

## Replication timing-associated variants

Locus-specific differences in DNA replication timing have been associated with various genomic features, including altered mutation rates. To study the impact of genetic variation on genome-wide patterns of human DNA replication timing, Steve McCarroll and colleagues (*CeIl* 159, 1015–1026, 2014) used a sequence-based approach to construct replication timing profiles for proliferating lymphoblastoid cell lines derived from 161 individuals from the 1000 Genomes Project. Overall, they identified 477 genomic regions exhibiting individual-level variation in DNA replication timing. They then tested whether variation in replication timing in these regions was associated with DNA sequence variation. Association analyses using the 161 discovery samples and an additional 334 independent samples identified 16 distinct genetic loci influencing DNA replication timing in *cis*. In addition, among the loci that fell just below the threshold for genome-wide significance, they observed a signal near *JAK2* that coincided with a haplotype previously associated with predisposition to somatic *JAK2* mutations and risk of myeloproliferative neoplasms. Notably, the haplotype associated with predisposition to somatic *JAK2* mutations is associated with earlier replication timing, with the elevated mutation rate possibly driven by interference between the replication and transcription machineries. **KV**

## Anopheline mosquito genomes

Daniel Neafsey, Nora Besansky and colleagues report the whole-genome sequencing of 16 anopheline mosquito species, representing a diversity of geographical locations, ecological conditions and vectorial capacities for the transmission of malaria (*Science* doi:10.1126/science.1258522; 27 November 2014). They also analyzed the transcriptomes from pooled male and female larvae, pupae and adults. For each species, they identify between 10,738 and 16,149 protein-coding genes. They find extensive shuffling of gene order, more frequent rearrangements on the X chromosome and high rates of gene gain or loss. In an accompanying report, Michael Fontaine, Nora Besansky and colleagues use the new genomic data to resolve the phylogeny of the *Anopheles gambiae* complex (*Science* doi:10.1126/science.1258524; 27 November 2014). They find that the ancestor of two primary malaria vectors, *An. gambiae* and *Anopheles coluzzii*, separated from other species about 2 million years ago and is distantly related to another primary malaria vector, *Anopheles arabiensis*. They identify pervasive autosomal introgression between these vectors, with the majority of genetic transfer in the direction from *An. arabiensis* into the ancestor of *An. gambiae* and *An. coluzzii*. **OB**

## Regulation of bitterness in cucumber

Sanwen Huang and colleagues report the identification of two transcription factors necessary for bitterness in the fruit and leaves of cucumber (*Science* 346, 1084–1088, 2014). The previously identified loci *Bitter* (*Bi*) and *Bitter fruit* (*Bt*) control bitterness, which is mainly due to cucurbitacin C (CuC), in cucumber and related plants. The authors performed genome-wide association analysis for variants associated with *Bi* using the cucumber genome variation map. They identified *Csa6G088690*, encoding an oxidosqualene cyclase, as the cucumber *Bi* gene. To further elucidate the mechanism for CuC

biosynthesis, the authors screened a cucumber ethyl methanesulfonate (EMS) library for plants with non-bitter leaves. They sequenced the genomes of two mutants and identified a SNP in *Csa5G156220*, which encodes a transcription factor expressed exclusively in leaves. They found that this gene, renamed *Bitter leaf* (*Bl*), regulates the expression of *Bi*. Eleven variants in a nearby homolog of *Bl* (*Csa5G157230*) associated with extreme bitterness in a local association analysis, and one variant cosegregated with *Bt* in an F<sub>2</sub> population, indicating that it is likely the *Bt* gene. The authors confirmed that *Bt* regulates *Bi* expression in fruits. Finally, the authors identified nine genes downstream of *Bl* and *Bt* in the CuC biosynthetic pathway through a candidate gene RNA interference screen. **BL**

## Clonal evolution of B-ALL

Stephen Quake and colleagues report a method for reconstructing the clonal architecture of pediatric B cell acute lymphoblastic leukemia (B-ALL) samples using a microfluidics platform and novel computational approach (*Proc. Natl. Acad. Sci. USA* doi:10.1073/pnas.1420822111; 25 November 2014). The authors performed exome sequencing on bone marrow and matched saliva samples collected during remission from six children diagnosed with B-ALL. Mutations in regions of increased tumor heterogeneity in the exome data were confirmed through targeted sequencing. The authors found 46 mutations per individual, on average, but a wide variability between individuals. They carried out whole-genome amplification for 1,479 single cells from the 6 children to separate mutations into specific clones and reconstruct the clonal evolution of the sample using, in parallel, a model-based probabilistic algorithm and data-driven clonal clustering. These methods, which were designed to filter out low-quality single-cell sequences, were validated using simulation tests. Among their findings, the authors determined that the majority of structural alterations occurred before point mutations and found that *KRAS* mutations occurred at later stages of disease. Evaluation of interclonal IgH sequences showed that clones within the same individual might arrest at different stages in B cell differentiation. **BL**

## Noncoding mutation in T-ALL

Overexpression of the *TAL1* oncogene occurs frequently in T cell acute lymphoblastic leukemia (T-ALL), but the mechanism driving overexpression is not known in some cases. Now, A. Thomas Look, Richard Young and colleagues show that somatic mutation of a noncoding element creates a binding site for the MYB transcription factor and drives expression of the *TAL1* oncogene (*Science* doi:10.1126/science.1259037; 13 November 2014). The authors first observed a broad region of aberrant histone H3 lysine 27 acetylation (H3K27ac) encompassing the *TAL1* transcriptional start site in Jurkat cells, a T-ALL cell line with unexplained overexpression of *TAL1*. They identified a 12-bp insertion in the region and further identified indels in the *TAL1* enhancer in 8 of 146 T-ALL clinical samples. They determined that all of the indels created MYB binding motifs, driving expression in reporter assays, and showed by chromatin immunoprecipitation that MYB bound to the *de novo* site in Jurkat cells. The authors also used CRISPR/Cas9 technology to show that disruption of the insertion site in Jurkat cells abolished *TAL1* expression and H3K27ac. This work provides an example of the *de novo* generation of an oncogenic enhancer. **EN**

Written by Orli Bahcall, Brooke LaFlamme, Emily Niemitz & Kyle Vogan