

How to cite this article:

Sartori FA, Pontes JV, Campregher PV, LaFramboise T. Mutational signatures in hematological malignancies. *einstein* (São Paulo). 2026;24:eRW1961.

Associate Editor:

Kenneth Gollob
Hospital Israelita Albert Einstein, São Paulo, SP, Brazil
ORCID: <https://orcid.org/0000-0003-4184-3867>

Corresponding Author:

Thomas LaFramboise
10900 Euclid Avenue Cleveland
Zip code: 44106-7078 - Cleveland, Ohio, United States
Phone: (1) 216 773 1079
E-mail: txl80@case.edu

Received on:

June 24, 2025

Accepted on:

Sep 10, 2025

Copyright the authors



This content is licensed under a Creative Commons Attribution 4.0 International License.

REVIEW

Mutational signatures in hematological malignancies

Felipe de Almeida Sartori^{1,2}, João Vitor Paes Pontes², Paulo Vidal Campregher^{2,3}, Thomas LaFramboise¹

¹ Case Western Reserve University, Cleveland, Ohio, United States.

² Hospital Israelita Albert Einstein, São Paulo, SP, Brazil.

³ Genesis Genomics, São Paulo, SP, Brazil.

DOI: [10.31744/einstein_journal/2026RW1961](https://doi.org/10.31744/einstein_journal/2026RW1961)

ABSTRACT

The set of somatic mutations present in a human tumor is a record of one or more mutational processes, each of which leaves distinct “signature” of mutation types. Mutation types can be classified in various ways, the most straightforward being the base change induced by a single-base substitution (e.g., C>A, T>G, etc.). The advent of high-throughput DNA sequencing has facilitated the comprehensive, genome-wide assessment of mutation types in human tumors. This has spurred the development of methodology to tease apart the relative contribution of each mutational process by decomposing the set of all mutations into individual signatures. Many mutational signatures have known etiologies. Therefore, mutational signature inference can shed light on the causes of cancer and inform patient treatment. To date, most studies in this area have been performed on solid tumors; consequently, the application of existing methods to hematological cancers has yielded limited results. In this review, we provide an overview of the history and methodology behind mutational signature inference. Here, we present the challenges inherent in its application to hematological cancers and survey the work performed thus far. We highlight how recent research analyzing mutational signatures in normal blood cells can elucidate the beginning of a continuum of mutational processes, from normal hematopoiesis through mature hematological malignancy. Accurate characterization of mutational signatures in cancer development may aid in clinical diagnosis, prognosis, and treatment decisions.

Keywords: Mutational signatures; Hematological malignancy

INTRODUCTION

Since the advent of high-throughput sequencing technologies, our understanding of the genetic underpinnings of human diseases has dramatically expanded.⁽¹⁾ Among the conditions most impacted by these advances is cancer, which is fundamentally driven by the accumulation of somatic mutations.⁽²⁾ The ability to decode entire genomes has propelled research, providing unprecedented insights into the molecular basis of tumorigenesis and opening avenues for personalized medicine.⁽³⁾

A particularly intriguing discovery, dating back to the early 2000s, was the observation that mutations tend to occur in distinct patterns that are often associated with specific exposures. For instance, tumors arising in sun-exposed skin commonly exhibit C>T and CC>TT transitions, reflecting DNA damage induced by ultraviolet (UV) light.⁽⁴⁾ These recurring mutation patterns, now known as mutational signatures, suggest that a tumor’s mutational landscape carries a historical imprint of the biological processes and external agents that shaped its evolution.^(3,5,6)

Although groundbreaking, early studies of mutational patterns have faced several limitations.⁽⁵⁻⁷⁾ Most focused exclusively on driver genes, which are under selective pressure, potentially obscuring the full spectrum of mutations. Moreover, these studies typically analyzed a limited number of mutations per sample and lacked robust computational tools to disentangle overlapping mutational processes. Consequently, these studies have only provided a partial view of the mutational processes operating in cancer genomes.

With the rise of whole-genome sequencing (WGS), it has become possible to catalog nearly all somatic mutations in a given tumor.⁽³⁾ This technological leap facilitated a more comprehensive approach to mutational analysis. A pioneering study by Nik-Zainal et al. applied a mathematical framework based on non-negative matrix factorization (NMF) to identify mutational signatures from large-scale sequencing data of 21 breast cancer genomes.⁽⁶⁾ Subsequent work by Alexandrov et al. refined and publicly released the model.⁽⁷⁾ By incorporating all mutations within each tumor sample and analyzing patterns across multiple samples, this method revealed distinct mutational signatures, some of which could be linked to specific biological mechanisms or exposures - such as UV light.

Initially, these efforts focused on single base substitutions (SBSs), categorized using a 96-mutation framework that considers the trinucleotide context of each mutation (the mutated nucleotide plus its immediate 5' and 3' neighbors).⁽⁵⁾ This allowed researchers to decompose complex mutational spectra into interpretable components. Over time, the scope of analysis broadened to include additional mutation types such as doublet base substitutions,⁽⁸⁾ small insertions and deletions,⁽⁸⁾ copy number alterations,^(9, 10) structural variants,⁽¹¹⁾ and even RNA-based mutations.⁽¹²⁾

As the field matured, the catalogue of known mutational signatures expanded significantly. New computational tools tailored to different data types and research goals have emerged, enabling more precise and personalized analyses.⁽¹³⁻²³⁾ Recent discoveries have further elucidated the links between certain signatures and their underlying etiologies.⁽⁸⁾ To support ongoing research and clinical use, the COSMIC (Catalogue of Somatic Mutations in Cancer) database⁽²⁴⁾ now maintains a curated and evolving collection of mutational signatures, currently in version 3.4, serving as a key resource for the scientific and medical communities.

This review provides an overview of the development and current state of mutational signature research. We describe the methodological advances that have enabled the creation of analytical tools for signature extraction,

compare existing approaches, and explore the applications of mutational signatures in both research and clinical contexts. Finally, we discuss the current state of mutational signature research in hematological malignancies and its potential in this domain.

EXISTING TOOLS TO EXTRACT MUTATIONAL SIGNATURES

A mutational signature, assumed to be caused by a single biological process or exogenous exposure, is a pattern of mutation types specified by the percentage of each mutation type. The simplest mutation types are SBSs. Considering the redundancy of the two complementary DNA strands, there are six possible substitutions: C>A, C>G, C>T, T>A, T>C, and T>G. As flanking nucleotides can also influence mutagenic processes, SBS signatures also consider the bases on the 3' and 5' sides of the mutation site, therefore yielding $4 \times 6 \times 4 = 96$ possible types of SBS mutations. Since the first tool was reported in 2012, tools for mutational signature extraction have undergone significant developments. In addition to refinements of NMF, alternative mathematical approaches have been introduced.^(16,18-21)

The original tool developed by Alexandrov et al., SigProfiler, has since been replaced by SigProfilerExtractor⁽¹⁵⁾ and SigProfilerAssignment,⁽¹³⁾ both of which were developed by the same laboratory but designed for distinct purposes. SigProfilerExtractor is intended for *de novo* extraction, whereas SigProfilerAssignment focused on signature fitting.

