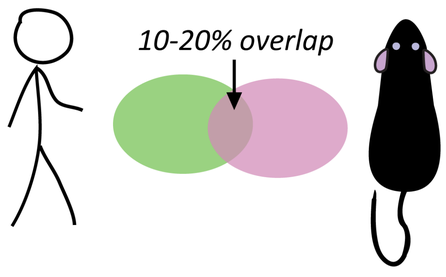
Evolutionary mechanisms and divergence rates of tissue-specific transcription factor binding revealed by analysis in multiple, closely spaced rodent species

David Thybert[1], Klara Stefflova[2]*, Michael D Wilson[2], John C Marioni[1], Dave Adams[3], Duncan T Odom[2], Paul Flicek[1]*

[1]European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, UK;
[3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK.

Introduction



TFBR overlap between mouse and human

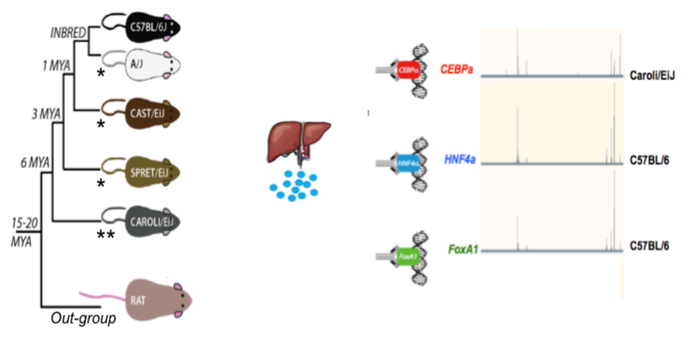
While the form and function of liver is well preserved through 80 million years (MY) of mammalian evolution, the binding of transcription factors (TFs) is not well conserved: only 10-20% of transcription factor binding regions (TFBR) is conserved between human and mouse

To understand the mechanisms behind this rapid divergence we investigate the earliest stages of TF binding divergence by interrogating TF binding evolution at short evolutionary scales in five inbred mouse lines separated by less than six million years of evolution.

System

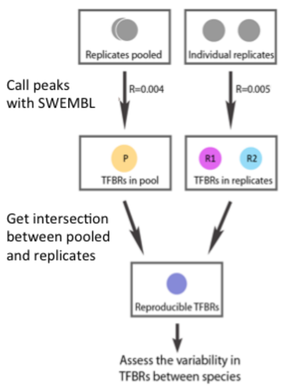
We use ChIP-Seq to investigate the binding of 3 TFs (CEBPa, HNF4a, FoxA1) in liver samples taken from 5 inbred mouse strains separated by < 6MY of evolution and rat used as outgroup

Comparing TF binding at this short evolutionary timescale allows for more precise estimates of the rate of TF binding evolution and provides insights into the underlying mechanisms



** From study Keane et al. 2011*
*** Sequenced specifically for this project*

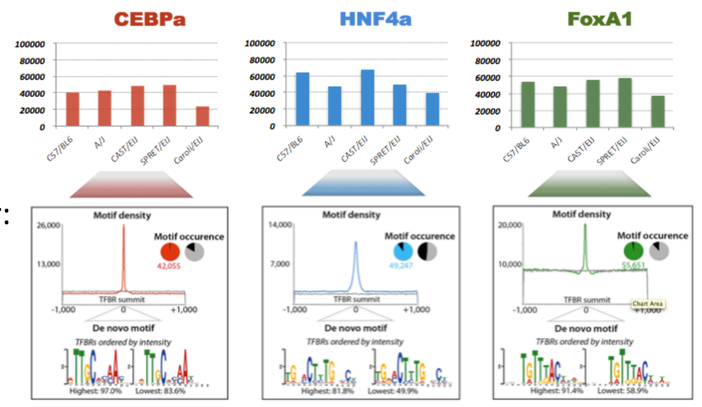
Between 40-50,000 binding region identified for each factor



