineRNA” produced by the tarsal-less/polished rice/mille-pattes gene turns out to encode small peptides that control epithelial morphogenesis in Drosophila and Tribolium (Savard et al., 2006; Galindo et al., 2007; Kondo et al., 2007). Intriguingly, this peptide promotes N-terminal processing of the transcription factor Shavenbaby, converting it from a repressor to an activator (Kondo et al., 2010).

The recent report from the Couso laboratory built on their previous work on tarsal-less to search for additional Drosophila transcripts that might encode small peptides. They identified a lincRNA that is expressed in muscle and encodes small peptides (Magny et al., 2013). Interestingly, these peptides are related to the vertebrate peptides Sarcolpin and Phospholamban in sequence and predicted structure. Mutants lacking the fly lineRNA, which they name sarcolamban, show a defect in cardiac function. Based on the known role of the human peptides in calcium uptake by this family of small peptides. The Schier and Giraldez/Rajewsky laboratories now bring us full circle, assigning a protein-coding function to lincRNAs (Magny et al, 2013; Pauli et al, 2014), (Bazzini et al, 2014) (Fig 1).
“ORFscore” method performed well with annotated Zebrafish RefSeq transcripts. Applying ORFscore to 2,450 potential non-coding RNAs identified 190 transcripts with the potential to encode small polypeptides (20–100 aa) and further 89 predicted to encode longer peptides. This report introduces a second computational tool, called micPDP, which uses a conservation-based approach to identify novel small peptides. Of 63 conserved short Zebrafish peptides identified by micPDP, 23 were also found by ribosome profiling. Interestingly, a similar analysis of human small ORFs by the two approaches yielded even more limited overlap: seven small ORFs were identified by both methods (out of 173 found by micPDP and 135 by ORFscore).

These studies point to an emerging biology of small peptides. Why have these remained relatively obscure until now? One reason is statistical. Genome annotation has tended to filter out potential open reading frames because they are simply too numerous. Conservation across genomes can help predict function. But, as geneticists, we know that conservation is not a prerequisite for function. Likewise, there can be reasons other than production of a peptide that might explain ribosome binding to RNA, and indeed phased binding. The methods reported in these studies are tantalizing, but clearly much work is needed to validate the predictions and to explore function. To date, functions have been assigned to only a few of these former lncRNAs, but the versatile tools available for genome manipulation should allow a rapid follow-through. We can expect to be hearing a lot about the small protein world in years to come.

Conflict of interest
The author declares that he has no conflict of interest.

References

Figure 1. lincRNAs have been thought to act as non-coding molecules. Recent studies identified several lincRNAs that, unexpectedly, encode for small polypeptides.