De novo extraction refers to the process of uncovering mutational signatures directly from a dataset (typically WGS data), without using previously-reported signatures.⁽²⁵⁾ This approach allows researchers to reveal both novel and known signatures. Computational comparisons, such as cosine similarity, are commonly employed to determine whether extracted signatures correspond to known signatures.

In contrast, the fitting strategy uses a predefined reference signatures, typically in the COSMIC database, to estimate the contribution of each known signature to the mutational profiles of individual samples. The COSMIC database⁽²⁴⁾ catalogs well over 100 human mutational signatures. Therefore, it is feasible to determine the composition of mutational signatures acting on a tumor by mathematically inferring the proportion of each known signature using the tumor's counts of each type of mutation as input data.

In addition to the contributions of Alexandrov et al., several other research groups have developed tools with various features and purposes. The main differences

among these tools include the type of extraction (*de novo* versus fitting), input format (usually VCF files or mutational count matrices), mathematical approach, and intended application (e.g., some are optimized for specific cancers, such as mmsig for hematologic malignancies⁽²²⁾).

Other tools available for signature extraction include PLDA,⁽¹⁹⁾ MCSM,⁽¹⁷⁾ SigTracer,⁽²⁰⁾ MutationalPatterns,⁽¹⁸⁾ YAPSA (Yet Another Package for Signature Analysis),⁽¹⁴⁾ mmsig,⁽²²⁾ MUSE-XAE⁽²¹⁾ and MuSiCal.⁽¹⁶⁾ A comparative description of these tools and their characteristics is provided in table 1.

Given the proliferation of tools and methods, two essential questions emerge: (i) What is the optimal strategy for extracting mutational signatures? (ii) Which tools are most effective? The literature reflects substantial variation in approaches, often leading to different results, even when applied to similar datasets.⁽¹⁶⁾ Two key studies have provided insight into these questions. Maura et al. proposed a structured workflow for signature extraction⁽²⁵⁾ with a focus on hematological malignancies, while Medo et al. conducted a comparative evaluation of different tools.⁽²⁶⁾

According to Maura et al., three main challenges complicate signature extraction:

- **Nonuniqueness:** Different combinations of signatures may explain the same mutational profile, especially for flat signatures that lack distinct peaks across trinucleotide contexts.

- **Localized processes:** Some mutational processes are region-specific and may be diluted when analyzing the genome as a whole.
- **Signature bleeding:** Signatures are erroneously assigned to samples in which they are not biologically present. This often results from pipelines that assume uniform exposure across heterogeneous cohorts.

To address these challenges, Maura et al. recommended a three-step workflow.

1. Begin with *de novo* extraction to identify cohort-specific signatures,
2. Map these to known references (e.g., COSMIC) using similarity metrics,
3. Perform fitting using the identified signatures as a reference to quantify their contribution in each sample.

They also suggested applying this workflow to specific genomic regions as a strategy to address the issue of localized mutational processes. By focusing on regions known to be differentially affected in the cohort of interest, it becomes possible to detect signatures that would be missed in a whole-genome approach.

Medo et al.⁽²⁶⁾ found that tool performance varies depending on several factors, such as mutation burden per sample, type of downstream analysis, and cancer type. Their results indicated that while SigProfilerSingleSample (now discontinued) performed

Table 1. Description of a selection of tools designed for mutational signature extraction

Tool name	Data types	Mathematical approach	Extraction type	Special trait
MCSM	WGS	MMM	Fitting	Evaluate strand or genomic region bias on genomic processes
Mix	Gene panel	MMM	Fitting	Use with clinical sequencing
mmsig	WGS	EM algorithm	Fitting	Built for hematologic malignancies
MUSE-XAE	WGS	Nonlinear encoder and linear decoder	<i>De novo</i> and fitting	-
MuSiCal	WES, WGS (preferably)	mvNMF and likelihood-based sparse NNLS	<i>De novo</i> and fitting	-
MutationalPatterns	WGS	NMF and NNLS	<i>De novo</i> and Fitting	Understand regional mutation spectra and strand asymmetry
Palimpsest	WES, WGS	NMF	<i>De novo</i> and Fitting	-
PLDA	WGS	PLDA	Fitting	Allowing the use of the same reference for all cancer types
SATS	Gene panel	pNMF and EM	<i>De novo</i> and fitting	Use with clinical sequencing
SigMA	Gene panel	NMF and NNLS	<i>De novo</i> and fitting	Mutational signature associated with HR deficiency
SigProfilerAssignment	WES, WGS (preferably)	NNLS	Fitting	-
SigProfilerExtractor	WES, WGS (preferably)	NMF and NNLS	<i>De novo</i> and fitting	-
SigTracer	WGS	Hierarchical Bayesian approach	Fitting	Analyze intra-tumor heterogeneity
YAPSA	WES, WGS (preferably)	NNLS	Fitting	-

EM: expectation-maximization; HR: homologous recombination; MCSM: mutation-level covariate signature model; MMM: multinomial mixture model; mvNMF: minimum-volume NMF; NMF: nonnegative matrix factorization; NNLS: nonnegative least squares; PLDA: parallelized latent Dirichlet allocation; pNMF: poisson NMF; WES: whole-exome sequencing; WGS: whole-genome sequencing; YAPSA: yet another package for signature analysis.

well in low-mutation samples, SigProfilerAssignment and MuSiCal were better-suited for high-mutation datasets.

Additionally, significant progress is being made in extracting mutational signatures from sequencing methods other than WGS, such as whole-exome sequencing (WES)⁽¹⁵⁾ and targeted panels.⁽²⁷⁻³⁰⁾ Given the clinical predominance of WES and targeted approaches, these developments are critical for translating mutational signature research into practical healthcare applications. Moreover, the increased depth of coverage achievable by these methods can uncover low-frequency mutations that may be missed by WGS.⁽³¹⁾ Tools capable of performing signature extraction in such contexts are listed in table 1, along with those designed for WGS data.

■ MATHEMATICS UNDERLYING THE MUTATIONAL SIGNATURE ANALYSIS

In their simplest form, algorithms to infer molecular signatures *de novo* take as input a $96 \times M$ matrix A , whose columns represent the M tumors sequenced, and whose rows represent all possible SBS mutation types. In its row i , column j entry, the input matrix A has the number of mutations of type i in tumor j .

The algorithm has two goals:

- 1) Infer the composition (*i.e.*, prevalence of each mutation type) of each individual mutational signature present in the sample set.
- 2) Infer what proportion each mutational signature contributes to each tumor's molecular landscape.

