Whole genome sequencing in acute myeloid leukemia

Torsten Haferlach

MLL Munich Leukemia Laboratory, Munich, Germany

E-mail: torsten.haferlach@mll.com

https://doi.org/10.3324/haematol.2026.289418

TITLE	Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia.
AUTHORS	Cancer Genome Atlas Research Network; Ley TJ, Miller C, Ding L, *et al.*
JOURNAL	The New England Journal of Medicine. 2013;368(22):2059-2074. doi: 10.1056/NEJMoa1301689.

When this study by The Cancer Genome Atlas (TCGA) Research Network on adult *de novo* acute myeloid leukemia (AML) appeared in 2013, it rewrote the model of the disease.[1] It showed that AML is not a genetic jungle, but a relatively "low-mutation" cancer whose complexity lies in which lesions co-occur, in which clones, and with which epigenetic consequences.

The authors sequenced 200 adults with *de novo* AML using whole-genome sequencing or whole-exome sequencing, combined with RNA and microRNA analyses, single nucleotide polymorphism arrays, and DNA methylation profiling. On average, each leukemia carried 13 coding mutations, of which five affected recurrently mutated genes, far fewer than in many solid tumors. Yet they identified 23 significantly mutated genes and another 237 genes mutated in at least two samples, revealing AML as combinatorially complex despite its modest mutation number. Almost every case harbored at least one non-synonymous mutation in one of nine functional categories – transcription-factor fusions, *NPM1*, tumor suppressors, DNA-methylation genes, activated signaling, chromatin modifiers, myeloid transcription factors, cohesin, or spliceosome components – turning a list of "interesting genes" into an organized map of leukemogenic pathways.

A particularly elegant element of the paper was the move from single mutations to rules of cooperation and mutual exclusivity. Transcription-factor fusions such as *PML*::*RARA*, *RUNX1*::*RUNX1T1*, or *CBFB*::*MYH11* were largely mutually exclusive with *NPM1* and *DNMT3A*, implying that these lesions can substitute as founding events. In contrast, *NPM1, DNMT3A*, and *FLT3* frequently clustered together and defined a molecularly coherent subgroup with distinctive RNA, microRNA, and methylation signatures, suggesting a true biological entity rather than a coincidental combination. Variant-allele-frequency clustering in the 50 cases studied by whole-genome sequencing showed that more than half of samples contained a founding clone and at least one subclone, making clonal architecture a part of AML biology (Figure 1).

Thus, the paper laid conceptual groundwork for genetically defined entities that later entered World Health Organization classifications and for risk models beyond "intermediate cytogenetics with or without *FLT3*." It also framed epigenetic regulation as central: *IDH1/2*-mutant cases showed broad gains in DNA methylation, while cases with *NPM1, DNMT3A,* or *FLT3* and some *KMT2A*-rearranged cases exhibited extensive hypomethylation, particularly in CpG-sparse regions. This helped to legitimize epigenetic therapies and pushed the field toward integrating methylation and expression information.

Today, the logic of that study still drives our ambitions, but we are no longer satisfied with static, diagnosis-only snapshots or genome-biased views. The interesting action often unfolds in non-coding regions, three-dimensional chromatin structure, small RNA, and clonal evolution under therapy, dimensions that only longitudinal whole-genome sequencing will fully capture.

As each patient's leukemia now generates millions of data points across molecular layers, human pattern recognition alone is no longer sufficient. This is where the path opened by the paper naturally extends into whole-genome sequencing and artificial intelligence assisted interpretation. Artificial intelligence models trained on TCGA-like multidimensional cohorts at scale can now discover and refine patterns of cooperation, exclusivity, and epigenetic consequence, linked to clinical trajectories. The paper gave us the first rigorous rules of the game. Our task

