Genome **Biology**

RESEARCH HIGHLIGHT

Dramatic changes in transcription factor binding over evolutionary time

Matthew T Weirauch[1] and Timothy R Hughes[1,2]*

Abstract

A recent study reveals a surprisingly high degree of change in the occupancy patterns of two transcription factors in the livers of five vertebrates.

It is a long standing hypothesis that alterations in transcriptional regulation are a major driving force in evolution, and the results of many recent studies offer corroborating evidence (reviewed in [1]). Recent studies also indicate that *cis*-regulatory sequence is the major determinant of differences in transcriptional output among related species, as opposed to other influences, such as changes in transcription factor (TF) DNA binding domains, other chromatin factors, or external signals. Wilson *et al.* [2] showed that mouse liver cells containing human chromosome 21 'read' the human DNA in much the same way as do human liver cells, with the TFs hepatocyte nuclear factor (HNF)1A, HNF4A, and HNF6 all binding the same chromosome 21 locations that they would in human, rather than the locations bound in the orthologous mouse chromosome. However, important details have remained elusive, including the degree to which regulatory interactions vary between species across the entire genome, the types of mutations that are responsible for regulatory changes, and whether striking differences in TF binding occupancy are observed more generally among species. In a recent issue of *Science*, Schmidt *et al.* [3] now show that individual regulatory elements are frequently gained and lost among vertebrates and that local *cis*-regulatory point mutations can account for much of the evolution of transcriptional regulation.

In this study, the authors [3] performed chromatin immunoprecipitation sequencing (ChIP-Seq) analysis in order to determine the genomic occupancy of the strongly conserved TFs CCAAT/Enhancer binding protein α (CEBPA) and HNF4A in the liver tissues of five vertebrates (human, mouse, dog, opossum, and chicken). Both TFs are known to have important roles in liver gene regulation; in addition, liver expression patterns are mostly conserved across mammals, and liver contains a relatively small number of cell types, providing an ideal setup to compare TF occupancy in functionally and structurally orthologous cells. Surprisingly, their results [3] reveal that most TF binding is species-specific: for both TFs, only 10 to 20% of binding events are present in at least two of the three placental mammals (Figure 1a). Furthermore, only 6 to 8% of opossum CEPBA-bound regions are also found in mouse, dog, or human (Figure 1b); this value drops to 2% for chicken (Figure 1c), consistent with continuous transcriptional rewiring roughly corresponding to evolutionary distance [3]. Indeed, very little intergenic sequence is conserved between mammals and chicken, suggesting that this result will probably hold for most TFs and will also extend to amphibians and fish, which have even less sequence conservation with mammals.
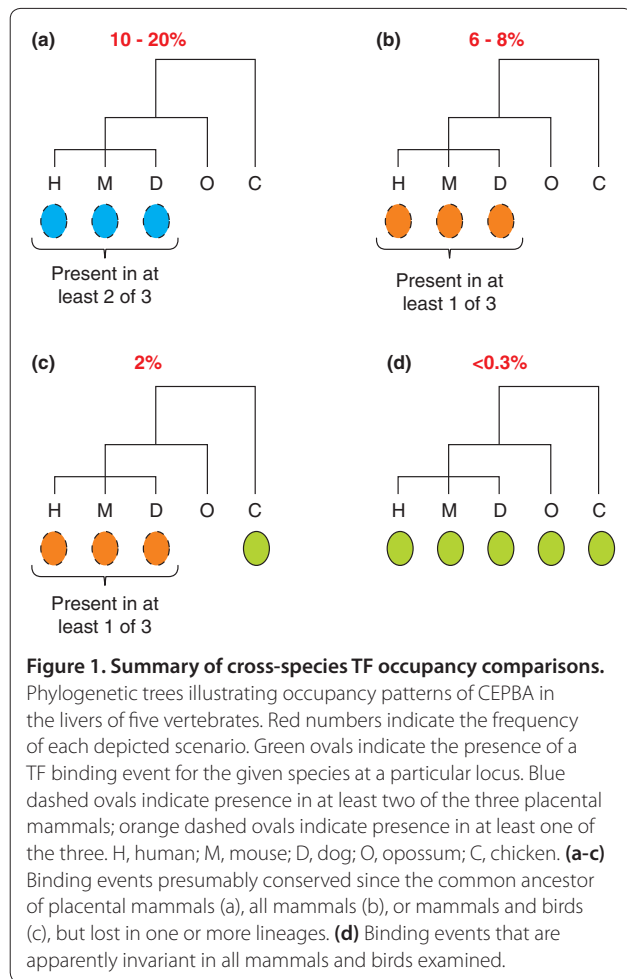
For both TFs, the majority of lineage-specific 'losses' (binding events not present in one placental mammal, but present at aligned, orthologous regions in the other two placental mammals) can be accounted for by either one or two point mutations (and not by insertions or deletions), suggesting that changes in TF occupancy are largely caused by the steady accumulation of small sequence changes [3]. Interestingly, a substantial proportion of losses (between 20% and 40%) occur at genomic locations with unchanged sequence composition at the TF binding site. Although changes in other *trans*-acting factors might have a role in these cases, another explanation could be the presence of local sequence changes that influence the chromatin state and/or the association of other factors (such as cofactors) with DNA.