The main mathematical concept underlying this algorithm is NMF.⁽³²⁾ Briefly (for technical details, see ⁽⁷⁾), the aim is to represent A as the product of two matrices W and H , that is $A \approx W * H$ where $*$ denotes matrix multiplication. W is a $96 \times k$ matrix whose rows represent the types of mutations and whose columns represent k distinct mutational signatures. W addresses Goal 1 above, as its columns give the prevalence of each mutational type in each of the k signatures. H is a $k \times M$ matrix that addresses Goal 2 by providing the relative contribution of each signature to an individual tumor in each column. The factorization $A \approx W * H$ (*i.e.* determining the matrices W and H) is typically performed using an iterative procedure that aims to find the entries for W and H that minimize the difference between A and $W * H$, computed as $\|A - W * H\|_F$, where $\|X\|_F$ represents the Frobenius norm of X , defined as

$$\|X\|_F = \sqrt{\sum_{i=1}^N \sum_{j=1}^M x_{ij}^2}$$

for any $N \times M$ matrix X with i, j entry x_{ij} .

Note that the factorization $A = W \times H$ as described here, assumes knowledge of the number k of distinct mutational signatures present in the set of tumors. In practice, k is unknown *a priori*. To determine the optimal value for k , the minimization procedure described above is typically run multiple times for each of several values of k . Each value of k is assessed for reproducibility and stability across the multiple runs. Also considered is the error between each tumor's actual mutation pattern, as catalogued in matrix A , and that approximated by $W \times H$. The value of k that is most reproducible, stable, and best mitigates the error rate is selected as the optimal value for k .

The methodology for extracting signatures using mutations other than SBS (DBS, ID, CN, SV, and RNA-based) is similar. There are also multiple alternative variants of the *de novo* extraction procedure; however, most are conceptually similar to the approach described here.

Unlike *de novo* extraction, fitting uses known and specific mutational signatures. Suppose there are n known SBS mutational signatures under consideration. They can be represented by a $96 \times n$ matrix S , the columns of which provide the proportion of each mutation type in the corresponding signature. The counts of different single-base substitutions in a tumor may be represented as vector v with 96 entries, each entry corresponding to the number of mutations of one of the 96 types described above. Determining the optimal assignment of relative mutational signature contributions to the tumor mathematically corresponds to determining the vector a of length n that minimizes the difference between v and $S \times a$. Here, the 96 entries in a give the estimated relative contributions of each mutational signature to the mutational spectrum of the tumor. The difference between v and $S \times a$ is computed as $\|v - S \times a\|$, where $\|x\|$ denotes the length of vector x , defined as

$$\|x\| = \sqrt{\sum_{i=1}^{96} x_i^2}$$

for a vector of length 96.

The naïve determination of the optimal value of a assumes that all n mutational signatures may be present, leading to overfitting. One approach to avoid overfitting

is to iteratively remove and add signatures and determine how the difference $||v - S \times a||$ is affected. The specific steps of this “forward stepwise” algorithm are outside the scope of this review, but details of one implementation may be found in Díaz-Gay et al.⁽¹³⁾ In the end, some subset of the n mutational signatures are omitted from consideration, which results in a more reasonable number of inferred signature contributions.

CLINICAL RELEVANCE OF MUTATIONAL SIGNATURE INFERENCE

Inferring mutational signature composition from tumor DNA sequence can reveal the estimated number of mutations caused by each contributing signature. Aside from basic scientific insights, what more practical value might mutational signature inference have, particularly in the context of myeloid malignancies?

Some of the mutational signatures curated by COSMIC have known or suspected etiology. The initial and most well-established of these are signatures that were found in the first whole cancer genomes to be sequenced. These signatures are associated with smoking (particularly in lung cancer)⁽³³⁾ and UV light (particularly in skin cancer).⁽³⁴⁾ The primary cause of a tumor with mutation patterns dominated by one of these signatures would presumably be the corresponding exogenous exposure. Subsequently, several other exogenous causes of mutational signatures were identified. Treatments (for cancer and otherwise) associated with specific mutational signatures include temozolomide, platinum chemotherapy, azathioprine, thiopurine, duocarmycin, and melphalan.⁽³⁵⁾ Some of these signatures are particularly pronounced in post-treatment tumors with underlying mismatch repair (MMR) deficiency. Therefore, treatment of MMR-deficient patients may actually induce mutations that drive relapse or resistance to the treatment itself. For instance, acute lymphoid leukemia (ALL) patients treated with thiopurine chemotherapy were shown to acquire known relapse-associated mutations in the genes *NR3C1*, *TP53*, and *NT5C2*.⁽³⁶⁾

MMR-deficiency signatures are among the many that arise endogenously, whether from specific gene mutations or by other means. Endogenously-derived signature etiologies include 5-methylcytosine deaminase and apolipoprotein B editing complex (APOBEC) activity, polymerase eta errors, base excision repair deficiency, reactive oxygen species (ROS) damage, and homologous recombination (HR) deficiency, among

others. Clustering based on some of these mutational signatures in multiple myeloma yielded groups with prognostic value.⁽³⁷⁾ Therefore, specific compositions of mutational signatures have the potential to inform clinical treatment and surveillance decisions.

It should also be noted that the presence of certain mutational signatures may indicate greater sensitivity to specific cancer therapeutics. The most well-established example is that the presence of HR-deficiency signatures is closely associated with sensitivity to both platinum-based chemotherapy and poly(ADP)-ribose polymerase (PARP) inhibitors in breast cancer. Other examples with potential clinical applications include APOBEC signatures and their association with sensitivity to ataxia telangiectasia and Rad3-related kinase (ATR) inhibition,⁽³⁸⁾ ROS damage signatures associated with *MYCN* amplification⁽³⁹⁾ and potentially with sensitivity to electron transport chain complex I inhibitors.⁽⁴⁰⁾ The potential link between ROS damage and therapeutic sensitivity may be particularly relevant in acute myeloid leukemia (AML), as McLeod et al. found a ROS damage mutational signature in a substantial proportion of pediatric and adult AML samples.⁽⁴¹⁾

ANALYTIC APPROACHES FOR MUTATIONAL SIGNATURE INFERENCE IN HEMATOLOGICAL MALIGNANCIES

The vast majority of work in cancer mutational signature analysis have been conducted in solid tumors.^(6,42-46) Indeed, there are inherent challenges in the accurate calling of mutational signatures in hematological cancers. One issue is that the mutation rate tends to be much lower than that in solid tumors (Figure 1). This is particularly true of myeloid malignancies. For example, AML typically has some 100-fold fewer mutations than lung cancer and melanoma. Even calling somatic mutations is more difficult in blood cancers than in solid tumors.⁽⁴⁷⁾ To distinguish between somatic and germline variants in tumor DNA, researchers often rely on matched normal DNA from the same patient. Blood is a convenient source of normal DNA in patients with solid malignancies; however, it is unsuitable for hematological malignancies. Buccal swabs, fingernails, hair, saliva, and skin can serve as alternative sources of normal DNA. However, all of these sources have drawbacks, such as inadequate amounts of DNA or contamination with malignant blood cells. Some researchers culture skin fibroblasts; however, this is not a scalable solution.⁽⁴⁸⁾

