Genomics

Transposable element evolution in mammals

Transposable elements (TEs) are thought to have contributed to the evolution of their host organisms. However, owing to the lack of suitable genomic resources and tools, TEs have been generally overlooked. Osmanski et al. aimed to remedy this trend by providing complete de novo TE annotations of 248 mammalian genome assemblies from the Zoonomia consortium. Overall, placental mammals share similar TE content and subtype proportions, with LINEs and SINEs representing the most abundant families. Nevertheless, the examination of younger insertions revealed substantial differences in TE accumulation from each category, with some lineages exhibiting no recent accumulation whereas others experiencing massive expansions in one or more groups. Young retrotransposons (particularly LINEs, LTRs and SINEs) positively correlate with genome size increase, whereas young DNA transposons are associated with smaller genomes. Interestingly, mammalian genomes accumulate individual dominating TE categories at any given period rather than multiple types simultaneously. Finally, carnivorous mammals tend to accumulate more recent DNA transposons than noncarnivores. However, no differences between herbivores and omnivores were detected. This in-depth TE curation represents an exceptional resource to deepen our knowledge of repetitive elements and their effect on genomic function and evolution.

Chiara Anania
Nature Genetics

Original reference: *Science* **380**, eabn1430 (2023)

Computational methods

Simulating scRNA-seq for benchmarks

The comparative benchmarking of computational methods uses both real-world and simulated data, each of which has respective strengths and weaknesses — the latter, for example, enables the generation of data with a defined ground truth to test performance in situations in which underlying model assumptions are not met. However, the history of genomics method development also demonstrates that these synthetic data do not fully recapitulate the complexity of real-world experiments. Crowell et al. comprehensively compared the outputs of 16 simulators across a wide variety of data, experimental designs and quality control and evaluation statistics. They discovered that most of these simulators are limited in applicability to real-world data, that performance is often over-stated and rankings are unreliable, and that the exact analysis task at hand should determine the statistics used to evaluate performance. These results emphasize that benchmarking is a nuanced task that requires careful consideration, and the authors' conclusion that "the most truthful model for real data is real data" underlines the importance of analyzing published data to demonstrate the capabilities of methods.

Michael Fletcher
Nature Genetics

Original reference: *Genome Biol.* **24**, 62 (2023)

Functional genomics

Single-cell CRISPR screen for GWAS loci

Most genetic variants associated with human complex diseases and traits that have been identified by genome-wide association studies (GWAS) lie within non-coding sequences. The effects of these variants on gene expression is often unclear, and it remains challenging to precisely pinpoint causal variants that are tagged by GWAS hits. Using an integrative multi-omic approach termed STING-seq, which includes CRISPR interference-based perturbation of GWAS-linked *cis*-regulatory elements and single-cell RNA sequencing analysis, Morris et al. identified 124 *cis*-target genes of 91 blood trait loci in human erythroid progenitors (K562 cell line). Predominantly, the identified target genes are the closest ones to the GWAS variant. Leveraging base editing, the authors validated the link between some noncoding variants and alterations in gene expression. In addition, they characterized *trans*-effect networks of noncoding loci for *cis*-target genes that encode transcription factors or microRNAs. They found an enrichment of GWAS variants in these networks, highlighting the polygenic architecture of complex traits. There is an urgent need to develop robust assays and analytical tools that can identify causal variants that regulate disease-relevant genes in a high-throughput manner. STING-seq will hopefully help to clarify the biological relevance of additional GWAS data.

Tiago Faial
Nature Genetics

Original reference: *Science* https://doi.org/10.1126/science.adh7699 (2023)