



OPEN ACCESS

EDITED BY
Dimitrios Kleftogiannis,
University of Bergen, Norway

REVIEWED BY
David Goode,
Peter MacCallum Cancer Centre,
Australia
Sayaka Miura,
Temple University, United States

*CORRESPONDENCE
Kimberly J. Bussey,
kbusse@midwestern.edu

†PRESENT ADDRESS
Luis H. Cisneros,
Biodesign Center for Biocomputing,
Security and Society, Arizona State
University, Tempe, AZ, United States
Charles Vaske,
Claret Bioscience, LLC, Santa Cruz, CA,
United States
Kimberly J. Bussey,
Precision Medicine, Midwestern
University, Glendale, AZ, United States

†These authors have contributed equally
to this work

SPECIALTY SECTION
This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

RECEIVED 30 April 2022
ACCEPTED 05 August 2022
PUBLISHED 06 September 2022

CITATION
Cisneros LH, Vaske C and Bussey KJ
(2022), Identification of a signature of
evolutionarily conserved stress-induced
mutagenesis in cancer.
Front. Genet. 13:932763.
doi: 10.3389/fgene.2022.932763

COPYRIGHT
© 2022 Cisneros, Vaske and Bussey.
This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Identification of a signature of evolutionarily conserved stress-induced mutagenesis in cancer

Luis H. Cisneros^{1,2†}, Charles Vaske^{1†} and
Kimberly J. Bussey^{1,2,3*†}

¹NantOmics, LLC, Santa Cruz, CA, United States, ²The Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ, United States, ³Precision Medicine, Midwestern University, Glendale, AZ, United States

The clustering of mutations observed in cancer cells is reminiscent of the stress-induced mutagenesis (SIM) response in bacteria. Bacteria deploy SIM when faced with DNA double-strand breaks in the presence of conditions that elicit an SOS response. SIM employs *DinB*, the evolutionary precursor to human trans-lesion synthesis (TLS) error-prone polymerases, and results in mutations concentrated around DNA double-strand breaks with an abundance that decays with distance. We performed a quantitative study on single nucleotide variant calls for whole-genome sequencing data from 1950 tumors, non-inherited mutations from 129 normal samples, and acquired mutations in 3 cell line models of stress-induced adaptive mutation. We introduce statistical methods to identify mutational clusters, quantify their shapes and tease out the potential mechanism that produced them. Our results show that mutations in both normal and cancer samples are indeed clustered and have shapes indicative of SIM. Clusters in normal samples occur more often in the same genomic location across samples than in cancer suggesting loss of regulation over the mutational process during carcinogenesis. Additionally, the signatures of TLS contribute the most to mutational cluster formation in both patient samples as well as experimental models of SIM. Furthermore, a measure of cluster shape heterogeneity was associated with cancer patient survival with a hazard ratio of 5.744 (Cox Proportional Hazard Regression, 95% CI: 1.824–18.09). Our results support the conclusion that the ancient and evolutionary-conserved adaptive mutation response found in bacteria is a source of genomic instability in cancer. Biological adaptation through SIM might explain the ability of tumors to evolve in the face of strong selective pressures such as treatment and suggests that the conventional ‘hit it hard’ approaches to therapy could prove themselves counterproductive.

KEYWORDS

cancer, stress induced mutagenesis, trans-lesion DNA synthesis, evolution, mutational clusters, cancer evolution, intratumor heterogeneity

Introduction

Genomic instability is a well-known hallmark of cancer manifested as higher than normal rates of genomic mutations. However, these mutations do not typically arise at uniformly random locations across the genome. Rather, they follow a non-uniform distribution resulting in mutational clustering (Drake, 2007; Wang et al., 2007; Ye et al., 2010; Nik-Zainal et al., 2012; Roberts et al., 2012; Alexandrov et al., 2013; Kamburov et al., 2015; Nik-Zainal et al., 2016; Chen et al., 2019). This phenomenon is observed in its extreme form as kataegis, consisting of 6 or more mutations with inter-mutational distances of 1 kb or less (Alexandrov et al., 2013; Nik-Zainal et al., 2016). Given that most mutations are either neutral or deleterious, the likelihood that randomly distributed mutations would result in gains in fitness is considered to be low (Ram and Hadany, 2019). But concerted patches of mutation, particularly when occurring within a gene, are more likely to result in alterations that could contribute to neofunctionalization and increased cellular fitness (Drake, 2007; Ram and Hadany, 2019; Cortés-Ciriano et al., 2020). Recent work has shown evidence of enrichment of clustered mutations in genes relative to intergenic spaces in cancer samples even though generally more mutations occur outside of genes than in them (Cisneros et al., 2017; Supek and Lehner, 2017). In particular, mutation clustering in non-coding regions has been associated with structural changes that possibly cause elevated mutation rates but by themselves very rarely constitute oncogenic drivers (Nik-Zainal et al., 2016; Rheinbay et al., 2020).

