

A sequencing revolution in cancer

By the early 2000s, the first draft of the human genome had been published (MILESTONE 1), and a small number in the cancer research community quickly recognized the promise of sequencing to transform research across the translational pipeline. The technique proved its clinical potential in 2004, when three publications showed that *EGFR* mutations were associated with sensitivity to *EGFR* inhibitors, demonstrating that sequencing cancer genes could guide precision medicine strategies. Others recognized that genomics — still a fledgling field — could be harnessed to provide answers to some fundamental questions about cancer, such as what is the repertoire of mutations in a tumour?

In 2006, Sjöblom et al. reported results of Sanger-based sequencing of 13,023 genes in 11 breast and 11 colorectal cancers. A main finding was that the repertoire of cancer-associated genes was much larger and more diverse than had been expected and that the mutational spectra between different tissues were surprisingly distinct. This landmark analysis showed that interrogating the cancer genome could identify disease-associated mutations. But it used a labour-intensive, targeted sequencing approach, which was likely to miss important coding and non-coding variants. The field was ripe for an unbiased, genome-wide approach.

Reporting in *Nature* in 2008, Ley et al. presented the first whole-genome sequence of a cytogenetically normal acute myeloid leukaemia (AML) sample using the Solexa/Illumina platform. Two samples taken from a woman aged in her 50s were sequenced; a tumour sample (sequenced at >30-fold coverage) and a normal skin sample. The normal sample allowed the team to filter out somatic mutations from germline ones, a crucial step towards distilling the genetic lesions that were driving the disease. A total of



Credit: Sabena Jane Blackbird / Alamy Stock Photo

2,647,695 single-nucleotide variants (SNVs) were discovered, of which 2,584,418 were also present in the skin genome. Further filtering for variants occurring in the coding sequences of annotated genes whittled the list down to eight novel non-synonymous mutations and two previously reported ones. The novel mutations occurred in genes with unknown roles in cancer, and their biological functions provided intriguing insight into the pathways disrupted in AML, including small-molecule and drug transport, cell–cell interactions, self-renewal and cell signalling. All mutations, except for a mutation in *FLT3*, were present in every tumour cell at both initial presentation and relapse, suggesting that the disease was driven by a single dominant clone and that the *FLT3* mutation had occurred most recently. Of note, sequencing in an independent cohort of 187 AML samples revealed no recurrent somatic mutations in these genes, providing the first hints that driver mutations are not always indigenous to a tumour type.

The ground-breaking work by Ley et al. confirmed what many had suspected: unbiased, genome-wide sequencing can unmask novel genes that likely contribute to tumorigenesis and offers a palette of potentially druggable targets.

“ the first whole-genome sequence of a cytogenetically normal acute myeloid leukaemia ”

And thus began a sequencing revolution in cancer. A parallel explosion in bioinformatics allowed scientists to make sense of swathes of sequencing data and identify an increasingly complex catalogue of mutations, from SNVs to complex structural rearrangements. The genomic landscapes of different cancer types revealed what Ley et al. had intimated: no two cancer genomes are alike.

In 2009, sequencing of the whole genome and transcriptome of a metastatic breast cancer sample offered insights into the evolution of the cancer genome during disease progression. Through its genome, a tumour’s history was laid bare, allowing researchers to identify not only the repertoire of driver mutations in a tumour but also the order in which they occurred.

The field continues to advance at an astonishing pace. As sequencing costs plummeted, small groups coagulated into large, international consortia, reporting on integrated, pan-cancer analyses of thousands of tumours. Sequencing-based assays can now identify disease-specific drivers, mutational signatures, tumour mutational burden and neo-antigens, offering tremendous promise to guide personalized patient care. Ever more powerful analytical tools are allowing us to interrogate these genetic features and infer their biological significance. While the fruits of these innovations are yet to dominate the treatment landscape, it is clear that the best is yet to come.

Safia Danovi,
Nature Genetics

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

FURTHER READING Lynch, T. J. et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004) | Paez, J. G. et al. *EGFR* mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004) | Pao, W. et al. *EGF* receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl Acad. Sci. USA* **101**, 13306–13311 (2004) | Sjöblom, T. et al. The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006) | Campbell, P. J. et al. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020) | Shah, S. P. et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009)