Despite widespread evidence of binding site loss and gain, a small number of binding events were found to be 'ultra-shared' (present in all five species; Figure 1d). The relative scarceness of such events emphasizes the low sensitivity of comparative techniques such as phylogenetic

*Correspondence: t.hughes@utoronto.ca
[1]Banting and Best Department of Medical Research and Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada
[2]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3E1, Canada

BioMed Central

Figure 1. Summary of cross-species TF occupancy comparisons.
Phylogenetic trees illustrating occupancy patterns of CEPBA in
the livers of five vertebrates. Red numbers indicate the frequency
of each depicted scenario. Green ovals indicate the presence of a
TF binding event for the given species at a particular locus. Blue
dashed ovals indicate presence in at least two of the three placental
mammals; orange dashed ovals indicate presence in at least one of
the three. H, human; M, mouse; D, dog; O, opossum; C, chicken. **(a-c)**
Binding events presumably conserved since the common ancestor
of placental mammals (a), all mammals (b), or mammals and birds
(c), but lost in one or more lineages. **(d)** Binding events that are
apparently invariant in all mammals and birds examined.

footprinting for identifying *in vivo* binding sites.
However, these events were found to be almost always
located near known liver-specific genes, suggesting that
deep conservation of a binding event is indeed indicative
of functionality, in agreement with the fact that highly
conserved sequence is known to specifically identify
functional regulatory sequence. In contrast, the authors
[3] did not find a tendency for stronger binding events to
be preferentially conserved: neither the strength of match
to the consensus sequence nor sequencing read depth
correlate with sequence conservation. If conservation is a
measure of functionality, these results suggest that
stronger binding does not necessarily imply functionality,
a result compatible with evidence that weaker binding
sites are functionally important and that TFs can often
bind to a wide range of sequences.

The finding that TF binding events have diverged
rapidly throughout the vertebrate lineage [3] is consistent
with recent results comparing related yeasts [4] and
different human and yeast individuals [5-7]. In contrast, a
recent study comparing the genome-wide binding of six
TFs among two closely related *Drosophila* species reports

[8] that 'where we observe binding by a factor in one
species, we almost always observe binding by that factor
to the orthologous sequence in the other species'. What
factors might contribute to such strikingly different
findings? One possible explanation is that the observed
differences might be attributable to discrepancies in the
evolutionary distance separating the species analyzed in
each study. The *Drosophila* species of Bradley *et al.* [8]
have neutral substitution rates of approximately one in
ten bases, a rate much lower than that of the vertebrates
of Schmidt *et al.* [3] (about one in three among placental
mammals) and the yeast species of Borneman *et al.* [4]
(about one in four). With such low *Drosophila* substi-
tution rates, perhaps there simply has not been enough
time for changes in the regulatory sequences to accu-
mulate. However, this notion is inconsistent with the data
comparing different human and yeast individuals [5-7].
Furthermore, recent results comparing the global binding
patterns of RNA polymerase II between human and
chimpanzee, which have substantially lower substitution
rates than the two *Drosophila* species, also indicate that
as many as 32% of genes have diverged regulatory
programs [5].

