

# Sequencing the secrets of the cancer genome

The advent of next-generation sequencing (NGS) in the first decade of the twenty-first century heralded a major change in cancer research. Similarly to how electronic miniaturization transformed computers from room-sized appliances into smartphones that fit in the palm of a hand, the progress from low-throughput Sanger sequencing to high-throughput NGS allowed an entire genome to be sequenced by an individual research group, in contrast to the globe-spanning industrial effort required by the Human Genome Project only a few years earlier. By that time, cancer's status as a genetic disease was well established, and catalogues of oncogenes and tumour suppressors were available. However, the number of mutations in any given tumour, as well as the full extent of the 'cancer-gene census', remained unknown; these and other questions could finally be answered with the newly developed sequencing technologies.

The first cancer whole-genome sequence, of an acute myeloid leukaemia (AML) from a woman in her mid-50s, was published in 2008 by Ley et al. This karyotypically normal, highly pure case was carefully selected to ensure adequate input material for sequencing and to facilitate interpretation of the results. Two samples from this patient were sequenced with Solexa technology on an Illumina Genome Analyzer: primary tumour and normal skin tissue.

Single-nucleotide variant (SNV) analysis demonstrated the importance of including the matched normal sample: 3.8 million

SNVs were identified in the tumour sample, of which 2.6 million also occurred in the skin and were thus categorized as germline variants. After further filtering, a total of eight SNVs were validated as novel somatic variants with a predicted effect on gene function. These eight genes had roles in functionally relevant pathways—cellular signalling (*PTPRT*, *KNDCl*, *ADGRA1*, *GPR183* and *GCOM2*), cell adhesion (*CDH24* and *CDHR2*) and transmembrane transport (*SLC15A1*)—thus demonstrating that genome sequencing can identify novel candidate cancer genes within key, potentially therapeutically targetable tumorigenic pathways. Most intriguingly, none of these eight somatic SNVs were observed in any of a cohort of 187 AML tumours, thus suggesting substantial heterogeneity in genes affected by mutation in a tumour type.

Beyond identifying a set of novel candidate cancer genes, this study attempted to track cancer evolution by comparing variant allele frequencies in amplicon sequencing data for ten somatic mutations (the eight SNVs, an *FLT3* internal tandem duplication and an *NPM1* insertion) in primary and post-relapse samples. This analysis suggested that nine of ten mutations were heterozygous in all tumour cells, but the last mutation—the *FLT3* internal tandem duplication—appeared to be subclonal and was suggested to be the most recent mutation.

In the following year, three additional cancer genomes from a metastatic breast tumour, and from lung and

metastatic melanoma cell lines were published. These were distinctly different from the AML results in terms of the number of somatic mutations observed: 32 in breast cancer, and 33,345 and 22,910 in the melanoma and lung cell lines, respectively. Notably, the mutations identified in the breast tumour were largely mutually exclusive from those found in an extended series of 192 tumours, as observed for the AML genome. The mutational signatures of exposure to tobacco smoke (C>A) and ultraviolet light (C>T) were detected in the lung and melanoma genomes, respectively.

Together, these four studies suggested that cancers have substantial genetic heterogeneity, both within and between individual tumour entities. Most importantly, these findings suggested a clear path for future research: whole-genome sequencing, and related techniques that profile transcriptomes and epigenomes, would need to be applied at scale to gain a better understanding of cancer's complex genetic basis and to provide insights into how these genetic discoveries might be clinically applied.

Large-scale studies such as The Cancer Genome Atlas (TCGA) and the Pan-Cancer Analysis of Whole Genomes (PCAWG) have now sequenced tens of thousands of cancer genomes across a wide variety of tumour types. The insights gained from these large cohorts include the identification of many more cancer genes associated with previously unanticipated cellular processes such as metabolism (*IDH1* and *IDH2*) and epigenetic regulation (*EZH2* and *PBRM1*), and the confirmation that most somatic mutations in tumours are not highly recurrent (the 'long tail' of mutation frequency). They have ignited the study of cancer evolution (MILESTONE 11), based on phylogenies of somatic mutations inferred through sequencing, which has shed light on the complex dynamics of tumourigenesis. Finally, the analysis of mutational signatures has identified many previously known and novel mutagenic processes, as well as the genomic signatures of prior exposures. Although the past decade has unequivocally demonstrated the value of genome sequencing in basic cancer research, clinical use remains restricted to well-resourced institutions equipped with the expertise to generate and interpret genomic data, thus suggesting that, at least for patients, the greatest impact of sequencing may be yet to come.

Michael Fletcher, *Nature Genetics*

**ORIGINAL ARTICLE** Ley, T. et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456, 66–72 (2008).

**FURTHER READING** Please visit the [online article](#) for a full list of further reading.



Credit: zmeel