The cancer genome and the birth of precision medicine

John F. DiPersio

Division of Oncology, Siteman Cancer Center/Washington University School of Medicine in St Louis, St Louis, MO, USA

E-mail: jdipersi@wustl.edu

https://doi.org/10.3324/haematol.2025.288784

TITLE	DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome.
AUTHORS	Ley TJ, Mardis ER, Ding L, *et al.*
JOURNAL	Nature. 2008:456(7218);66-71. doi.10.1038/nature07485.

Sequencing of the first cancer genome by Ley and colleagues[1] stood on the shoulders of the magnificent work of giants in the field of human genetics and their first draft sequence of the human genome.[2,3] This work has led to the birth of precision medicine and the use of modern genomics and epigenomics to understand mutations associated with the generation of cancer cells (in the case of Ley *et al.*, acute myelogenous leukemia [AML]) but has also provided the platform for using these cancer-associated mutations to predict progression-free survival, resistance to therapy and monitoring of minimal residual disease. They also led to the discovery of clonal hematopoiesis and to the identification of potential targets for effective therapeutic interventions. The public international Human Genome Project[2] and the private biotechnology Celera project[3] set the stage for the transformative study by Ley *et al.* proving that unbiased whole genome sequencing of a single patient's cancer genome could be performed and decoded, transforming medicine and cancer care forever.

Understanding the etiology and factors associated with the initiation and progression of hematologic and solid tumor malignancies has been a quest of many over the decades and had been largely limited by our understanding of the genetics or genomics of cancer. This has radically changed over the last 50-75 years with the discovery of chromosomes as the "packages" carrying the encrypted dictionary of heredity, the double helix as the structure of this package and the invention of recombinant DNA technologies allowing for cloning, genetic manipulation and function of specific genes controlling both heredity and the dysfunction and mutation of these genes resulting in cancer. The overwhelming effort by many in the field was to use these emerging DNA sequencing technologies to generate the draft sequence of the first human genome in an effort to generate a framework and comparator for future efforts to understand how this sequence was altered contributing to many diseases including cancer.

The work of the Human Genome Project began in 1990 and consisted of a collaborative effort of over 20 institutions in six countries led by MIT/Broad, Washington University and the Sanger Center. The work was completed in 2000 with the first draft sequence of the human genome (~90% of the human genome), enhanced in 2003 (93%) and finished in October 2022 by the Telomere-to Telomere consortium (T2T).[4] The final cost was over $3 billion.[2] In 1998 Celera was formed with the goal of applying an alternative whole genome sequencing approach called hierarchical shotgun sequencing *versus* the whole genome shotgun sequencing used in the Human Genome Project.

There were many who thought attempts at sequencing a single patient's cancer genome would not be feasible or even interpretable. Ley and colleagues proved that this was incorrect. The first cancer genome sequenced was from a 50-year-old female patient with AML. It was a monumental feat that cost $2 million and took only 2 years and was one of the foundational events giving birth to the era of modern precision medicine. The first challenge was to compare the ~3 billion base pairs of the sequences of tumor DNA and germline (skin) DNA. Coverage was 32.7-fold for the tumor DNA and 13.9-fold for the normal skin DNA from the same patient. This was an essential visionary insight allowing Ley and his colleagues to distinguish between mutations and single nucleotide variants (SNV) in the tumor DNA. Of the 2,647,695 SNV found in the tumor genome, 2,584,418 (97.6%) were also detected in the patient's germline skin genome. Without the sequencing of matched tumor-germline DNA from the same patient the identification of mutations *versus* SNV would not have been possible. What was incredibly surprising was that after subtracting germline SNV only eight non-synonymous somatic SNV were identified in this first cancer (AML