An alternative explanation is that Bradley *et al.* [8]
focus on early embryogenesis, a developmental stage that
might be expected to be under stronger selection
constraints, whereas the other studies [3,5,6] analyze
samples taken from adult tissues. It is also possible that
some of the differences between conclusions reached by
different studies are due to differences in methodology of
data collection and analysis. For example, Bradley *et al.*
[8] identified binding event losses as those present in one
species (using a stringent threshold) and completely
absent in the other species (using a lenient threshold).
Accordingly, a binding event that is strong in one species
and weak in the other would be considered a 'conser-
vation' event by Bradley *et al.* [8] but a 'loss' event by
Schmidt *et al.* [3]. Other discrepancies might arise from
differences in false negative rates. If one study has a false
negative rate of 5%, the expected divergence rate for two
species with completely conserved binding events would
be 10% - a second study with a different false negative
rate would have a different expected divergence rate.
Finally, simulation studies have shown that TF binding
sites cannot be aligned accurately at many of the
divergence distances considered in the above studies,
resulting in the manifestation of binding site loss events
simply as a result of alignment errors. In the end, an
unbiased, methodologically uniform assessment compar-
ing the results of these studies would be greatly beneficial.
Ideally, such a study would address whether there is
evidence for selection acting to preserve binding events -
it is currently unclear how many conserved binding
events would be expected by chance alone.

Central to the significance of all of these studies [2-8] is the question of what proportion of individual TF binding sites are functional. Results from several recent ChIP-microarray (ChIP-chip) and ChIP-Seq studies (reviewed in [9]) demonstrate that many TFs bind promiscuously genome-wide, but that most binding events seem to have little influence on gene expression, echoing earlier results from yeast. Given the large number of binding events and mounting evidence supporting the transient nature of TF binding events, it is possible that most individual TF binding sites have limited functional importance. Furthermore, given that 30 to 50% of CEBPA and HNF4A binding site sequences overlap in the genome, many binding events might be non-functional interactions with accessible motifs in regions of open chromatin - in yeast, nucleosome depletion is a strong predictor of where TFs will bind.

Deciphering the determinants of TF binding and their relationship to gene expression output will be important for understanding both the function and the evolution of transcriptional regulatory mechanisms. Nonetheless, the findings of Schmidt *et al.* [3] offer intriguing insights not only into the evolution of transcriptional regulation, but into evolution itself. At first glance, it might seem somewhat surprising that something as important as TF binding sites is evolving so rapidly. However, assuming that gene regulation occurs by ensembles of modules that act largely independent of one another - a model that is supported by a wealth of evidence [10] - most losses (and gains) of individual binding sites are likely to have a small effect on overall transcriptional output. In such a model, the vast majority of individual TF binding sites would be disposable over the long term, because compensatory sites would also arise frequently, resulting in the accumulation of point mutations disrupting individual binding sites at near-neutral rates. The ability to tolerate such changes could also increase an organism's capacity to generate heritable phenotypic variation, and so increase overall 'evolvability'. The fluidity of eukaryotic transcriptional regulatory regions may therefore enable the exploration of potentially beneficial new regulatory sequence configurations.

References
1. Carroll SB: Evolution at two levels: on genes and form. *PLoS Biol* 2005, **3**:e245.
2. Wilson MD, Barbosa-Morais NL, Schmidt D, Conboy CM, Vanes L, Tybulewicz VL, Fisher EM, Tavare S, Odom DT: Species-specific transcription in mice carrying human chromosome 21. *Science* 2008, **322**:434-438.
3. 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: Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 2010, **328**:1036-1040.
4. Borneman AR, Gianoulis TA, Zhang ZD, Yu H, Rozowsky J, Seringhaus MR, Wang LY, Gerstein M, Snyder M: Divergence of transcription factor binding sites across related yeast species. *Science* 2007, **317**:815-819.
5. Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM, Habegger L, Rozowsky J, Shi M, Urban AE, Hong MY, Karczewski KJ, Huber W, Weissman SM, Gerstein MB, Korbel JO, Snyder M: Variation in transcription factor binding among humans. *Science* 2010, **328**:232-235.
6. McDaniell R, Lee BK, Song L, Liu Z, Boyle AP, Erdos MR, Scott LJ, Morken MA, Kucera KS, Battenhouse A, Keefe D, Collins FS, Willard HF, Lieb JD, Furey TS, Crawford GE, Iyer VR, Birney E: Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* 2010, **328**:235-239.
7. Zheng W, Zhao H, Mancera E, Steinmetz LM, Snyder M: Genetic analysis of variation in transcription factor binding in yeast. *Nature* 2010, **464**:1187-1191.
8. Bradley RK, Li XY, Trapnell C, Davidson S, Pachter L, Chu HC, Tonkin LA, Biggin MD, Eisen MB: Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related *Drosophila* species. *PLoS Biol* 2010, **8**:e1000343.
9. Farnham PJ: Insights from genomic profiling of transcription factors. *Nat Rev Genet* 2009, **10**:605-616.
10. Arnosti DN, Kulkarni MM: Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* 2005, **94**:890-898.