Large mutational loads in human cancer have been associated with replication repair deficiency (Campbell et al., 2017; Ma et al., 2018; Campbell et al., 2021), and thus underlying defects in the DNA repair machinery are thought to lead to biases in the types and locations of passenger mutations and structural events acquired during the progression of cancer. These general ideas justify targeting DNA repair and checkpoint inhibitors in cancer therapies (Murai, 2017; Forment and O'Connor, 2018; Ubhi and Brown, 2019; Zhu et al., 2020). Previous studies have identified the action of the AID/APOBEC family of cytosine deaminases as well as the action of Pol- η as contributing mechanisms to the phenomenon of mutational clustering (Lada et al., 2012; Roberts et al., 2013; Taylor et al., 2013; Supek and Lehner, 2017; Buisson et al., 2019; Roper et al., 2019; Shi et al., 2020) and underlying kataegis in particular. However, these processes only explain a subset of the mutational clusters observed and thus a more general mechanism remains to be determined.

Stress-induced mutagenesis (SIM) in bacteria occurs when double-strand break damage (DSB) happens in the context of sufficient cellular stress to initiate the SOS response (McKenzie et al., 2000; McKenzie et al., 2001; Foster, 2007; Janion, 2008; Rosenberg, 2010; Shee et al., 2012). SIM has been shown to increase the mutation rates locally around DNA lesions as cells

strive to adapt to the challenging environment (Foster, 2007; Rosenberg, 2010; Fitzgerald et al., 2017). During a double-strand break mediated mutagenesis in bacteria, DNA repair switches from high-fidelity homologous recombination to a repair mechanism that relies on the error-prone DNA polymerase, *DinB*. The result of this mechanism is a spectrum of both single nucleotide variants (SNV) and copy number amplifications, with a molecular signature consisting of clustering of SNVs spanning several kilobases in size and with a decaying frequency as a function of the distance from the DSB site. This pattern remains above the background neutral noise for up to a megabase (Rosenberg et al., 2012; Shee et al., 2012; Fitzgerald et al., 2017). The evidence of mutational clustering combined with the observation of intra-tumor chromosomal structural heterogeneity that characterizes many cancers (Roschke et al., 2002; Roschke et al., 2003; Roschke et al., 2005) prompted us to inquire whether a process comparable to bacterial SIM takes place during carcinogenesis. This idea was previously suggested by Fitzgerald, Xia, Rosenberg, and others (Fitzgerald et al., 2017; Xia et al., 2019). Expression of adaptive mutagenesis has been shown in the context of the emergence of drug resistance, with evidence of down-regulation of mismatch repair (MMR) and homologous recombination (HR), and up-regulation of error-prone polymerases in drug-tolerant colorectal tumor cells (Russo et al., 2019). Furthermore, mTOR stress signaling has been shown to facilitate SIM in multiple human cancer cell lines exposed to non-genotoxic drug selection (Cipponi et al., 2020).

In humans, the orthologous genes to *DinB* have become specialized for translesion synthesis (TLS). The closest orthologous protein to *DinB* is Pol- κ , one of several DNA polymerases involved in TLS (Waters et al., 2009). These error-prone DNA polymerases are capable of high-fidelity synthesis against the damaged bases they recognize but exhibit orders of magnitude less fidelity when the template is undamaged. TLS is employed during normal replication as a mechanism to bypass DNA damage (Waters et al., 2009) and as part of microhomology-mediated breakage-induced repair (Sakofsky et al., 2015), two processes active in cancer. The dysregulation of cell cycle and DNA repair that characterizes most tumors would also logically increase the need for TLS in cancer.