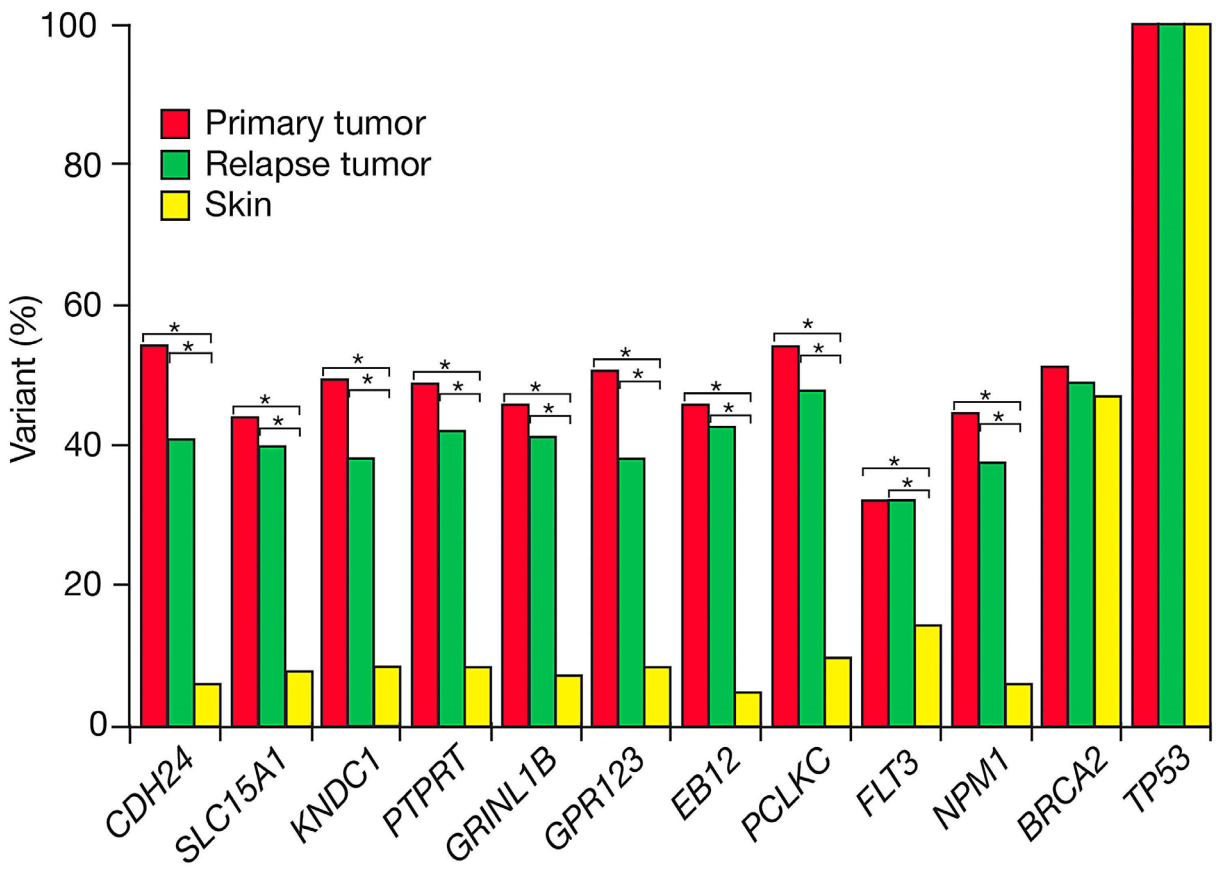

Figure 1. Summary of Roche/454 FLX readcount data obtained for ten somatic mutations and two validated single nucleotide polymorphisms in the primary tumor, relapse tumor and skin specimens. The readcount data for the variant alleles in the primary tumor sample and relapse tumor sample are statistically different from those of the skin sample for all mutations (P<0.000001 for all mutations, Fisher's exact test, denoted by a single asterisk in all cases). Note that the normal skin sample was contaminated with leukemic cells containing the somatic mutations. The patient's white blood cell count was 105,000/μL (85% blasts) when the skin punch biopsy was obtained. Figure reproduced from Ley TJ *et al.*[1]

genome). There were homozygous germline variants of *TP53* and a heterozygous germline variant of *BRCA2*. Aside from heterozygous missense mutations in two known AML genes (*NPM1* and *FLT3*) most of the somatic mutations identified were in a small number of genes that had no known association with AML and were assumed to be innocent bystander mutations that were present in the cell that initiated the malignancy and were swept forward in this "founder clone". Relapse of this patient generated a very limited number of new mutations after both standard chemotherapy and after allogeneic stem cell transplant (Figure 1). Additional efforts in subsequent studies did identify new genes clearly associated with AML such as *DNMT3A*, *TET2* and *AXL1* (DTA mutations) as well as *IDH1* and *IDH2* against which clinically effective small molecule therapies have been developed and approved by the Food and Drug Administration.

The success with the unbiased sequencing of the first cancer genome set the stage for an assault on sequencing the genomes of multiple cancers. This effort has resulted in the generation of important targeted therapies and immunotherapies for those diseases with mutation-dense genomes such as melanoma and smoker's non-small cell lung cancer. These studies helped to define genes associated with poor outcomes, resistance to conventional and novel therapies and genes that could be tracked before and after treatment to predict progression-free survival, early relapse or those patients who might require early and more aggressive therapies such as stem cell transplant. Some mutations have been seen to accumulate during aging (clonal hematopoiesis of indeterminate potential [CHiP]) and may predict increased risk of not only myelodysplastic syndromes and AML, but also other inflammatory diseases such as cardiovascular disease, liver disease and chronic renal diseases. Tracking these mutations before and after therapy not only has provided accurate predictors of relapse and survival but also has allowed us to better understand how different therapies can shape the clonal architecture and progression through critical bottlenecks that therapies exert on these cancers.

The unbiased sequencing of the first human cancer genome has opened the field up to critical insights into the causes of these diseases, the development of intelligent therapeutic interventions, myriads of commercial platforms tracking cancer by next-generation sequencing, and assessment of cell-free tumor DNA, all resulting in the insertion of precision medicine into the everyday life of all in the worlds of medicine and biomedical sciences.

Disclosures

No conflicts of interest to disclose.

References

1. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008;456(7218):66-71.
2. Landes ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
3. Venter CJ, Adams MD, Myers EW, et al. Shotgun sequencing of the human genome. Science. 2001;291(5507):1304-1351.
4. Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. Science. 2022;376(6588):44-53.