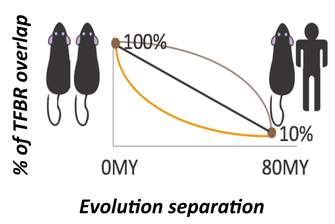
High confident and reproducible TFBR

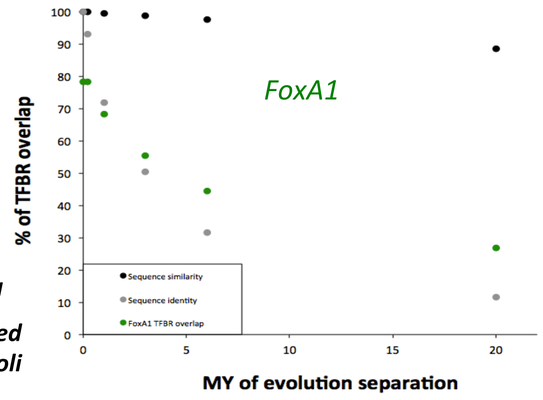
Average of peak by factor:

40,962 CEBPa
53,808 HNF4a
50,551 FoxA1



Rapid Decay of TFBR overlap at short time evolutionary scale

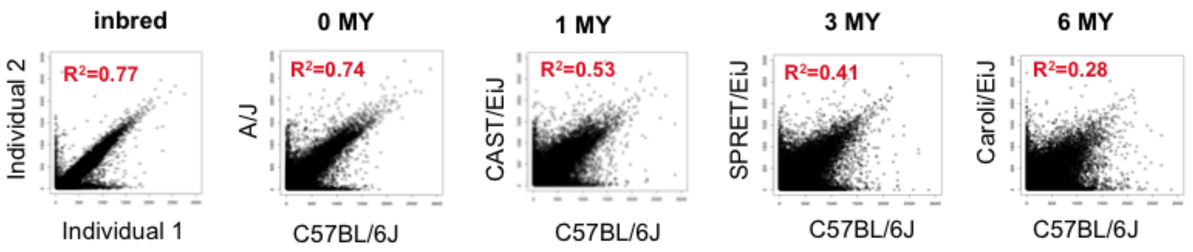




We observe an exponential decay of TF binding

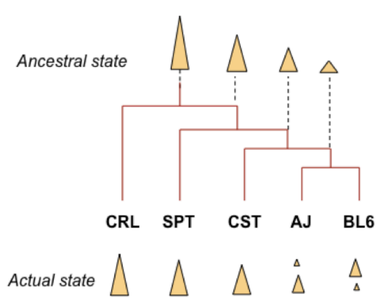
Less than 45% overlap for three liver-expressed TFs in the livers of Mus musculus and Mus caroli species.

How TF binding intensity correlation evolve at this evolutionary time scale



- Intra-strain variability in TF binding intensity is minimal
- Across 6 MY of evolution the correlation decays rapidly
- This is robust to differences in genome quality as assessed by mapping the C57BL/6J data to all of the different genomes

Is the loss/gain of TF binding a stochastic or gradual evolutionary process?



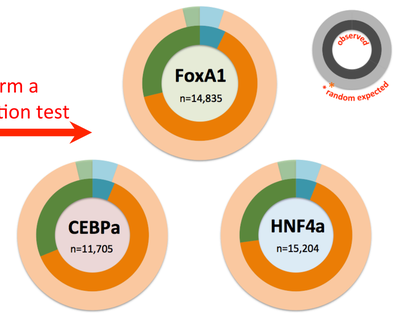
More TF binding than randomly expected have a conserved intensity.
→ TF binding intensity have a sense during evolution