We investigated SNV distributions, observed by whole genome sequencing of non-inherited mutations in normal samples and a wide variety of solid tumors, for evidence of mutational clustering. We inquired whether the molecular signal of SIM can be identified by measuring cluster geometry, and how these observations relate to clinical outcomes. We found clear evidence of mutational clustering as demonstrated by enrichment of closer than expected mutations, particularly for samples with low mutational loads. By characterizing the distributions of clusters, we observed that there is a greater consistency of cluster locations across normal samples than in

cancer samples, suggesting a degree of regulation control for mutations in normal tissue that breaks down during carcinogenesis. We found that clusters displayed the mutational geometry that characterizes SIM in bacteria. We also studied the potential mechanisms that could have resulted in the observed somatic mutation profile. We concluded that TLS is the primary driving force behind clusters with mutational geometries characteristic of SIM in both cancer and non-inherited mutation in normal individuals. Furthermore, we associated these findings with clinical outcomes determining that the diversity of SNV distribution within clusters in the tumor is a poor prognostic factor for patients with cancer.

Methods

Data

We obtained variant calls for normal and cancer samples from public repositories where all cases had been called by a standard pipeline. For non-inherited mutations in normal tissue, we used whole-genome sequencing data (WGS) from the Complete Genomics Indices database in the 1000 Genomes Project ([The 1000 Genomes Project Consortium et al., 2015](#)) release 20130502 (RRID:SCR_006828, [Supplementary Table S7 in Cisneros et al. \(2017\)](#) for a list of donors). We refer to this set as CGI. These data have an average genome coverage of 47X. The variant call tables (VCF) of 129 trios were analyzed using the `vcf_contrast` function from the VCFTools analysis toolbox to compare each child with the two corresponding parents. The resulting potential novel variants were then filtered such that the child and both parents must be flagged as PASS (i.e., the variant passed all filters in the calling algorithm); the child must have a read depth of at least 20, and the alternative (aka novel) allele frequency was $VAF \geq 0.35$.

For cancer samples, we analyzed simple somatic mutations and corresponding clinical data from the PCAWG coordinated WGS calls for 1,950 tumor samples from 1,830 donors representing 14 different primary sites ([Campbell et al., 2017](#)). Somatic variants for all data sets were classified as previously published ([Cisneros et al., 2017](#)). In addition to the pan-cancer data, we obtained experimental data published in a recent study about the role of MTOR in adaptive evolution in cancer by [Cipponi et al. \(2020\)](#) (available in the NCBI's BioProject database accession: PRJNA623123). This corresponds to WGS calls (average coverage 116X) on single cell-derived clonal populations from SKMEL28 human melanoma and the 94T778 human liposarcoma cell lines exposed to different treatments: SKMEL28 and 94T778 naïve and exposed to drug lines (vemurafenib and tunicamycin correspondingly), plus a line of 94T778 with MTOR silenced (via FRAP1 knockdown) and the corresponding control. Each branch of the study consists of 5 samples giving a total of 30 samples. We call this set

KCCCCG. Vemurafenib is an inhibitor of the V600E mutation in BRAF. Tunicamycin inhibit protein glycosylation, leading to an unfolded protein response. The MTOR signaling pathway is an evolutionarily conserved sensor of environmental and endogenous stress generally expressed in human tumors ([Saxton and Sabatini, 2017](#); [Cipponi et al., 2020](#)). These data allow us to isolate the effect of SIM independent of DNA damaging agents.

For our theoretical data, we generated 500 replicates for eight groups of simulated mutations defined by their total mutational load ($N_{SNV} = 500, 1000, 2500, 5000, 10000, 25000, 50000, 100000$). We modeled a uniform, random distribution of SNVs across the genome as a one-dimensional Bernoulli Process. These simulations correspond to our null model used to assess genomic heterogeneity in observed data.

Detection of mutation clusters and cluster shapes

A group of $n > 3$ SNVs is deemed a “cluster” if it is set of consecutive mutations with interevent distance closer than $D^* = 15$ kb (tuple) and its probability, according to the negative binomial test ([Roberts et al., 2012](#); [Cisneros et al., 2017](#)), is less than 1% ([Supplementary Materials](#)). In other words, a cluster is a group of variations that is statistically unlikely in the mutational background of the sample. The specific value $D^* = 15$ kb was chosen because it is an adequate balance between signal and noise ([Supplementary Materials](#)).

To check if variations in background mutational density are responsible for the observed clustering, we looked for an association of clustering with various descriptions of chromatin domain structure. We use chromatin domain annotation as defined by [Libbrecht et al. \(2015\)](#) and topologically associating domain annotation along with boundaries as defined by [Akdemir et al. \(2020a\)](#). These two annotations integrate multiple epigenetic marks with transcriptional activity and replication timing across multiple tissue types. In both normal and cancer data, while we did see the previously established differences in mutational background as a function of domain, there was no difference in the distribution of SNVs found in our clusters versus those not residing in clusters and their presence in various chromatin domains as defined by Libbrecht, et al. (Chi-square = 30, simulated p -value after 1000 repetitions = 1 for both normal and cancer). The same result was seen using the TAD classification data from Akdemir et al. Therefore, the background effect of non-uniform mutational density is sufficiently small as to not influence our detection of clusters. Difference in mutational loads across the genome reflects the activity of repair mechanisms and the ability of cells to tolerate mutation at particular locations but does not in and of itself predispose to mutations that are closer than would be expected according to our null hypothesis.