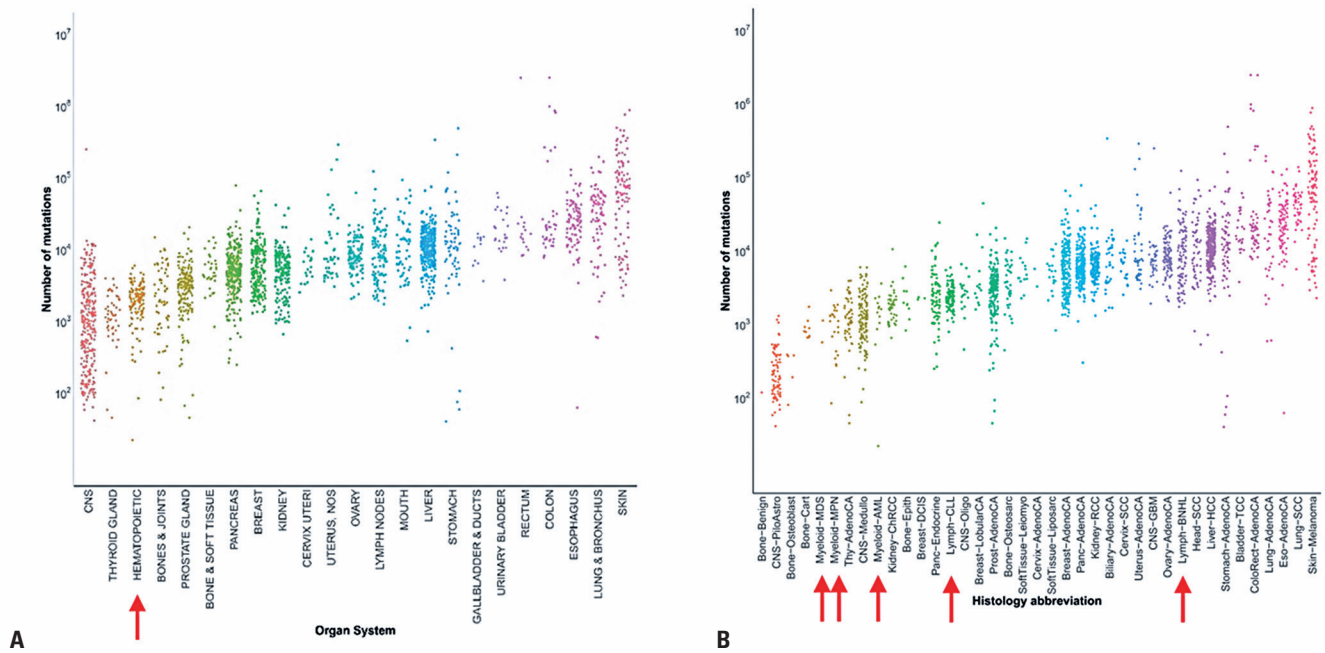


Figure 1. Somatic mutation rates tend to be lower in hematological malignancies. (A) Total numbers of mutations for each tumor in the Pan-Cancer Analysis of Whole Genomes, grouped by tissue type. Hematological tumors are indicated with red arrow. (B) Same as panel (A), but here, tumors are grouped by specific cancer type. Red arrows indicate hematological cancers. Note that all hematological cancers are among the less frequently mutated types, save B-cell non-Hodgkin lymphoma (BNHL)

Another challenge is that *de novo* signature extraction requires a large patient cohort.⁽²⁵⁾ Hematological cancers are generally much rarer than many solid tumor types, which limits the number of available whole-genome sequences. In existing literature, the dominant signatures in blood cancers are “clock-like” signatures that are closely linked to aging.⁽²²⁾ As a result, lower-abundance signatures can be difficult to detect accurately.

Despite these challenges, several groups have developed strategies to identify mutational signatures in hematological malignancies using standard tools. Researchers at the Memorial Sloan Kettering Cancer Center developed a platform, termed MSK-IMPACT Heme⁽⁴⁷⁾ which sequences 400 targeted genes that are known to play roles in hematological cancers. The use of a targeted panel greatly reduces costs compared to whole-genome sequencing and makes the sequencing of matched normal DNA economically feasible. In their study, the authors used either nail clippings or saliva as germline DNA sources. Owing to the limited number of mutations that can be detected by targeted sequencing, SigProfiler was applied to only the 11% of tumors with the highest mutational burden, all but two of which were lymphoid malignancies. In addition to age-related signatures, a variety of other signatures were detected,

largely corresponding to the type of treatment that the patient received prior to sequencing.

Diamond et al.⁽³⁵⁾ obtained similar results using whole-genome and whole-exome sequencing of myeloid malignancies, also applying SigProfiler. They exclusively found age-related signatures in treatment-naïve AML and patients undergoing radiation, but found treatment-related signatures in patients undergoing either high-dose melphalan- or platinum-containing therapies. In both cases, the study used standard methods to infer mutational signatures without tailoring them to hematological malignancies.

Hoang et al.⁽³⁷⁾ used the Palimpsest method,⁽⁴⁹⁾ primarily developed for solid tumors, to infer mutational signatures from whole-genome and whole-exome sequencing of multiple myeloma samples. The authors identified multiple “flat” signatures, including one associated with homologous recombination repair deficiency, which was subsequently criticized as being a false positive finding.⁽²²⁾ As mentioned previously, flat signatures are difficult to distinguish from one another.

In response to some of these challenges, Rustad et al.⁽²²⁾ developed a method called *mmsig*, which was specifically designed for hematological malignancies. The authors demonstrated its efficacy on 82 multiple myeloid samples from the CoMMpass project,⁽⁵⁰⁾

142 chronic lymphocytic leukemia (CLL) patients from the CLL Genome Project,⁽⁵¹⁾ and AML patients from TCGA.⁽⁵²⁾ The approach uses a dynamic error-suppression procedure rather than relying on a hard percentage cutoff for signature contributions. They were able to show the ability to accurately infer mutational signatures, even from low numbers of mutations.

MUTATIONAL SIGNATURES IN HEMATOLOGICAL MALIGNANCY AND THEIR PRECURSORS: A SURVEY OF DISCOVERIES THUS FAR

All hematological malignancies arise in cells with normal blood cell progenitors. Deep sequencing of hematopoietic stem cells (HSCs) has revealed distinct mutational signatures, even at very early stages of life, including in cord blood.⁽⁵³⁾ Some mature malignancies pass through precursor stages during which cells acquire additional mutations that confer a proliferative advantage.