Infer ancestral TFBR state using parsimony approach

3 categories of BR

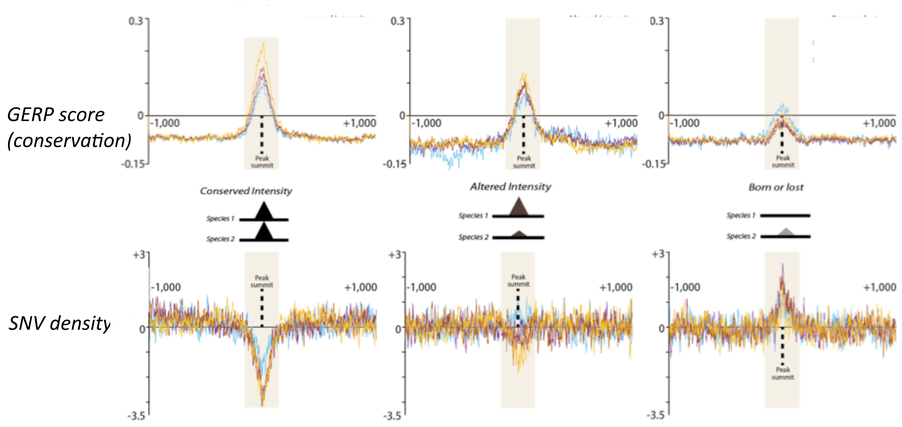
- Directional
- Stochastic
- Conserved

Perform a permutation test



Intensity has been scaled in 10 intensity classes using a factorial scale . Two binding regions are considered to have different intensity if they have at least two classes of difference

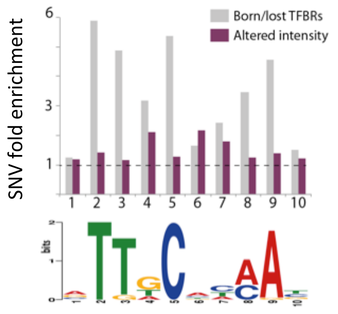
Is Cis-effect related with TF binding variation



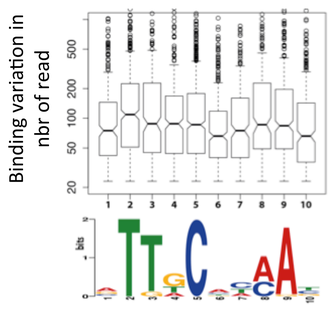
We can observe a strong correlation of the SNV density around the peak summit and the degree of TFBR variability between species.
→ Proximal cis effects are clearly involved in TFBR turnover

How Single Nucleotide Variation on motif affect TF binding

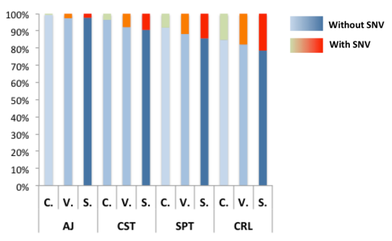
SNV Enrichment in the CEBPa motif of variable TF binding



Effect of SNV in CEBPa binding intensity variation



Proportion of CEBPa Binding Region with at least 1 SNV in the motif

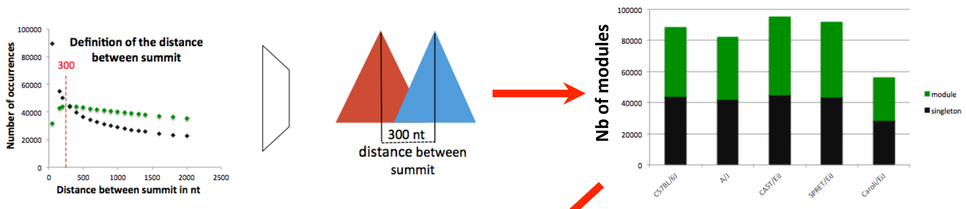


C: Shared with common intensity
V: Shared with variable intensity
S: Unshared

We observe a strong relation of SNV on conserved position of the motif and the TF binding variability. However SNV on motif explain a very small part of the variability.

