RNA splicing in common disease

Altered RNA splicing is implicated in many human diseases, although how common this phenomenon is and to what extent genetic variation contributes to it is not known. To address these questions, Brendan Frey and colleagues developed a machine learning algorithm that predicts the impact of single-nucleotide variants (SNVs) on RNA splicing and analyzed the association of variants with disease (*Science* doi:10.1126/science.1254806; 18 December 2014). Using this approach, they detected thousands of variants that might be involved in a myriad of diseases, including spinal muscular atrophy, colorectal cancer and autism spectrum disorder (ASD). Indeed, a genome-wide analysis of several individuals with ASD showed that, for the genes predicted to be misspliced, there was enrichment for transcripts that are highly expressed in brain tissue and that are associated with neurodevelopmental conditions. They also discovered that intronic mutations linked to disease are nine times more likely to affect splicing than common ones and that missense exonic mutations with little impact on protein function are five times more likely to perturb normal splicing. It will be interesting in the future to incorporate additional layers of splicing regulation into computational models, such as epigenetic modifications and chromatin remodeling. ***TF***

Microexons on the brain

Coregulated alternative splicing events involved in developmental and disease processes are largely uncharacterized. Manuel Irimia, Benjamin Blencowe and colleagues developed a computational pipeline to systematically identify all alternative splicing events, including the use of microexons, in RNA sequencing data to better understand which biological processes are affected by specific alternative splicing programs (*Cell* **159**, 1511–1523, 2014). The pipeline was applied to data from over 50 tissue and cell types in human and mouse and was designed to detect microexons—exons less than 27 nt in length—which are often missed in genome annotations. The authors found a group of ~2,500 neural-regulated splicing events and an enrichment of microexons among neural-specific alternative exons. Moreover, most microexon inclusion events were neural specific and highly conserved between humans and mice. Through analysis of publicly available RNA sequencing data, the authors found that the splicing factor nSR100 (SRRM4) was responsible for the regulation of most alternatively spliced microexons. They experimentally demonstrated that microexon regulation is switched on late during neural differentiation and that microexons likely function to enhance specific protein-protein interactions when spliced in. Finally, the authors found that 30% of alternatively spliced microexons were misregulated in the brains of some individuals with autism spectrum disorder (ASD), and the inclusion of neural-regulated microexons was correlated with *nSR100* expression levels across all individuals with ASD analyzed. *BL*

Zebrafish mutants versus morphants

The use of antisense morpholinos to perform targeted knockdowns in zebrafish embryos can be a powerful method to assess the developmental functions of specific genes of interest, but the usefulness of this approach is hampered by the potential for off-target effects. To explore this issue further, Nathan Lawson and colleagues (*Dev. Cell* doi:10.1016/j.devcel.2014.11.018; 18 December 2014) used genome editing techniques to introduce mutations into 24 zebrafish genes suspected to be involved in early vascular development. They found that mutations in only 3 of the 24 genes caused vascular phenotypes, whereas the other mutations produced no overt defects in vascular or overall morphology. Notably, previous morpholino-based studies had reported developmental phenotypes for more than half of these genes. The low concordance rate between morphant and mutant phenotypes observed in this study indicates that morpholinos may potentially produce very high rates of false positive phenotypes. Given the recent development of efficient methods for generating targeted mutations using site-specific endonucleases, the authors recommend that the generation and characterization of mutants be considered the standard approach for defining gene function in zebrafish. *KV*

AID mistargeting to enhancers

The activation-induced cytidine deaminase (AID) targets DNA mutations and breaks to immunoglobulin loci in B cells during the processes of somatic hypermutation and class switch recombination, but aberrant targeting of non-immunoglobulin loci can lead to oncogenic mutations. Now, James Bradner, Frederick Alt and colleagues (*Cell* **159**, 1538–1548, 2014) and Marei Dose, Rafael Casellas and colleagues (*Cell* **159**, 1524–1537, 2014) report new insights into the causes of mistargeting by AID. Bradner, Alt and colleagues mapped AID off-target sites in mouse B cells by using high-throughput genome-wide translocation sequencing to identify double-strand break hotspots and characterized the B cell transcriptome using global run-on sequencing. They found that AID targets are characterized by overlapping sense and antisense transcription, termed convergent transcription, at intragenic enhancers with high levels of acetylation at lysine 27 of histone H3 (H3K27ac). Dose, Casellas and colleagues mapped AID targets across the genome in mouse B cells, using replication protein A chromatin immunoprecipitation, and in human B cells and B cell tumors, using deep sequencing. They found that AID preferentially targets promoters and enhancers tethered by long-range interactions and transcribed enhancers with high chromatin accessibility. These concordant studies provide new insights into the genomic determinants of AID off-target mutagenic activity. *EN*

Genetic modules for autism

Autism spectrum disorder (ASD) is genetically heterogeneous, with most ASD-associated variants accounting for very few cases. Michael Snyder and colleagues used a systems biology approach to identify common pathways dysregulated in ASD (*Mol. Syst. Biol.* 10, 774, 2014). Highly interacting modules were extracted from a protein interaction network. Two modules were significantly enriched for known ASD susceptibility genes, and one contained multiple genes involved in synaptic transmission. Whole-exome or whole-genome sequencing of 25 ASD and 5 control brains, followed by comparison to the 1000 Genomes Project database, identified 38 genes in this module that were significantly affected by rare, coding variants in individuals with ASD, 28 of which had not been previously linked to ASD. Expression analysis of all genes in the module found that they clustered into two groups, one preferentially expressed in the corpus callosum and the other more ubiquitously expressed. The corpus callosum is involved in communication between the brain hemispheres. The authors generated RNA sequencing data from the corpus callosum of six individuals with ASD and matched controls and found that the expression levels of genes in the module were significantly altered in ASD cases. This association did not hold for synaptic genes or ASD-associated genes in general. ***BL***

Written by Tiago Faial, Brooke LaFlamme, Emily Niemitz & Kyle Vogan