## Transcription factor binding

**Most transcription factor binding events differ between closely related species despite conservation of functional targets. Now, Paul Flicek and Duncan Odom and colleagues report an evolutionary comparison of transcription factor CEBPA binding among five vertebrate species (*Science* published online, doi:10.1126/science.1186176, 8 April 2010). The authors compared genome-wide binding of CEBPA as determined by ChIP-sequencing in livers of human, mouse, dog, short-tailed opossum and chicken. The CEBPA binding motif was virtually identical between the five species, but most binding events were species-specific. Only 10% to 22% of binding events were shared between any two placental mammals, and there was even greater divergence of binding events in opossum and chicken. Only 35 binding events were shared by all five species. The authors were able to account for more than half the lost binding events (those absent in one placental mammal and present in two other placental mammals) by identifying base-pair substitutions and indels that disrupted alignable binding motifs. About half the lost binding events showed evidence of nearby compensatory binding events. The authors suggest that many, if not most, interactions between transcription factors and DNA are evolving neutrally.** ***EN***

## Heritable chromatin signatures

Allele-specific gene expression is thought to contribute to phenotypic variation, but the genetic mechanisms responsible are not well understood. Ewan Birney and colleagues analyzed the heritability of individual-specific and allele-specific binding of the transcription factor CTCF and DNase I in two unrelated trio families (*Science* **328**, 235–239, 2010). Binding sites were categorized as 'constant' (present in the four parents), 'individual-specific' (present in at least two of the parents but absent in the other two), or 'singleton' (present in one person). The signals of the two children were more similar to those of their respective parents at sites not present in the unrelated parents, providing evidence that chromatin signatures are heritable. The authors next assessed allele-specific binding (at sites containing a heterozygous SNP) and found that 7% of DNase I and 11% of CTCF sites showed significant allele specificity. To analyze the heritability of allele specificity, the authors looked at allele-specific signals in each child at sites where each parent was homozygous for the different alleles. The allele with a stronger CTCF signal in the child was most often (65%) the same allele carried by the parent with stronger CTCF binding. These results suggest a genetic basis for allele-specific chromatin structure. *PC*

## Lgr6 marks epidermal stem cells

The different compartments of the mammalian skin are thought to be maintained by distinct stem cell populations, but the location and identity of these cells have been heavily debated. Hans Clevers and colleagues (*Science* **327**, 1385–1389, 2010) now show that one such stem cell population in mouse skin is marked by expression of Lgr6, a homolog of the well studied stem cell marker Lgr5. The authors generated knock-in alleles at the *Lgr6* locus, allowing them to visualize Lgr6$^+$ cells and their progeny. Shortly after birth, Lgr6 expression in the skin becomes restricted, eventually marking a unique population located at the central isthmus just above the bulge region in resting hair follicles. Through a series of lineage-tracing studies in adult mice, they showed that the progeny of Lgr6$^+$ cells contribute not only to hair follicles but also to sebaceous glands and interfollicular epidermis. They also showed that transplanted Lgr6$^+$ stem cells could reconstitute fully formed hair follicles and contribute to all three lineages in host skin. Finally, they showed that Lgr6$^+$ cells contribute to wound healing. These findings identify Lgr6 as a marker of a distinct epidermal stem cell population contributing to all three skin lineages. *KV*

## Identifying nonobvious phenologs

Conservation of gene function among organisms is a fundamental principle of biology that allows the use of model organisms to investigate problems in human biology and disease. Edward Marcotte and colleagues now report a systematic approach (*PNAS* **107**, 6544–6549, 2010) for identifying orthologous phenotypes (phenologs), which are two phenotypes in different organisms that have a significant degree of overlap in associated genes. The authors obtained large sets of gene–phenotype associations from the literature in human, mouse, *Caenorhabditis elegans*, *Arabidopsis thaliana* and yeast and quantified the overlap in orthologs in each interorganism phenotype pair. After correction for multiple testing, the authors found >6,200 significant phenologs. These phenologs may suggest nonobvious models for human disease, and the authors demonstrated that a phenolog between defective angiogenesis in mice mutants and reduced growth rate in yeast mutants grown in the hypercholesterolemia drug lovastatin can effectively predict genes required for angiogenesis in vertebrates. The authors discovered another nonobvious phenolog between genes involved in plant response to gravitational cues and genes mutated in Waardenburg syndrome, which arises from defects in neural crest cell migration. Using this phenolog, the authors suggest that *SEC23IP* is a promising Waardenburg syndrome gene. *PC*

## Personal genome

**In the past year, we have seen a handful of newly reported individual genome sequences, with the suggestion that whole genome sequencing could prove useful in genetic and clinical diagnoses. Richard Gibbs and colleagues now report the whole genome sequence and genetic diagnosis of an individual with Charcot-Marie-Tooth disease (CMT) (*N. Engl. J. Med.* 36, 1181–1191, 2010). The authors studied a family with an inherited neuropathy that included four affected siblings who had previously been screened for mutations in several of the more common CMT-associated genes. They sequenced the genome of the proband at ~30× coverage with the SOLiD system (Applied Biosystems) and identified compound heterozygous mutations in *SH3TC2*, a gene with a known role in CMT. Sequencing of *SH3TC2* in family members showed segregation of the mutations with disease and confirmed these alleles as causative for CMT in this family. The authors also examined copy number variation (CNV) using several different methods. They identified 234 CNVs in the proband compared to a male control, but found no CNVs affecting genes with a known role in CMT or other neuropathies.** ***OB***

*Written by Orli Bahcall, Pamela Colosimo, Emily Niemitz and Kyle Vogan*