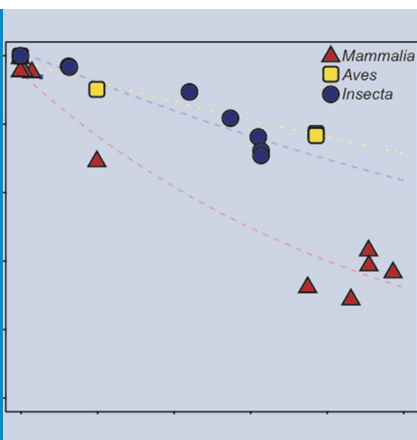


## COMPARATIVE GENOMICS

# One for all

**Using a common analysis pipeline to compare data from three major lineages of complex eukaryotes reveals that transcription seems to evolve at a common rate.**

DUNCAN T ODOM



**Related research article** Carvunis A-R, Wang T, Skola D, Yu A, Chen J, Kreisberg JF, Ideker T. 2015. Evidence for a common evolutionary rate in metazoan transcriptional networks. *eLife* 4:e11615. doi: [10.7554/eLife.11615](https://doi.org/10.7554/eLife.11615)

**Image** Genome sequences evolve more rapidly in mammals (red) than in birds (yellow) and insects (blue): x-axis is evolutionary separation from reference species; y-axis is a measure of genomic similarity

There is an old saying in computational circles that researchers in bioinformatics would rather use someone else's toothbrush than use someone else's code. One example of this adage being true can be seen in previous attempts to compare the rates at which differences in the mechanisms that control DNA accumulate in different species and lineages.

The information contained in DNA is first accessed by dedicated proteins called transcription factors (TF) that bind to preferred sequence of bases in the DNA. This sequence is typically short, between 8 and 20 bases in length (Vaquerizas *et al.*, 2009), although some can be as long as 35 bases (Filippova *et al.*, 1996). After transcription factor binding has taken place, the basal transcription machinery and its associated complexes open the region's chromatin and begin transcribing DNA into RNA. These crude transcripts must undergo extensive processing and maturation before they can be

exported to the cytoplasm as mature messenger RNA (mRNA). Understanding the rate at which all these steps (notably transcription factor binding and the production of mRNA) change during evolution is a long-standing goal in genetics (Wray, 2007; Wittkopp and Kalay, 2012).

Technically, it is (relatively) easy to map all the contacts between the transcription factors and the DNA, and also to map all the mRNA molecules, in a biological sample using high-throughput sequencing technologies. A number of research groups have compared the amount of transcription factor binding in many species of flies and mammals (He *et al.*, 2011; Paris *et al.*, 2013; Schmidt *et al.*, 2010; Ballester *et al.*, 2014). Based on this work it seemed as if transcription factor binding evolved rapidly in mammalian tissues (Weirauch and Hughes, 2010), but only very slowly in fruit flies (He *et al.*, 2011). However, it can be difficult to compare the first results generated in an entirely novel field of study because different groups often use very different approaches. And in this case this difficulty is further compounded by the toothbrush issue.

Now, in eLife, Trey Ideker and colleagues at the University of California San Diego – including Anne-Ruxandra Carvunis, Tina Wang and Dylan Skola as joint first authors – report that they used a new analysis pipeline to study the raw data for more than 25 species of complex eukaryotes across three animal lineages (mammals, birds and insects) that previously had only been studied in isolation (Carvunis *et al.*, 2015). In other words, they have cleaned everyone's teeth with the same toothbrush. Moreover, their pipeline could be tweaked to vary the analysis parameters for all the datasets across three

lineages at once, thus allowing them to make like-with-like comparisons.

This intellectual scrubbing resulted in two major insights. First, it appears that transcription factor binding (which dictates the function of the genome) and mRNA both evolve at a shared (and perhaps even fundamental) rate in complex eukaryotes. This result is somewhat surprising since most evolutionary geneticists think that the mechanisms that influence genome or functional evolution for the lineages studied by Carvunis et al. are radically different.

Second, particularly in mammals, the evolution of the genome sequence en masse is much more rapid than the evolution of transcription factor binding and transcription. This disconnect may be linked to the instability of the large number largely-silent repeat elements in mammalian genomes, and/or to the fact that insects and birds have more stable genomes.

Moreover, Carvunis et al. have powerfully demonstrated why it is important for all of us in the functional genomics community to meticulously curate our raw data and to make it readily available for others to analyse. None of the insights reported in this work would have been possible without easy access to carefully annotated sequencing reads from the original studies.

**Duncan T Odom** is an *eLife* Reviewing Editor and is in the Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom  
Duncan.Odom@cruk.cam.ac.uk

**Competing interests:** The author declares that no competing interests exist.

**Published** 11 February 2016

## References

**Ballester B**, Medina-Rivera A, Schmidt D, González-Porta M, Carlucci M, Chen X, Chessman K, Faure AJ, Funnell APW, Goncalves A, Kutter C, Lukk M, Menon

S, McLaren WM, Stefflova K, Watt S, Weirauch MT, Crossley M, Marioni JC, Odom DT, Flicek P, Wilson MD. 2014. Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. *eLife* **3**:e02626. doi: [10.7554/eLife.02626](https://doi.org/10.7554/eLife.02626)

**Carvunis A-R**, Wang T, Skola D, Yu A, Chen J, Kreisberg JF, Ideker T. 2015. Evidence for a common evolutionary rate in metazoan transcriptional networks. *eLife* **4**:e11615. doi: [10.7554/eLife.11615](https://doi.org/10.7554/eLife.11615)

**Filippova GN**, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenkov VV. 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Molecular and Cellular Biology* **16**:2802–2813. doi: [10.1128/MCB.16.6.2802](https://doi.org/10.1128/MCB.16.6.2802)

**He Q**, Bardet AF, Patton B, Purvis J, Johnston J, Paulson A, Gogol M, Stark A, Zeitlinger J. 2011. High conservation of transcription factor binding and evidence for combinatorial regulation across six drosophila species. *Nature Genetics* **43**:414–420. doi: [10.1038/ng.808](https://doi.org/10.1038/ng.808)

**Paris M**, Kaplan T, Li XY, Villalta JE, Lott SE, Eisen MB. 2013. Extensive divergence of transcription factor binding in drosophila embryos with highly conserved gene expression. *PLoS Genetics* **9**:e1003748. doi: [10.1371/journal.pgen.1003748](https://doi.org/10.1371/journal.pgen.1003748)

**Schmidt D**, Wilson MD, Ballester B, Schwalie PC, Brown GD, Marshall A, Kutter C, Watt S, Martinez-Jimenez CP, Mackay S, Talianidis I, Flicek P, Odom DT. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**:1036–1040. doi: [10.1126/science.1186176](https://doi.org/10.1126/science.1186176)

**Vaquerizas JM**, Kummerfeld SK, Teichmann SA, Luscombe NM. 2009. A census of human transcription factors: function, expression and evolution. *Nature Reviews Genetics* **10**:252–263. doi: [10.1038/nrg2538](https://doi.org/10.1038/nrg2538)

**Weirauch MT**, Hughes TR. 2010. Dramatic changes in transcription factor binding over evolutionary time. *Genome Biology* **11**:122. doi: [10.1186/gb-2010-11-6-122](https://doi.org/10.1186/gb-2010-11-6-122)

**Wittkopp PJ**, Kalay G. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics* **13**:59–69. doi: [10.1038/nrg3095](https://doi.org/10.1038/nrg3095)

**Wray GA**. 2007. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics* **8**:206–216. doi: [10.1038/nrg2063](https://doi.org/10.1038/nrg2063)