For each WGS sample in our database, all possible clusters were identified and the “center of mass” (genomic location of cluster centroid) in each case was calculated, along with other properties such as start and end locations, length, and size (number of variations) (Cisneros et al., 2017). We treated cluster centroids as likely locations of the DSBs that induced the accumulation of variations. Therefore, the expected signature for stress-induced mutagenesis should be evident as a concentration of mutations around these centroids that decays with distance from it. Thus, for each cluster i , we computed the cumulative distribution of SNV events $F_i(X)$ as a function of the distance X from the cluster centroid up to 250 kb and in both the 3' and the 5' directions. This window size was chosen based on the observations in E.coli that SIM led to elevated mutation rates up to 1 MB away from the double-strand break (Shee et al., 2011). By aggregating all cumulative distributions observed in each sample we generated a representative overall curve $F(X) = \sum F_i(X)$ that conveys the probability of finding a mutation at a given distance from the cluster center. If the distribution of SNV events is uniformly random (and therefore does not typically decay) then $F(X)$ is expected to increase proportionally with X . This assumption gives us a background of mutations against which we can compare the observed distribution pattern. To define a useful score, we normalize X by 250 kb and F by the number of events closer than 250 kb, thus mapping all cluster-associated cumulative distribution curves to a unit box:

$$\frac{X}{250\text{kb}} \rightarrow x \text{ with } x \in [0, 1]$$

$$\frac{F(X)}{F(250\text{kb})} \rightarrow f(x) \text{ with } f(x) \in [0, 1]$$

If the null hypothesis were correct for these events, $f(x) = x$. We define a measure of the degree of deviation from the null hypothesis by integrating the difference between the normalized cumulative distribution $f(x)$ and the expected value x as follows:

$$S(f(s)) = 2 \cdot \sum_{x=0}^{x=1} (f(x) - x)$$

The value of S is a signed statistic with a range $S \in [-1, 1]$ (Figure 1). As S approaches to one, smaller windows close to the origin (i.e., cluster center) contain more events than expected from a random uniform distribution, indicating that SNV events concentrate near the center of the clusters and sharply decay with the distance. A negative S value indicates that the events are typically depleted from the center and concentrated on the edges of the cluster, and S values close to zero indicate that events are mostly uniformly distributed across the 250 kb interval length, supporting the null hypothesis. We call this the **Overall Stress Introduced Heterogeneity** (SIth) score of the distribution of somatic SNVs and use it to represent the typical cluster geometry in a sample.

Following the same definition for individual clusters we can estimate a **Cluster SIth** score using the function $F_i(X)$ instead of $F(X)$, thus leading to $S_i = S(f_i(x))$. This definition is statistically less robust than the overall measure but allows us to assess the diversity of behaviors in clusters within a sample. We do this by estimating the quartile statistics on the ensemble of S_i values for each sample. We use the interquartile range of Cluster SIth scores in each sample, called SIth IQR, to score this diversity.

