Regulatory genomics

Linking GWAS to gene regulation

Genome-wide association studies (GWASs) have shown that many traits are highly polygenic, and the general assumption is that trait-associated variants operate through gene-regulatory mechanisms such as expression quantitative trait loci (eQTLs). However, it remains frustratingly difficult to link genetic variants to their target, trait-causal genes. Connally et al. delved into this problem by analyzing a curated set of 220 putative causal genes that harbor both GWAS-identified common, non-coding and rare, coding genetic variation. Strikingly, eQTL colocalization, transcriptome-wide association, and regulatory annotations nominated a candidate target for <10% of analyzed non-coding variants. The authors concluded that there is a major portion of so-called "missing regulation" that may be due to eQTL context dependence, non-linear and/or non-homeostatic gene-expression effects, and the like. This study highlights the complexities of assigning target genes to GWAS-identified variants, and suggests that more-nuanced approaches beyond trusting in increasingly large data-generation efforts will be needed to definitively address this fundamental problem of genetics.

Michael Fletcher

Nature Genetics

Original reference: *eLife* **11**, e74970 (2022)

Cancer genetics

Sarcoma predisposition

Sarcomas are a group of rare cancers that have been relatively understudied compared with epithelial malignancies. To gain more insight into the etiology of these mesenchymal tumors, Ballinger et al. carried out whole-genome germline sequencing on 1,644 individual people, along with family members and matched control participants. Their analyses identified mutations in 14 new putative sarcoma-predisposing genes and, importantly, highlighted two pathways involved in mitosis and telomere integrity. Specifically, the list includes genes related to the centrosome complex (prevention of chromosome mis-segregation and aneuploidy) and to the shelterin complex (telomere protection). Notably, sarcomas are often aggressive cancers and tend to affect a younger population. This study is therefore a welcome contribution to this research field. It should clear the path toward better screening for sarcoma predisposition and will hopefully aid in the development of more-effective therapies.

Tiago Faial

Nature Genetics

Original reference: *Science* **379**, 253–260 (2022)

Epigenomics

Human cell-type methylomes

Human methylome studies across different cell types have been essential for understanding cell-type-specific methylation patterns and gene regulation. Loyfer et al. performed deep whole-genome bisulfite sequencing at an average depth of 30× on 39 human cell-type groups sorted from 205 healthy tissue samples. These methylomes showed large similarities between replicates of the same cell type, which suggests that cell-type-specific programs rather than environmental factors are the main determinants of DNA methylation. The authors identified methylation blocks of CpG sites that reflect human developmental lineage of cell types. They also characterized cell-type-specific unmethylated genomic regions and identified the top 250 unmethylated markers for each cell type. Fragment-level analysis led to the identification of cell-type-specific enhancer regions and regulatory transcription factors. Cell-type-specific hypermethylated regions were found to be enriched for CpG islands, Polycomb targets and CTCF binding sites. Using the atlas, interestingly, the authors detected a high concentration of endothelial-cell-derived cell-free DNA in patients with COVID-19. The study provides a valuable human methylome atlas of normal cell types, as well as cell-type-specific biomarkers with potential clinical applications.

Wei Li

Nature Genetics

Original reference: *Nature* **613**, 355–364 (2023)

Disease epidemiology

Linking viruses with neurodegeneration

Following on recent reports linking prior infection with Epstein–Barr virus to a risk of developing multiple sclerosis, Levine et al. leveraged available data from two large population-based studies, FinnGen and UK Biobank, to explore associations between a range of viral exposures and risk of neurodegenerative diseases. In FinnGen, they identified 45 significant associations between viral exposures and neurodegenerative diseases, 22 of which were replicated in UK Biobank. The strongest hazard ratio was observed for the association between viral encephalitis and Alzheimer's disease. Notably, no protective effects were seen for any viral exposures. The authors further analyzed longitudinal data available in each biobank to explore the relationship between timing of viral exposure and disease diagnosis. Although hazard ratios were found to be highest within 1 year of infection, effects of viral exposure on neurodegenerative disease risk persisted several years after initial infection. The authors also found that risks of viral infection following a diagnosis of neurodegenerative disease were generally lower than risks of developing a neurodegenerative disease after viral infection. Collectively, these epidemiological studies suggest widespread effects of viral exposure on neurodegenerative disease risk and highlight potential strategies for the prevention and treatment of these diseases.

Kyle Vogan

Nature Genetics

Original reference: *Neuron* https://doi.org/10.1016/j.neuron.2022.12.029 (2023)