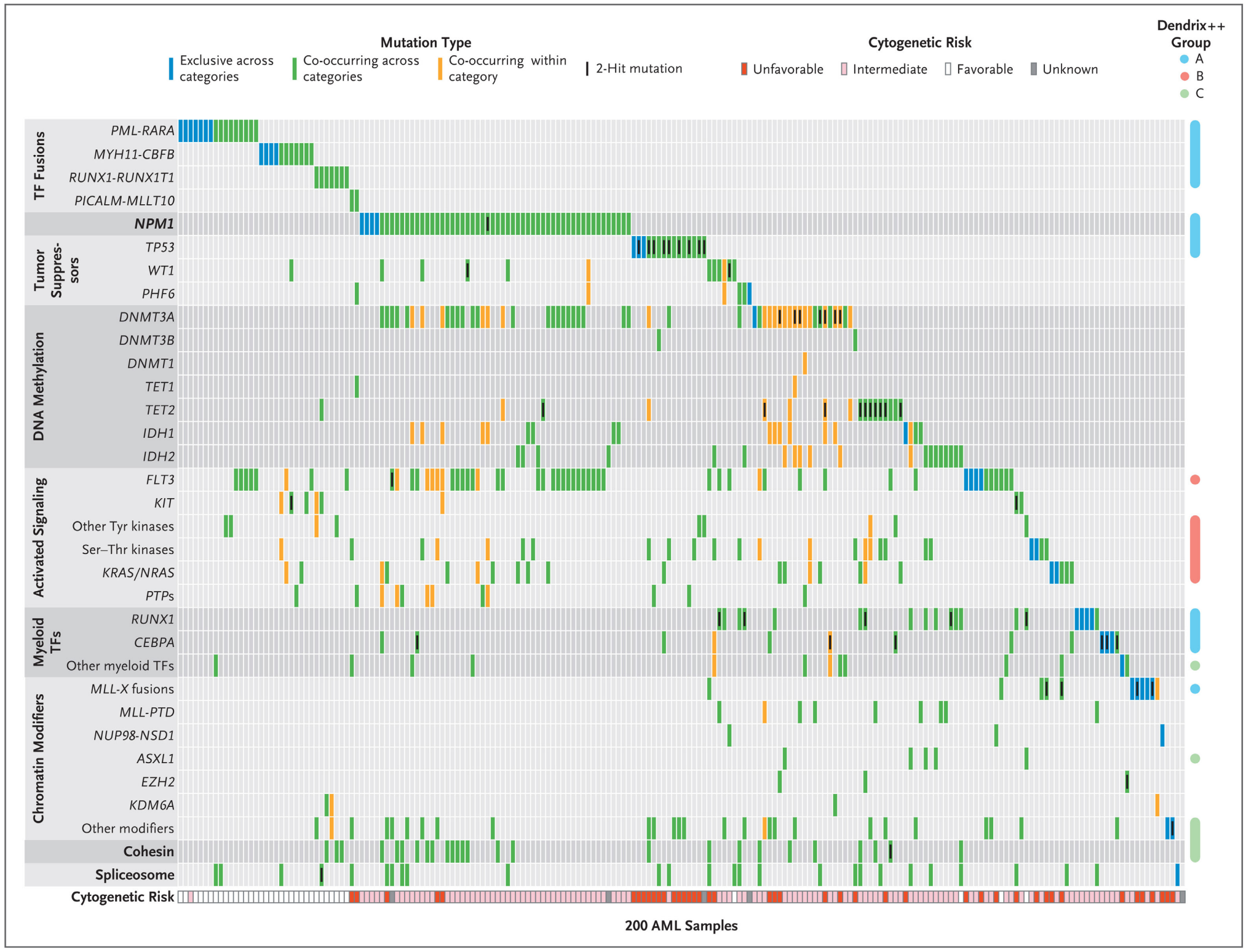

Figure 1. Organization of mutations into categories of related genes. The figure shows somatic, non-synonymous mutations in individual genes and sets of genes, grouped into nine categories, including one single-gene category, as labeled on the left. Of the 200 samples evaluated, 199 (>99%) had at least one mutation in one of the listed genes or sets. Blue boxes indicate mutations that are exclusive across all categories; green boxes, mutations that co-occur in the same sample across different categories; and orange boxes, mutations that co-occur in the same sample in the same category. Computational analysis with the use of the Dendrix++ algorithm identified three significant, mutually exclusive groups of genes, annotated on the right as groups A, B, and C. The cytogenetic risk for each patient is shown at the bottom of the chart. Additional information about data in this figure is provided in Tables S17 through S20 in the Supplementary Appendix of the original paper.[1] Ser–Thr: serine–threonine; TF: transcription factor; Tyr: tyrosine. Figure reproduced with permission from the paper by the Cancer Genome Atlas Research Network, *et al*. N Engl J Med. 2013;368(22):2059-2074.

now is to let whole-genome sequencing and artificial intelligence help us apply them in patients' care with the resolution and speed this disease demands.

Disclosures

TH is part owner of MLL Munich Leukemia Laboratory.

References

1. Cancer Genome Atlas Research Network; Ley TJ, Miller C, Ding L, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059-2074.