In the myeloid realm, one stepwise path is from normal HSCs to clonal hematopoiesis of indeterminate potential (CHIP) to clonal cytopenia of unknown significance (CCUS) to myelodysplastic syndrome (MDS) to AML. Along this continuum, mutational signatures have only been characterized in normal HSCs, CHIP, and AML. In individuals without mutagenic exposures such as smoking and chemotherapy, mutational signatures in all these entities tend to be dominated by the clock-like signatures SBS1 and SBS5. (Note that here we use the nomenclature adopted from the COSMIC database <https://cancer.sanger.ac.uk/signatures/>). SBS1 is caused by spontaneous deamination of methylated cytosines into thymines and is very closely tied to the cell cycle. The cause of the SBS5 signature is unknown, although it is strongly correlated with age. SBS32 has also been reported to be common in normal HSCs and AML.⁽⁵³⁾ However, its association with azathioprine treatment⁽⁵⁴⁾ raises questions regarding its validity. The observation of a dominant and consistent mutational pattern in normal HSCs led some researchers to define the overall pattern as its own signature, termed SBS-HSC.^(31,35) Studies querying mutational signatures in CHIP have, unsurprisingly, revealed SBS-HSC or its components as dominant for the majority of individuals.^(55,56) As compared to healthy individuals without CHIP, mutational patterns were found to be enriched in SBS4 and SBS6.⁽⁵⁷⁾ SBS4 is associated with tobacco smoking, which is also linked to CHIP.⁽⁵⁸⁾ On the other hand, SBS6 is associated with defective DNA mismatch repair, raising the possibility that this deficiency is partially responsible for CHIP mutations in some individuals.

Although they also harbor pronounced clock-like mutational signatures, many lymphoid malignancies have signatures that are not typically found in myeloid malignancies. These include those attributed to APOBEC, activation-induced cytidine deaminase activity, DNA polymerase ϵ , UV light exposure, and MMR deficiency.⁽⁴⁷⁾

Recently, Alberge et al.⁽⁵⁹⁾ characterized mutational signatures in a cohort of patients with multiple myeloma ($n = 812$) and its precursors: Monoclonal Gammopathy of Undetermined Significance (MGUS; $n = 37$) and smoldering multiple myeloma ($n = 120$). They found 21 different signatures across the cohort, of which eight were novel. Observed signatures with known etiologies included those associated with somatic hypermutation, activation-induced cytidine deaminase (SHM/AID), APOBEC, and ROS, along with the expected clock-like signatures. Clonal hierarchy analysis allowed the authors to determine the temporal order of the different classes of signatures. They concluded that early mutations were largely driven by clock-like signatures and an AID signature before the onset of MGUS and smoldering multiple myeloma. APOBEC-driven mutational processes take hold after the initiation of these precursor conditions and then gradually develop into multiple myeloma.

Similar to solid tumors, patients with hematological malignancies who have undergone treatment typically carry signatures of the therapy. Indeed, analysis reveals signatures specifically attributed to thiopurine chemotherapy,⁽³⁶⁾ platinum chemotherapy,^(35,47,60) melphalan,⁽⁶¹⁾ and/or radiation treatment⁽³⁵⁾ that the patient had undergone.

Table 2 catalogs the mutational signatures that have been reported to date in hematological malignancies.

CONCLUSION AND FUTURE WORK

Despite the significant progress made in the study of mutational signatures over the past decade, substantial room remains for further discovery and refinement. One of the main challenges is the variability in the outputs generated by different signature extraction tools, which makes it difficult to determine the most accurate or biologically meaningful results. Currently, most reference signature sets are designed to represent mutational processes across the entire human body. However, it is well established that mutational processes vary significantly across tissues. The development of tissue-specific reference signatures could help address issues such as signature bleeding and improve the biological relevance of signature assignments.

Table 2. Description of the mutational signatures most commonly found in hematologic malignancies

Mutational process	SBS signatures	Malignancy type
Aging (clock-like signatures)	SBS1 and 5	Ubiquitous signatures
Activity of APOBEC cytidine deaminase	SBS2 and 13	MM, MGUS, SMM, DLBCL, FL, B-ALL
Defective homologous DNA repair	SBS3	MM
Defective DNA base excision repair	SBS30	MM
AID	SBS84 and 85	MM, MGUS, SMM, DLBCL, FL, B-ALL, CLL
Ultraviolet light exposure	SBS7	DLBCL and mature T and NK neoplasms
Polymerase eta somatic hypermutation	SBS9	MM, MGUS, SMM, DLBCL
Defective DNA MMR	SBS6, 15, 20, 21, 26 and 44	MM, DLBCL, FL, BLL, mature T and NK neoplasms
ROS	SBS18	MM, MGUS, SMM, B-ALL, T-ALL and AML
Unknown chemotherapy exposure	SBS86	ALL and B-ALL
Thiopurine chemotherapy exposure	SBS87	ALL and B-ALL
Platinum based chemotherapy exposure	SBS31 and 35	AML and MM
Melphalan exposure	SBS-MM1 (or 99)	AML
Unknown aetiology	SBS8, 16 and 17	MM, MGUS, SMM

AID: activation-induced cytidine deaminase; ALL: acute lymphoblastic leukemia; AML: acute myeloid leukemia; B-ALL: B cell ALL; CLL: chronic lymphocytic leukemia; DLBCL: diffuse large B cell lymphoma; FL: follicular lymphoma; MGUS: monoclonal gammopathy of undetermined significance; MM: multiple myeloma; MMR: mismatch repair; NK: natural killer; SBS: single base substitution; SMM: smoldering multiple myeloma.

Table 3. Exemplary publicly-available data sources for mutational signature inference

Data source	Data types	Disease subtype(s)	Number of patients
TARGET	WES, WGS	ALL	1006
TARGET	WES, WGS	AML	249
TCGA	WES, WGS	AML	200
TCGA	WES, WGS	Diffuse Large B-cell Lymphoma	58
BeatAML	WES	AML	534
MMRF-CoMMpass	WES, WGS	Multiple Myeloma	975
MP2PRT	WES, WGS	ALL	1507
CGCI	WGS	Burkitt Lymphoma	252
CGCI	WGS	Diffuse Large B-cell Lymphoma	66
The CLL Genome Project	WES, WGS	Chronic Lymphocytic Leukemia	284
MDS Sequencing Project	WES	Myelodysplastic Syndrome	42
All of Us	WGS	AML, MDS	1020
OHSU	WES	Chronic Myeloproliferative Disorders	158

ALL: acute lymphocytic leukemia; AML: acute myeloid leukemia; CGCI: the Cancer Genome Characterization Initiative; CLL: chronic lymphocytic leukemia; MDS: myelodysplastic syndrome; MMRF-CoMMpass: Multiple Myeloma Research Foundation-Relating Clinical Outcomes in Multiple Myeloma to Personal Assessment of Genetic Profile; MP2PRT: Molecular Profiling to Predict Response to Treatment; OHSU: Oregon Health Sciences University; TARGET: Therapeutically Applicable Research to Generate Effective Treatments; TCGA: the Cancer Genome Atlas; WES: whole-exome sequencing; WGS: whole-genome sequencing.

In addition to improving the reference sets, it would be valuable to establish guidelines for tool selection based on cancer type or tissue context, as different tools show varying levels of accuracy depending on the setting. Similarly, a better understanding of the minimum number of mutations required to reliably extract each signature is required. Some signatures, particularly the flatter ones, are inherently more difficult to detect, and applying signature analysis to low-mutation samples without this knowledge may lead to misleading conclusions.

