Cancer genetics

Pan-cancer atlas of intratumour heterogeneity

Transcriptional differences between tumours have been extensively studied using bulk RNA sequencing. More recently, single-cell RNA-sequencing (scRNA-seq) has led to the discovery of marked transcriptional heterogeneity between malignant cells of the same tumour that has important implications in terms of prognosis and response to therapy. Individual scRNA-seq studies of intratumour heterogeneity (ITH) have so far been limited by sample size; reporting in *Nature*, Gavish et al. integrate data from 77 scRNA-seq studies to define a comprehensive pan-cancer atlas that characterizes 11 'hallmarks' of transcriptional ITH.

The authors incorporated published scRNA-seq data from human tumours, unpublished data from neuroendocrine tumours, head and neck cancer and schwannoma, and selected datasets from mouse models or cell lines, in total covering 1,456 samples from 24 cancer types. For each dataset, 38 cell types were annotated on the basis of canonical marker expression, and malignant cells were distinguished from non-malignant cells on the basis of inferred copy-number alterations.

Based on previous scRNA-seq studies showing that sets of genes (ITH programmes) have coordinated variability within a tumour, the authors set out to identify related ITH programmes from different tumours. For each tumour, they used non-negative matrix factorization (NMF) to characterize ITH programmes of malignant cells as defined by their 50 top-scoring genes. 'Robust' ITH programmes that were consistently identified in individual tumours using multiple parameter values were grouped across tumours into 41 clusters according to shared genes. 83% of clusters were derived from several cancer types and the clusters covered 66% of all ITH programmes, which indicates that most ITH can be described by recurrent patterns common to different types of tumour. For each cluster, a meta-programme (MP) of ITH was defined as the most commonly shared genes in that cluster. MPs were given a functional annotation based on those genes, and MPs with related functions were grouped into 11 'hallmarks' of transcriptional ITH. Twenty-one of the 41 MPs (mainly those of low frequency) had not been detected previously in smaller scRNA-seq studies, indicating the increased sensitivity of the combined dataset. Using bulk RNA-seq data from The Cancer Genome Atlas (TCGA), the authors identified associations between expression of the 41 MPs and tumour metastasis, therapy resistance and overall survival, some of which were consistent across cancer types, which indicates the functional relevance of the MP classification.

Seven of the MPs were found at medium or high frequency in most cancer types — including the family of cell-cycle-associated MPs — and these were denoted as 'general' MPs. Most other MPs were negatively correlated with cell-cycle MPs, suggesting that highly proliferative tumour cells may repress other gene expression programmes.

Thirteen of the MPs were found in only one or two cancer types and were denoted as 'context-specific' MPs. For example, MP38 (glutathione genes) and MP39 (metal response genes) were identified as being variable only in kidney clear cell carcinoma, despite these genes being expressed by other cancer types.

The remaining 21 'shared' MPs were detected in 3–12 cancer types. For example, MP30, which contains genes characteristic of pancreatic ductal adenocarcinoma (PDAC), was shown to be variable also in lung, colorectal, liver and head and neck cancers. Further analysis of data from TCGA showed that a subset of lung adenocarcinoma classified as invasive mucinous adenocarcinoma has high expression of MP30 and of other PDAC-enriched genes. Both MP30 and PDAC gene signatures contain genes associated with mucin production, which suggests that MP30 represents an aberrant mucin production programme shared by several cancer types.

Gavish et al. extended their MP analysis to non-malignant cells from the same original dataset. They show that most of the malignant MPs resemble non-malignant MPs of epithelial cells, which suggests that much of the transcriptional heterogeneity of tumours is already present in the healthy tissue. Where differences exist between malignant and non-malignant MPs, they may provide information about the coupling or uncoupling of related pathways in cancer. Furthermore, correlations between MPs of different cell types can be used to infer interactions between cells within a tumour, for example inferring a network of cells associated with a particular function such as angiogenesis or leukocyte migration, a shared response of several cell types to a specific factor such as interferon, or direct interactions between cell types — such as correlation between a cytotoxic T cell response and the proteasome MP of malignant cells (which is required for degradation of antigen and presentation to T cells).

> "most [intratumour heterogeneity] can be described by recurrent patterns common to different types of tumour"

The analysis of select gene expression programmes described by the authors highlights the enormous value of this combined dataset — available through the Curated Cancer Cell Atlas — for further discovery of the clinical and therapeutic relevance of patterns of ITH.

Kirsty Minton

Original article: Gavish, A. et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* https://doi.org/10.1038/s41586-023-06130-4 (2023)

CREDIT: JON HICKS/STONE/GETTY