The chromatin of cancer

Taipale, Jussi

2018-10-26


http://hdl.handle.net/10138/273448
https://doi.org/10.1126/science.aav3494

Downloaded from Helda, University of Helsinki institutional repository.
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
The chromatin of cancer

Understanding tumorigenesis and inherited risk for cancer requires a multiomic approach

By Jussi Taipale

Developments in modern genomics tools have led to rapid progress in our understanding of the genetic basis of cancer. Recent large-scale efforts have primarily focused on two types of analysis: mapping acquired somatic mutations by whole-exome and whole-genome sequencing (1, 2), and identification of common inherited variants that increase cancer risk using genome-wide association studies (GWAS) (3). Despite the power of these technologies, we are still far from understanding how the variants and mutations found in individual tumors precisely drive the oncogenic process. A large number of genetic variants increase risk for cancer, but most explain only a very small fraction of the risk. Furthermore, although acquired somatic mutations are found in almost all tumors, most do not carry complete sets of mutations that, according to our present mechanistic understanding, would be sufficient to cause cancer. On page 420 of this issue, Corces et al. (4) show how a third type of genomics approach—functional genomic analyses of primary human tumors—can begin to bridge this gap in our mechanistic understanding of the tumorigenic process.

The authors analyzed chromatin accessibility using ATAC-seq (assay for transposase-accessible chromatin using sequencing) of 410 primary tumors representing 23 different types of human cancer. Analysis of chromatin accessibility measures stable binding of proteins to the genome; regions that are unbound are accessible to enzymes such as deoxyribonuclease I (DNase I) (5) or Tn5 transposase (4). The ATAC-seq method used by Corces et al. utilizes Tn5, which inserts a linker sequence to accessible DNA and cuts it, allowing highly efficient isolation and sequencing of the liberated fragments. Most of the human genome is relatively inaccessible because it is wound around histone proteins, forming nucleosomes, each of which contains 147 base pairs of DNA. In less than 1% of the genome, the histones are replaced by other proteins that regulate chromosome structure, or that function as transcription factors to direct gene expression. Tn5 can insert DNA linkers between such proteins; if the proteins are bound tightly, their binding position also leaves a “footprint” that is narrower than that formed by a nucleosome. DNA accessibility is known to correlate with the presence of active gene regulatory elements such as promoters and enhancers, and is thus commonly used as a proxy for gene regulation. Motif mining of the accessible regions and analysis of sequences under the footprints can then be used to infer which sequence-specific DNA binding proteins are bound to the accessible regions. The power of the approach of Corces et al. derives from the combination of deep sequencing that allows footprinting with the analysis of a large number of samples representing different types of cancer. Importantly, the samples used are sequenced for mutation mapping in The Cancer Genome Atlas (TCGA) project, facilitating comparative multiomic analyses between different data types.

The motif mining and footprinting analyses reveal many transcription factors that are strongly active in the different cancer types. For example, the authors detect androgen receptor in prostate cancer and microphthalmia-associated transcription factor (MITF) in melanoma, indicating that ATAC-seq can pinpoint known cancer type–specific transcription factors. The identification of accessible chromatin across multiple cancer types, together with detection of expressed genes by RNA sequencing, allows inference of DNA elements that may regulate gene expression (5). This analysis is based on correlation, but the authors also validate a subset of the potential enhancer-promoter links by targeting a repressor to the regulatory elements using CRISPR-Cas9 interference. Compared to RNA sequencing, the analysis of accessible chromatin also gives a more detailed “fingerprint” of the tumor, facilitating classification of tumors and analysis of their cellular composition. The chromatin accessibility data can also be used to locate elements that contain variants that may contribute to inherited cancer risk, and to identify somatic noncoding mutations that affect chromatin accessibility. Given the scale of the dataset and its multiomic character, there is great potential for new discoveries. Most of the individual findings reported by Corces et al. need further validation. However, the large number of interesting initial discoveries, such as the link between elements near the MECT1 gene and adverse outcome in kidney cancer, highlights the value of the dataset as a reference and as a data-mining resource for future studies.

The analysis of chromatin accessibility in primary tumor cells also extends the known repertoire of potential gene regulatory elements. Of those identified by Corces et al., 35% were not previously known. Many of the elements identified from primary tumors are likely to be important for normal developmental and homeostatic processes. However, there is good reason to suspect that at least some may not be so benign. Our genome is likely to encode a large number of potentially pathological transcription factor–DNA interactions (6, 7). This is because cancer-causing mutations can directly affect transcription factor binding sites, leading to activation of normally silent regulatory elements (8). Mutations can also activate transcription factors...
Traditionally, there has been a disconnect between cancer genomics and large-scale efforts to map the functional genome. The Roadmap Epigenomics (10) and Genotype-Tissue Expression (GTEX) (11) projects primarily focus on normal tissues, whereas the main drive of ENCODE (5) is to identify functional genomic elements; although cancer cells are used as models in some of these projects, the cell lines used do not adequately represent major forms of human cancer. Previous epigenomic studies of cancer, in turn, have mainly focused on targeted DNA methylation analysis (12), transcription factor binding analyses in a few cell lines (13), or analysis of histone modifications in a particular type of cancer (14). In this context, the study by Corces et al. is particularly welcome because it paves the way toward a large-scale effort to map the functional genome of cancer cells. To understand how individual tumors form, it is necessary to map their genomic features such as germline variants, somatic mutations, chromosomal content, and allelic imbalance (15), together with functional genomic features such as genes that are essential for growth and survival, three-dimensional (3D) chromosome conformation, the DNA methylome, chromatin modification state, and accessible chromatin landscape (see the figure). Comparing cancer types to each other can yield interesting results but suffers from the disadvantage that all cancers share key phenotypic characteristics, such as unrestricted growth. A better comparison would be between cancers and their cell types of origin. However, the cell type of origin of many cancers is unknown, and many tumors are thought to originate from relatively rare cells (for example, stem or progenitor cells). Therefore, it will also be necessary to develop analytical methods that can detect genomic features from minor cell populations or from single cells. Without such multimodal maps at the cell-type level, it will be exceedingly difficult to move from genomics toward understanding the main drivers of the phenotype of individual tumors. Without such understanding, we may not be able to conquer cancer.

REFERENCES

**REFERENCES**


10.1126/science.aav3494
The chromatin of cancer
Jussi Taipale

Science 362 (6413), 401-402.
DOI: 10.1126/science.aav3494