Another promising avenue for future research is the application of mutational signatures to noncancerous tissues. As this methodology expands beyond oncology, it opens up a broad range of opportunities for exploring mutational processes in other biological contexts such as aging, inflammation, and disease precursor states.

Specifically within hematologic malignancies, research on mutational signatures remains relatively underdeveloped, particularly for myeloid neoplasms. Given the rapid evolution of tools and the increasing availability of large-scale public datasets, there is an opportunity to close this gap, even in the face of relatively low mutation rates. Table 3 provides a sample

of publicly-available whole-exome and whole-genome data sets from hematological malignancies. Most of these data sets have not been mined for mutational signatures. Applying mutational signature analysis to these diseases may also shed light on pre-leukemic conditions such as clonal hematopoiesis of indeterminate potential and clonal cytopenia of unknown significance, which are well studied in the literature but are not yet fully understood from a mutational process standpoint. Indeed, viewing mature hematological malignancy as the terminus of a continuous evolution from normal hematopoietic stem cells to precursor conditions to frank malignancy, examination of mutational signatures at each of these stages may provide clues as to which signatures are associated with progression. Thus, accurate characterization of mutational signatures in the early stages of leukemia development may aid in clinical diagnosis, prognosis, and treatment decisions.

DATA AVAILABILITY

The underlying content is contained within the manuscript.

ACKNOWLEDGMENT

US National Institutes of Health grants R01LM013067, R01CA217992, and R21CA249138.

AUTHORS' CONTRIBUTION

Felipe de Almeida Sartori: contributed to study conceptualization, conducted the literature review, and drafted and revised the manuscript. João Vitor Paes Pontes: contributed to the literature review and participated in critical revision of the manuscript. Paulo Vidal Campregher: contributed to the study conceptualization and provided clinical guidance during manuscript revision. Thomas LaFramboise: supervised the project, contributed to study conceptualization, participated in the literature review, and assisted in drafting and revising the manuscript.

AUTHORS' INFORMATION

Sartori FA: <http://orcid.org/0009-0001-4616-0304>

Pontes JV: <http://orcid.org/0009-0007-7145-6637>

Campregher PV: <http://orcid.org/0000-0001-6005-405X>

LaFramboise T: <http://orcid.org/0000-0003-1793-4777>

REFERENCES

- Dorado G, Gálvez S, Rosales TE, Vásquez VF, Hernández P. Analyzing Modern Biomolecules: The Revolution of Nucleic-Acid Sequencing - Review. *Biomolecules*. 2021;11(8):1111.
- Mardis ER. The Impact of Next-Generation Sequencing on Cancer Genomics: From Discovery to Clinic. *Cold Spring Harb Perspect Med*. 2019;9(9):a036269.
- Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009;458(7239):719-24.
- Pfeifer GP, You YH, Besaratinia A. Mutations induced by ultraviolet light. *Mutat Res*. 2005;571(1-2):19-31.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale AL, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörð JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Illicic T, Imbeaud S, Imielinski M, Jäger N, Jones DT, Jones D, Knappskog S, Kool M, Lakhani SR, López-Otín C, Martin S, Munshi NC, Nakamura H, Northcott PA, Pajic M, Papaemmanuil E, Paradiso A, Pearson JV, Puente XS, Raine K, Ramakrishna M, Richardson AL, Richter J, Rosenstiel P, Schlesner M, Schumacher TN, Span PN, Teague JW, Totoki Y, Tutt AN, Valdés-Mas R, van Buuren MM, van 't Veer L, Vincent-Salomon A, Waddell N, Yates LR; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MML-Seq Consortium; ICGC PedBrain; Zucman-Rossi J, Futreal PA, McDermott U, Lichter P, Meyerson M, Grimmond SM, Siebert R, Campo E, Shibata T, Pfister SM, Campbell PJ, Stratton MR. Signatures of mutational processes in human cancer. *Nature*. 2013;500(7463):415-21. Erratum in: *Nature*. 2013;502(7470):258. Imielinsk, Marcin [corrected to Imielinski, Marcin].
- Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ, McLaren S, Butler AP, Teague JW, Jönsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerød A, Tutt A, Martens JW, Aparicio SA, Borg A, Salomon AV, Thomas G, Børresen-Dale AL, Richardson AL, Neuberger MS, Futreal PA, Campbell PJ, Stratton MR; Breast Cancer Working Group of the International Cancer Genome Consortium. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012;149(5):979-93.
- Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep*. 2013;3(1):246-59.
- Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, Islam SMA, Lopez-Bigas N, Klimczak LJ, McPherson JR, Morganella S, Sabarinathan R, Wheeler DA, Mustonen V; PCAWG Mutational Signatures Working Group; Getz G, Rozen SG, Stratton MR; PCAWG Consortium. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94-101. Erratum in: *Nature*. 2023;614(7948):E41.
- Steele CD, Pillay N, Alexandrov LB. An overview of mutational and copy number signatures in human cancer. *J Pathol*. 2022;257(4):454-65.
- Steele CD, Abbasi A, Islam SM, Bowes AL, Khandekar A, Haase K, et al. Signatures of copy number alterations in human cancer. *Nature*. 2022;606(7916):984-91.
- Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, Khurana E, Waszak S, Korbel JO, Haber JE, Imielinski M; PCAWG Structural Variation Working Group; Weischenfeldt J, Beroukhi R, Campbell PJ; PCAWG Consortium. Patterns of somatic structural variation in human cancer genomes. *Nature*. 2020;578(7793):112-21. Erratum in: *Nature*. 2023;614(7948):E38.
- Martínez-Ruiz C, Black JRM, Puttick C, Hill MS, Demeulemeester J, Larose Cadieux E, Thol K, Jones TP, Veeriah S, Naceur-Lombardelli C, Toncheva A, Prymas P, Rowan A, Ward S, Cubitt L, Athanasopoulou F, Pich O, Karasaki T, Moore DA, Salgado R, Collier E, Castignani C, Dietzen M, Huebner A, Al Bakir M, Tanić M, Watkins TBK, Lim EL, Al-Rashed AM, Lang D, Clements J, Cook DE, Rosenthal R, Wilson GA, Frankell AM, de Carné Trécesson S, East P, Kanu N, Litchfield K, Birkbak NJ, Hackshaw A, Beck S, Van Loo P, Jamal-Hanjani M; TRACERx Consortium; Swanton C, McGranahan N. Genomic-transcriptomic evolution in lung cancer and metastasis. *Nature*. 2023;616(7957):543-52.
- Díaz-Gay M, Vangara R, Barnes M, Wang X, Islam SM, Vermees I, et al. Assigning mutational signatures to individual samples and individual somatic mutations with SigProfilerAssignment. *Bioinformatics*. 2023;39(12):btad756.
- Hübschmann D, Jopp-Saile L, Andresen C, Krämer S, Gu Z, Heilig CE, et al. Analysis of mutational signatures with yet another package for signature analysis. *Genes Chromosomes Cancer*. 2021;60(5):314-31.
- Islam SM, Díaz-Gay M, Wu Y, Barnes M, Vangara R, Bergstrom EN, et al. Uncovering novel mutational signatures by *de novo* extraction with SigProfilerExtractor. *Cell Genom*. 2022;2(11):None.
- Jin H, Gulhan DC, Geiger B, Ben-Isy D, Geng D, Ljungström V, et al. Accurate and sensitive mutational signature analysis with MuSiCal. *Nat Genet*. 2024;56(3):541-52.
- Kahane I, Leiserson MD, Sharan R. A mutation-level covariate model for mutational signatures. *PLOS Comput Biol*. 2023;19(6):e1011195.
- Manders F, Brandsma AM, de Kanter J, Verheul M, Oka R, van Roosmalen MJ, et al. MutationalPatterns: the one stop shop for the analysis of mutational processes. *BMC Genomics*. 2022;23(1):134.
- Matsutani T, Hamada M. Parallelized Latent Dirichlet Allocation Provides a Novel Interpretability of Mutation Signatures in Cancer Genomes. *Genes (Basel)*. 2020;11(10):1127.
- Matsutani T, Hamada M. Clone decomposition based on mutation signatures provides novel insights into mutational processes. *NAR Genom Bioinform*. 2021;3(4):lqab093.
- Pancotti C, Rollo C, Codicé F, Birolo G, Fariselli P, Panavia T. MUSE-XAE: MUTational Signature Extraction with eXplainable AutoEncoder enhances tumour types classification. *Bioinformatics*. 2024;40(5):btac320.

