Single-haplotype genome assembly

Karyn Meltz Steinberg, Richard Wilson and colleagues produced a single-haplotype assembly of the human genome using genomic DNA from an essentially haploid hydatidiform mole (*Genome Res.* doi:10.1101/gr.180893.114; 4 November 2014). They used a short read–based, reference-guided method to assemble the complete genome of the CHM1 hydatidiform mole, using 100× whole-genome shotgun sequence reads and the GRCh37 reference assembly. The assembly was then improved through the integration of high-quality finished sequence from 382 BAC clones, selected to overlap complex regions with segmental duplications or where the reference had known gaps. They demonstrate that the CHM1_1.1 assembly is of high quality, with a contig N50 length of 144 kb and a scaffold N50 length of 50 Mb, making it more contiguous than previous human whole-genome assemblies. Although CHM1_1.1 is of high quality, the authors note errors in it, including ones due to the use of the GRCh37 reference. They stress the value of using multiple sequencing technologies and assembly-independent resources, demonstrated by alignment to a more recently available long-read data set, which allowed the identification of specific assembly errors. ***OB***

Modeling Ebola hemorrhagic fever in mice

Studies of Ebola hemorrhagic fever (EHF) pathogenesis have been hindered by the lack of suitable mouse models that recapitulate key aspects of the disease course in humans. Michael Katze and colleagues (*Science* doi:10.1126/science.1259595; 30 October 2014) have now used the Collaborative Cross, a genetically diverse panel of recombinant inbred mice, to identify lines of mice that develop a severe EHF-like pathology following infection with a mouse-adapted strain of Ebola virus (MA-EBOV). The authors infected 47 distinct mouse lines by intraperitoneal injection with MA-EBOV and observed a spectrum of phenotypes ranging from complete resistance to lethal infection with symptoms that included internal hemorrhage, prolonged blood coagulation and severe liver pathology. Further analyses showed that the susceptible lines were characterized by high levels of infectious virus in the spleen and liver, with widespread infection in hepatocytes. Conversely, the resistant lines exhibited low levels of infectious virus, with viral antigen in the liver restricted to endothelial cells and resident macrophages. Genetic mapping studies in these mouse lines will provide insights into the host factors influencing susceptibility to specific phenotypes following Ebola virus infection, which could aid understanding of human disease pathogenesis. *KV*

Cohesin and chromosome loops

Chromosomes form local structures such as gene loops, but the influence of these structures on transcriptional regulation is not fully understood. Now, Keji Zhao, Richard Young and colleagues report the identification of chromosome structures, called insulated neighborhoods, that regulate the expression of local genes in embryonic stem cells (ESCs) (*Cell* **159**, 374–387, 2014). The authors used cohesin chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), which maps DNA-DNA interactions at cohesin-occupied sites. Cohesin is known to associate with enhancer-promoter loops and with CTCF-bound regions. Accordingly, the authors found that 92% of interacting cohesin-occupied sites occurred at enhancers, promoters and CTCF-binding sites. They identified a structure consisting of a loop formed by two interacting CTCF sites co-occupied by cohesin that was associated with enhancers, in particular, super-enhancers that regulate the expression of key cell identity genes, and with repressed lineage-specifying genes that are marked by trimethylation of histone H3 at lysine 27 (H3K27me3). They showed that deletion of CTCF sites at five key ESC genes led to the altered expression of nearby genes, providing evidence that these structures have a functional role. The authors suggest that association of enhancers and target genes in insulated neighborhoods might have a role in preventing off-target effects. *EN*

Fine mapping with function

Bogdan Pasaniuc, Alkes Price, Peter Kraft and colleagues report a new integrative statistical method for the fine mapping of associations assisted by functional annotation, which is useful in prioritizing variants to include in following functional studies (*PLoS Genet.* **10**, e1004722, 2014). Their method, PAINTOR (Probabilistic Annotation INTegratOR), provides a framework for incorporating functional annotations with association statistics to assign a probability of causality for SNPs at a previously associated locus. The authors used an empirical Bayes prior to integrate functional annotation data and maximum-likelihood estimation to simultaneously estimate model parameters over all fine-mapping loci. Importantly, their method allows for multiple causal variants at any locus. The authors demonstrate in simulations that using posterior probabilities to prioritize variants provides greater accuracy than other fine-mapping methods in identifying causal variants, in part because the method allows for multiple causal variants. They apply their approach to a large meta-analysis for 4 lipid traits in combination with 450 cell type–specific annotations. They demonstrate that their approach improves the prioritization of causal variants, reducing the 90% confidence set from 17.5 to 13.5 SNPs per locus. The software is publicly available at http://bogdan.bioinformatics.ucla.edu/software/paintor. *OB*

Persimmon sex determination

Ryutaro Tao, Luca Comai and colleagues report the discovery of a small RNA–mediated sex determination pathway in the persimmon, *Diospyros lotus* (*Science* 346, 646–650, 2014). The authors sequenced the genomes of 32 female and 25 male plants and identified approximately 800 male-specific contigs, from which they inferred the male-specific Y chromosome region (MSY). Combining their map with RNA sequencing of developing male and female buds, they identified seven male-specific transcripts that mapped to the MSY. One candidate, a predicted homeodomain transcription factor they named *OGI*, meaning 'male tree' in Japanese, encoded a 21-bp small RNA and showed similarity to an autosome-derived female-specific transcript, which the authors named *MeGI* ('female tree'). Both genes are orthologs of the barley *Vrs1* gene, which is involved in flower development. Analysis of transgenic tobacco and *Arabidopsis thaliana* plants expressing *MeGI* and *OGI* suggested that *OGI* targets and inhibits expression of *MeGI*, thereby suppressing feminization. The results also indicated that *MeGI* might act to sterilize androecia in a dose-dependent manner. ***BL***

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