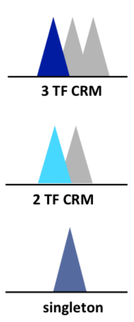
How TF binding evolve in the context of cis regulatory module?

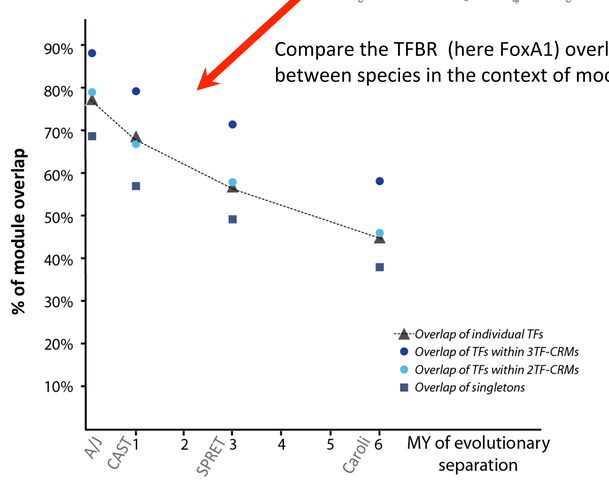
The clusters of TF binding (CRM) are defined by TFBR separated by less than 300 nt summit to summit



Compare the TFBR (here FoxA1) overlap between species in the context of module

Classification of TFBR regarding the number of different factor in the CRM

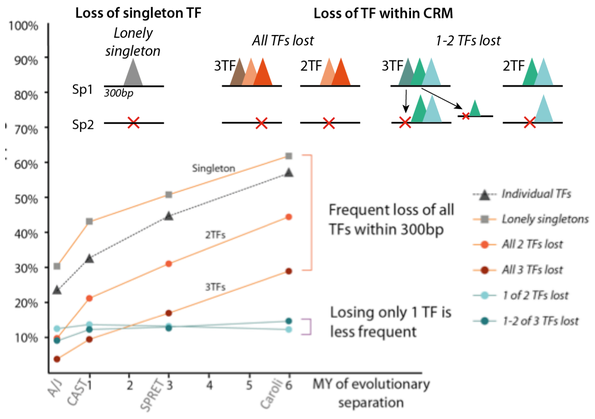




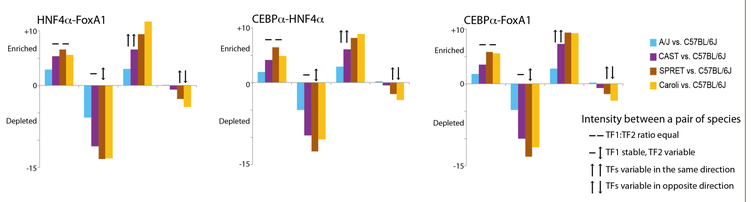
More a TF binding is associated in module to other factor and more conserved it is. However the evolution rate stays the same.

Is TF binding co-evolving within modules?

Whether the gain/lost of a TFBR is associated to the gain/lost of other TFBR in the same CRM?



Do TF binding intensity coevolve within a module?



TF bindings of one factor are coevolving with the other TF binding in the same CRM in term of gain/lost and intensity of TF binding.

Conclusion

By taking advantage of both the system and the high quality of ChIP-seq data, this study of TFBR evolution in five inbred mouse lines separated by less than 6 MY provides insight into the very fast turnover of TFBR of single factors and modules. The analysis of TF binding intensity shows that a fast loss of TF binding intensity correlation occurs between species within 6 MY of evolution. This TF binding intensity variation seems to happen in a stochastic way rather than a progressive way such that gain/loss of TFBR seems to happen in a stochastic manner. The SNV analysis shows clearly that a *cis* effect on motif is involved in the TFBR turnover mechanism but explain a very small number of variation. The co-variation of TF binding within the same CRM suggest that cooperativity between factor or chromatin accessibility are significantly involve in the process of turnover of TF binding during the evolution.