22. Rustad EH, Nadeu F, Angelopoulos N, Ziccheddu B, Bolli N, Puente XS, et al. mmsig: a fitting approach to accurately identify somatic mutational signatures in hematological malignancies. *Commun Biol.* 2021;4(1):424.
23. Wong JK, Aichmüller C, Schulze M, Hlevnjak M, Elgaafary S, Lichter P, et al. Association of mutation signature effectuating processes with mutation hotspots in driver genes and non-coding regions. *Nat Commun.* 2022;13(1):178.
24. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 2019;47 D1:D941-7.
25. Maura F, Degasperis A, Nadeu F, Leongamornlert D, Davies H, Moore L, et al. A practical guide for mutational signature analysis in hematological malignancies. *Nat Commun.* 2019;10(1):2969.
26. Medo M, Ng CK, Medová M. A comprehensive comparison of tools for fitting mutational signatures. *Nat Commun.* 2024;15(1):9467.
27. Gulhan DC, Lee JJ, Melloni GE, Cortés-Ciriano I, Park PJ. Detecting the mutational signature of homologous recombination deficiency in clinical samples. *Nat Genet.* 2019;51(5):912-9.
28. Lee D, Hua M, Wang D, Song L, Zhang T, Hua X, Yu K, Yang XR, Chanock SJ, Shi J, Landi MT, Zhu B. Pan-cancer mutational signature analysis of 111,711 targeted sequenced tumors using SATS. *medRxiv [Preprint].* 2024:2023.05.18.23290188. .
29. Sason I, Chen Y, Leiserson MD, Sharan R. A mixture model for signature discovery from sparse mutation data. *Genome Med.* 2021;13(1):173.
30. Yaacov A, Ben Cohen G, Landau J, Hope T, Simon I, Rosenberg S. Cancer mutational signatures identification in clinical assays using neural embedding-based representations. *Cell Rep Med.* 2024;5(6):101608.
31. Pich O, Reyes-Salazar I, Gonzalez-Perez A, Lopez-Bigas N. Discovering the drivers of clonal hematopoiesis. *Nat Commun.* 2022;13(1):4267.
32. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature.* 1999;401(6755):788-91.
33. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature.* 2010;463(7278):184-90.
34. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature.* 2010;463(7278):191-6.
35. Diamond B, Ziccheddu B, MacLachlan K, Taylor J, Boyle E, Ossa JA, et al. Tracking the evolution of therapy-related myeloid neoplasms using chemotherapy signatures. *Blood.* 2023;141(19):2359-71.
36. Li B, Brady SW, Ma X, Shen S, Zhang Y, Li Y, et al. Therapy-induced mutations drive the genomic landscape of relapsed acute lymphoblastic leukemia. *Blood.* 2020;135(1):41-55.
37. Hoang PH, Cornish AJ, Dobbins SE, Kaiser M, Houlston RS. Mutational processes contributing to the development of multiple myeloma. *Blood Cancer J.* 2019;9(8):60.
38. Buisson R, Lawrence MS, Benes CH, Zou L. APOBEC3A and APOBEC3B Activities Render Cancer Cells Susceptible to ATR Inhibition. *Cancer Res.* 2017;77(17):4567-78.
39. Brady SW, Liu Y, Ma X, Gout AM, Hagiwara K, Zhou X, et al. Pan-neuroblastoma analysis reveals age- and signature-associated driver alterations. *Nat Commun.* 2020;11(1):5183.
40. Dalton KM, Lochmann TL, Floros KV, Calbert ML, Kurupi R, Stein GT, et al. Catastrophic ATP loss underlies a metabolic combination therapy tailored for MYCN-amplified neuroblastoma. *Proc Natl Acad Sci USA.* 2021;118(13):e2009620118.
41. McLeod C, Gout AM, Zhou X, Thrasher A, Rahbarinia D, Brady SW, et al. St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem. *Cancer Discov.* 2021;11(5):1082-99.
42. Steele CD, Tarabichi M, Oukrif D, Webster AP, Ye H, Fittall M, et al. Undifferentiated Sarcomas Develop through Distinct Evolutionary Pathways. *Cancer Cell.* 2019;35(3):441-456.e8.
43. Dziubańska-Kusibab PJ, Berger H, Battistini F, Bouwman BA, Iftikhar A, Katainen R, et al. Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. *Nat Med.* 2020;26(7):1063-9.
44. Chen H, Chong W, Teng C, Yao Y, Wang X, Li X. The immune response-related mutational signatures and driver genes in non-small-cell lung cancer. *Cancer Sci.* 2019;110(8):2348-56.
45. Chopra N, Tovey H, Pearson A, Cutts R, Toms C, Proszek P, et al. Homologous recombination DNA repair deficiency and PARP inhibition activity in primary triple negative breast cancer. *Nat Commun.* 2020;11(1):2662.
46. Wang S, Li H, Song M, Tao Z, Wu T, He Z, et al. Copy number signature analysis tool and its application in prostate cancer reveals distinct mutational processes and clinical outcomes. *PLoS Genet.* 2021;17(5):e1009557.
47. Ptashkin RN, Ewalt MD, Jayakumar G, Kiecka I, Bowman AS, Yao J, et al. Enhanced clinical assessment of hematologic malignancies through routine paired tumor and normal sequencing. *Nat Commun.* 2023;14(1):6895.
48. DeRoin L, Cavalcante de Andrade Silva M, Petras K, Arndt K, Phillips N, Wanjar P, et al. Feasibility and limitations of cultured skin fibroblasts for germline genetic testing in hematologic disorders. *Hum Mutat.* 2022;43(7):950-62.
49. Shinde J, Bayard Q, Imbeaud S, Hirsch TZ, Liu F, Renault V, et al. Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer. *Bioinformatics.* 2018;34(19):3380-1.
50. Skerget S, Penaherrera D, Chari A, Jagannath S, Siegel DS, Vij R, Orloff G, Jakubowiak A, Niesvizky R, Liles D, Berdeja J, Levy M, Wolf J, Usmani SZ; MMRF CoMMpass Network; Christofferson AW, Nasser S, Aldrich JL, Legendre C, Benard B, Miller C, Turner B, Kurdoglu A, Washington M, Yellapantula V, Adkins JR, Cuyugan L, Boateng M, Helland A, Kyman S, McDonald J, Reiman R, Stephenson K, Tassone E, Blanski A, Livermore B, Kirchhoff M, Rohrer DC, D'Agostino M, Gambella M, Collison K, Stumph J, Kidd P, Donnelly A, Zaugg B, Toone M, McBride K, DeRome M, Rogers J, Craig D, Liang WS, Gutierrez NC, Jewell SD, Carpten J, Anderson KC, Cho HJ, Auclair D, Lonial S, Keats JJ. Comprehensive molecular profiling of multiple myeloma identifies refined copy number and expression subtypes. *Nat Genet.* 2024;56(9):1878-1889. Erratum in: *Nat Genet.* 2025;57(7):1789.
51. Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature.* 2011;475(7354):101-5.
52. Cancer Genome Atlas Research Network; Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, Hoadley K, Triche TJ Jr, Laird PW, Baty JD, Fulton LL, Fulton R, Heath SE, Kalicki-Veizer J, Kandoth C, Klotz JM, Koboldt DC, Kanchi KL, Kulkarni S, Lamprecht TL, Larson DE, Lin L, Lu C, McLellan MD, McMichael JF, Payton J, Schmidt H, Spencer DH, Tomasson MH, Wallis JW, Wartman LD, Watson MA, Welch J, Wendl MC, Ally A, Balasundaram M, Birol I, Butterfield Y, Chiu R, Chu A, Chuah E, Chun HJ, Corbett R, Dhalla N, Guin R, He A, Hirst C, Hirst M, Holt RA, Jones S, Karsan A, Lee D, Li H, Marra MA, Mayo M, Moore RA, Mungall K, Parker J, Pleasance E, Plettner P, Schein J, Stoll D, Swanson L, Tam A, Thiessen N, Varhol R, Wye N, Zhao Y, Gabriel S, Getz G, Sougnez C, Zou L, Leiserson MD, Vandin F, Wu HT, Applebaum F, Baylin SB, Akbani R, Broom BM, Chen K, Motter TC, Nguyen K, Weinstein JN, Zhang N, Ferguson ML, Adams C, Black A, Bowen J, Gastier-Foster J, Grossman T, Lichtenberg T, Wise L, Davidsen T, Demchok JA, Shaw KR, Sheth M, Sofia HJ, Yang L, Downing JR, Eley G. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med.* 2013;368(22):2059-74. Erratum in: *N Engl J Med.* 2013;369(1):98.
53. Osorio FG, Rosendahl Huber A, Oka R, Verheul M, Patel SH, Hasaart K, et al. Somatic Mutations Reveal Lineage Relationships and Age-Related Mutagenesis in Human Hematopoiesis. *Cell Rep.* 2018;25(9):2308-2316.e4.
54. Inman GJ, Wang J, Nagano A, Alexandrov LB, Purdie KJ, Taylor RG, et al. The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature. *Nat Commun.* 2018;9(1):3667.
55. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med.* 2014;371(26):2477-87.

56. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014;371(26):2488-98.
57. Bick AG, Weinstock JS, Nandakumar SK, Fulco CP, Bao EL, Zekavat SM, Szeto MD, Liao X, Leventhal MJ, Nasser J, Chang K, Laurie C, Burugula BB, Gibson CJ, Lin AE, Taub MA, Aguet F, Ardlie K, Mitchell BD, Barnes KC, Moscati A, Fornage M, Redline S, Psaty BM, Silverman EK, Weiss ST, Palmer ND, Vasan RS, Burchard EG, Kardia SLR, He J, Kaplan RC, Smith NL, Arnett DK, Schwartz DA, Correa A, de Andrade M, Guo X, Konkle BA, Custer B, Peralta JM, Gui H, Meyers DA, McGarvey ST, Chen Y, Shoemaker MB, Peyser PA, Broome JG, Gogarten SM, Wang FF, Wong Q, Montasser ME, Daya M, Kenny EE, North KE, Launer LJ, Cade BE, Bis JC, Cho MH, Lasky-Su J, Bowden DW, Cupples LA, Mak ACY, Becker LC, Smith JA, Kelly TN, Aslibekyan S, Heckbert SR, Tiwari HK, Yang IV, Heit JA, Lubitz SA, Johnsen JM, Curran JE, Wenzel SE, Weeks DE, Rao DC, Darbar D, Moon JY, Tracy RP, Buth EJ, Rafaels N, Loos RJF, Durda P, Liu Y, Hou L, Lee J, Kachroo P, Freedman BI, Levy D, Bielak LF, Hixson JE, Floyd JS, Whitsel EA, Ellinor PT, Irvin MR, Fingerlin TE, Raffield LM, Armasu SM, Wheeler MM, Sabino EC, Blangero J, Williams LK, Levy BD, Sheu WH, Roden DM, Boerwinkle E, Manson JE, Mathias RA, Desai P, Taylor KD, Johnson AD; NHLBI Trans-Omics for Precision Medicine Consortium; Auer PL, Kooperberg C, Laurie CC, Blackwell TW, Smith AV, Zhao H, Lange E, Lange L, Rich SS, Rotter JI, Wilson JG, Scheet P, Kitzman JO, Lander ES, Engreitz JM, Ebert BL, Reiner AP, Jaiswal S, Abecasis G, Sankaran VG, Kathiresan S, Natarajan P. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature*. 2020;586(7831):763-768. Erratum in: *Nature*. 2021;591(7851):E27.
58. Dawoud AA, Tapper WJ, Cross NC. Clonal myelopoiesis in the UK Biobank cohort: ASXL1 mutations are strongly associated with smoking. *Leukemia*. 2020;34(10):2660-72.
59. Alberge JB, Dutta AK, Poletti A, Coorens TH, Lightbody ED, Toenges R, et al. Genomic landscape of multiple myeloma and its precursor conditions. *Nat Genet*. 2025;57(6):1493-503.
60. Pich O, Cortes-Bullich A, Muiños F, Pratcorona M, Gonzalez-Perez A, Lopez-Bigas N. The evolution of hematopoietic cells under cancer therapy. *Nat Commun*. 2021;12(1):4803.
61. Rustad EH, Yellapantula V, Leongamornlert D, Bolli N, Lederger G, Nadeu F, et al. Timing the initiation of multiple myeloma. *Nat Commun*. 2020;11(1):1917.