The chromatin of cancer

Understanding tumorigenesis and inherited risk for cancer requires a multiomic approach
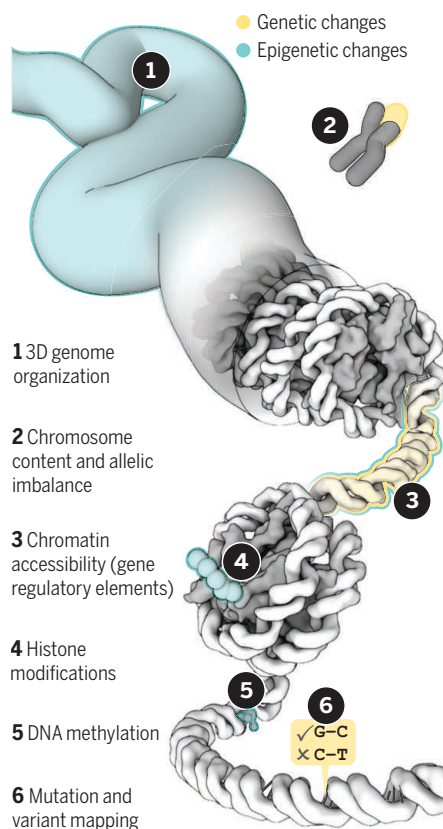
By **Jussi Taipale**[1,2,3]

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,

1Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK. 2Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. 3Genome-Scale Biology Program, University of Helsinki, Helsinki, Finland. Email: ajt208@cam.ac.uk

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

Sequence and function of the cancer genome

Multiomic mapping and comparison between genetic and epigenetic features are required for mechanistic understanding of cancer and for providing a "fingerprint" of the tumor. Multiomic analyses are likely to be important for cancer diagnosis and prediction of outcome, as well as for guiding treatment decisions and drug development.

● Genetic changes
● Epigenetic changes

1 3D genome organization

2 Chromosome content and allelic imbalance

3 Chromatin accessibility (gene regulatory elements)

4 Histone modifications

5 DNA methylation

6 Mutation and variant mapping

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 tissue, 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 *MECOM* 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

that have low or no activity in the corresponding normal tissues, leading to activation of large sets of gene regulatory elements (9). In addition, many driver mutations affect chromatin modifiers or alter levels of CpG methylation of DNA, leading to destabilization of the entire chromatin regulatory system (1, 2). It is difficult for evolutionary processes to remove large sets of potentially harmful elements from our genome, as each individual element has limited impact at the population level, and cancer generally affects individuals who are above reproductive age. Therefore, it is likely that elements that are specifically activated in cancer are present in our genome. Identification of such elements will facilitate improved diagnosis and prognosis, and also allow investigations of new therapeutic modalities to target oncogenic gene regulation.

The mapping of accessible chromatin landscapes is also important for the mechanistic understanding of tumorigenesis. It is known that altered activities of transcription factors and/or their binding regions drive the major forms of human cancer. Cancer can thus be considered a disease of gene expression, where a combination of mutations locks the gene regulatory network of a single cell into a state that drives unrestricted cell proliferation (6). Although mutations in some oncogenes and tumor-suppressor genes are commonly found across many forms of cancer, most driver genes are mutated in a more restricted set of tumors. Some of the differences in oncogene composition can be explained by differences in mutational mechanisms and proto-oncogene expression between the cell types of origin of the tumors. However, many oncogenes cannot transform fibroblastic cells in standard cell-based assays, suggesting that cell lineage–determining transcription factors collaborate in some way with oncogenes. The mechanisms of such collaboration are currently poorly understood, but given that lineage-determining factors commonly define chromatin states, it is likely that accessibility of chromatin at specific regulatory sites contributes to this process. An important contribution of the study by Corces et al. is the identification of candidate sets of such lineage-specific regulatory elements that are critical for the cancer phenotype.

Cancer genome sequencing efforts have revealed that a large number of genes can cause cancer. Because most of the driver genes are mutated infrequently, making mechanistic sense of the cancer genotype by straightforward genetic interaction analysis requires extremely large sample sizes. Combining genomic data with phenotypic information is thus an attractive alternative. 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 multiomic 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

1. L. A. Garraway, E. S. Lander, Cell 153, 17 (2013).
2. B. Vogelstein et al., Science 339, 1546 (2013).
3. A. Sud et al., Nat. Rev. Cancer 17, 692 (2017).
4. M. R. Corces et al., Science 362, eaav1898 (2018).
5. R. E. Thurman et al., Nature 489, 75 (2012).
6. I. Sur, J. Taipale, Nat. Rev. Cancer 16, 483 (2016).
7. J. Taipale, EMBO J. 37, e96114 (2018).
8. M. R. Mansour et al., Science 346, 1373 (2014).
9. K. Gangwal et al., Proc. Natl. Acad. Sci. U.S.A. 105, 10149 (2008).
10. P. Polak et al., Nature 518, 360 (2015).
11. M. Melé et al., Science 348, 660 (2015).
12. K. A. Hoadley et al., Cell 173, 291 (2018).
13. J. Yan et al., Cell 154, 801 (2013).
14. B. Akhtar-Zaidi et al., Science 336, 736 (2012).
15. K. Palin et al., Nat. Commun. 9, 3664 (2018).

EVOLUTION

Our shallow-water origins
Coastal habitats represent a cradle of diversification for early vertebrates

By Catalina Pimiento[1,2]

Vertebrates encompass all animals with a backbone, from fish to humans. How and when they evolved are questions that have been studied for centuries, revealing the origins and processes involved in anatomical innovations such as jaws, teeth, and paired appendages (1). A less explored, but equally important question is where they evolved. On page 460 of this issue, Sallan et al. (2) compile a new database of early occurrences (mid-Paleozoic, 490 to 360 million years ago) and site-specific environmental information to reconstruct vertebrate ancestral habitats. They report that all major vertebrate clades originated in restricted, shallow-water environments.

The environmental context of vertebrate evolution had remained a gap in our knowledge. Understanding the habitat constraints

> "What was so special about the shallow-water habitats where vertebrates diversified?"

present when key traits evolved is necessary to answering fundamental questions in macroevolution such as the extent to which the environment can drive anatomical transformations. Reaching this level of understanding has been limited mostly by a lack of data in available compendia (3). The examination of primary data on early fish (e.g., from the mid-Paleozoic) revealed that their fossil record accumulated in shallow waters (4). However, it has been recognized that this might be an artifact of a poor fossil record; in other words, the habitats from where ancient fish have been recovered might reflect outcrop (the exposure of rocks)

1Department of Biosciences, Swansea University, Swansea, UK. 2Smithsonian Tropical Research Institute, Panama. Email: cpimiento@swansea.ac.uk

Science

The chromatin of cancer

Jussi Taipale

ARTICLE TOOLS	http://science.sciencemag.org/content/362/6413/401
RELATED CONTENT	http://science.sciencemag.org/content/sci/362/6413/eaav1898.full http://stm.sciencemag.org/content/scitransmed/4/156/156ra140.full http://stm.sciencemag.org/content/scitransmed/5/209/209ra153.full http://stm.sciencemag.org/content/scitransmed/7/283/283ra54.full http://stm.sciencemag.org/content/scitransmed/2/38/38ra47.full
REFERENCES	This article cites 15 articles, 7 of which you can access for free http://science.sciencemag.org/content/362/6413/401#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions