

Hypermutation in breast cancer

Whereas the mutational patterns of cancers related to specific carcinogens are known, the mutational spectra of many other cancers are not well understood. Now, Michael Stratton and colleagues report the sequencing of 21 breast cancer genomes and a sophisticated analysis investigating the mutational forces shaping these cancer genomes (*Cell* 149, 979–993, 2012). After cataloging somatic mutations, the authors considered the nucleotide context in which the mutations occurred. By systematically looking at the nucleotides both 5' and 3' to the mutated base pair, the authors found an overrepresentation of C>T substitutions at XpCpG trinucleotides. The authors attribute these mutations to the deamination of methylated cytosine to thymine. Other mutational contexts were overrepresented, although the underlying mechanisms are not known. The authors employed a mathematical approach called nonnegative matrix factorization (NMF) and modeled selection to extract mutational signatures from the 21 cancers, finding evidence for five biologically distinctive mutational signatures. They also calculated intermutational distances and observed regional hypermutation in some cancers, a phenomenon they call 'kataegis' (Greek for shower or thunderstorm). These mutations were present in *cis*, such that they occurred on the same DNA strand, perhaps simultaneously or in a sequential manner over a short time period. **PF**

Targeting cancer stem cells

The concept that some cancers are sustained by rare cancer stem cells (CSCs) has led to efforts to develop drugs that selectively target these CSCs. Previously, a human pluripotent stem cell (hPSC) line was developed that has characteristics of cancer stem cells, including enhanced self-renewal and low differentiation potential. Now, Mickie Bhatia and colleagues use this neoplastic hPSC line to screen several hundred known compounds for their ability to induce differentiation and overcome neoplastic self-renewal (*Cell*, published online 24 May 2012; doi:10.1016/j.cell.2012.03.049). The authors set out to identify compounds that would induce differentiation in neoplastic hPSCs but would have little effect on normal hPSCs. Of 590 compounds screened, only thioridazine significantly induced differentiation in neoplastic hPSCs but not in normal hPSCs. The authors tested 2,446 additional compounds, whereby they identified 26 candidates, including thioridazine and 2 other phenothiazine compounds. Of these three compounds, thioridazine still showed the strongest effect on neoplastic hPSCs. The authors conducted xenotransplantation studies *in vivo* and treated leukemic disease-initiating acute myeloid leukemia (AML) leukemic stem cells (LSCs) with thioridazine. This treatment reduced the leukemic disease potential of these cells. **PF**

STAT3 activation in rare T-cell leukemia

Large granular lymphocytic leukemia is a rare disorder marked by expansion of cytotoxic T lymphocytes, often accompanied by autoimmune disease manifestations. Satu Mustjoki and colleagues (*N. Engl. J. Med.* 366, 1905–1913, 2012) now show that this leukemia subtype frequently harbors somatic activating mutations in *STAT3*. The authors performed exome sequencing on leukemic and non-leukemic cells from a subject with typical T-cell large granular lymphocytic leukemia and

identified a missense mutation affecting the SH2 domain of *STAT3*. Targeted sequencing in 76 additional cases identified protein-altering *STAT3* mutations in 40% of subjects with large granular lymphocytic leukemia. The altered residues cluster at the dimerization interface of *STAT3* and are predicted to result in cytokine-independent signaling activity. Consistent with this, *STAT3* mutation-positive leukemic cells showed nuclear localization of *STAT3* and upregulation of *STAT3* transcriptional targets. Transfected cells expressing mutant *STAT3* proteins also had increased expression of a *STAT*-responsive reporter construct. These findings suggest that a substantial proportion of individuals with large granular lymphocytic leukemia could benefit from treatment with *STAT3* inhibitors, which are currently being tested in clinical trials for other hematological disorders. **KV**

Mapping 5-hydroxymethylcytosine

Peng Jin, Bing Ren, Chuan He and colleagues report a method called Tet-assisted bisulfite sequencing (TAB-seq) to map 5-hydroxymethylcytosine (5hmC) across the genome at single-base resolution (*Cell* 149, 1368–1380, 2012). The method can discriminate 5hmC from 5-methylcytosine (5mC), as 5hmC is selectively protected by β -glucosyltransferase, thereby rendering 5mC uniquely susceptible to TET-mediated oxidation to 5-carboxylcytosine. Bisulfite treatment then converts cytosine and 5-carboxylcytosine to uracil and 5-carboxyluracil, respectively, which are read as thymine during sequencing, whereas β -glucosyl-5-hydroxymethylcytosine is read as cytosine. The authors applied TAB-seq to mouse and human embryonic stem cell genomes and found that 5hmCs are present in clusters and that 5mC and 5hmC can coexist at the same cytosine in a population of cells. Unlike 5mC, 5hmC is asymmetric at CG sequences and shows a bias for G-rich sequences. The authors noted enrichment of 5hmC immediately adjacent to transcription factor-binding sites and also found that approximately half of 5hmCs are located in distal regulatory elements, such as p300-binding sites, enhancers, CTCF-binding sites and DNase I hypersensitive sites. **EN**

Rare variation in human populations

Michael Bamshad and colleagues report sequencing of 15,585 protein-coding genes to an average depth of 11 \times in 2,440 individuals of European and African ancestry (*Science*, published online 17 May 2012; doi:10.1126/science.1219240). They identify over 500,000 single-nucleotide variants (SNVs), the majority of which are rare, novel and population specific. The overall site frequency spectra (SFS) were highly skewed, showing a large excess of rare variants compared to the number expected with the standard neutral model. The authors fit a modified out-of-Africa demographic model that provides support for a recent and rapid acceleration in population growth. In characterizing functional variation using seven different computational methods, the authors found that ~47% of SNVs were predicted to be deleterious by at least one of these methods, with limited overlap in predictions between methods. The authors estimated a mean of 13,595 SNVs and 549 novel SNVs per individual, with both of these numbers higher in individuals of African ancestry relative to those of European ancestry. Further, they estimated an average of 318–580 predicted functional protein-coding SNVs per individual, depending on the definition of functional variant, and this estimate was higher in individuals of African ancestry than in those of European descent. **OB**

Written by Orli Bahcall, Pamela Feliciano, Emily Niemitz & Kyle Vogan