Mutation motif analysis

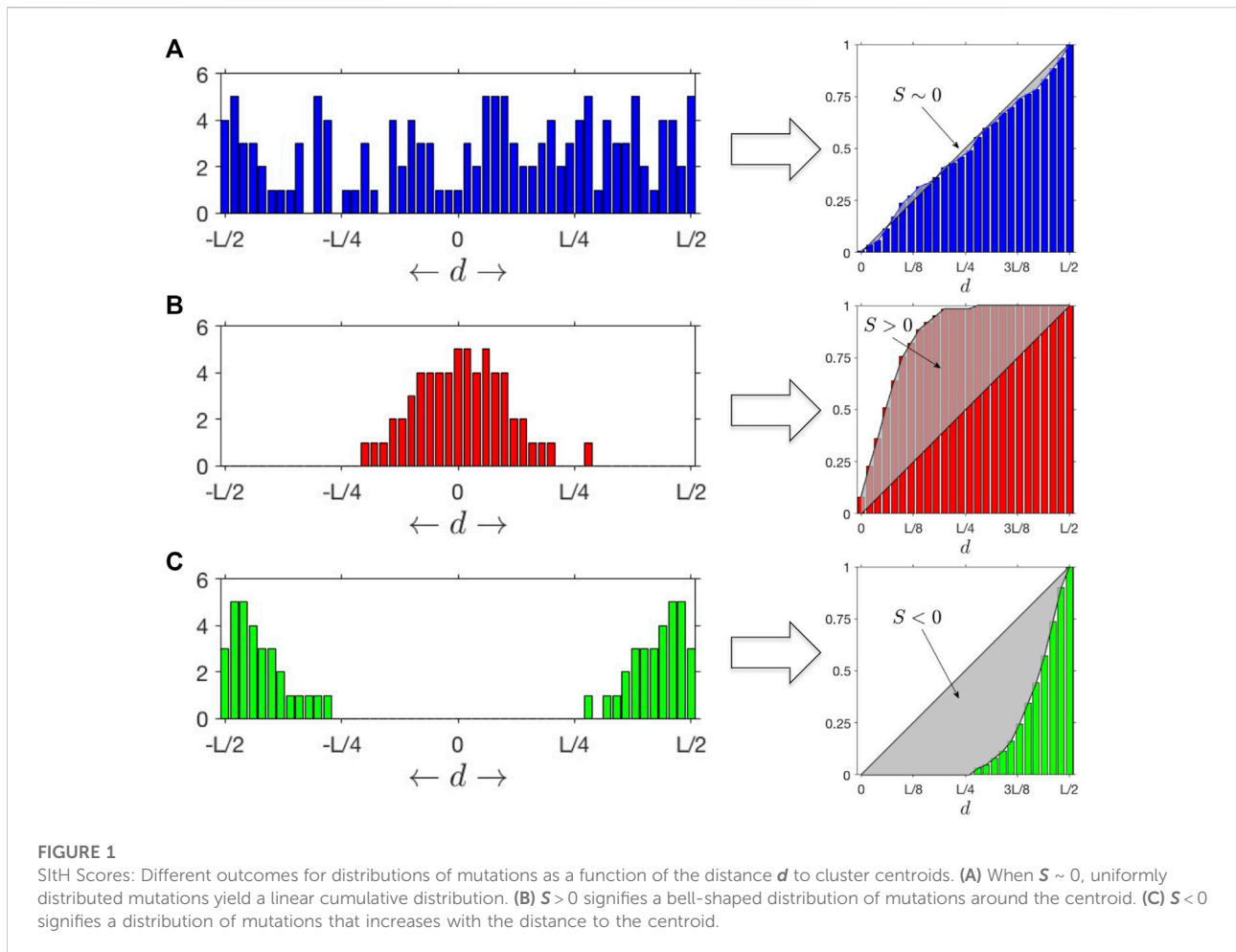
SNV mutant variants were compared to their corresponding wild-type reference sequence to match each contextual mutational pattern to motifs specific to TLS (REV1, POLH, POLK, POLQ, POLM and POLI), APOBEC and AID mechanisms according to the rules shown in Table 1 (Livneh, 2001; Goodman, 2002; Waters et al., 2009; Goodman and Woodgate, 2013) Because the rules for REV1, POLH_1, POLH_2, and POLK can also result from a failure of mismatch repair, we do not include those motifs in our analysis. Therefore, our results slightly underestimate the TLS contribution to the mutational load.

Mutational signature analysis

Mutational signature analysis of clustered SNVs was done in R (version 4.1.2) using Bioconductor (version 3.14), the MutationalPatterns package (version 3.14) and the reference genome BSgenome.Hsapiens.UCSC.hg19. As the purpose of this analysis was to compare signatures found in the clustered SNVs to known catalogs of signatures COSMIC v2 [https://cancer.sanger.ac.uk/signatures/signatures_v2/, (Alexandrov et al., 2013)] and the 82 substitution reference signatures from the SIGNAL project [<https://signal.mutationalsignatures.com/explore/cancer>, (Degasperis et al., 2022)], we did not optimize for *de novo* signature extraction, but designated 30 (the number of COSMIC v2 signatures) as the number of signatures to extract to facilitate comparison to the two reference signature profiles.

Statistical analysis

All statistical analysis was done in R (version 4.1.2). Two tailed Fisher tests, ANOVA, and Benjamini-Hochberg multiple comparison adjustments to p -values were done using the stats package (version 4.1.2). Survival analysis was performed using the survival package (version 3.2-13). Kaplan-Meier curves and Cox Proportional Hazard Regressions were calculated using the survminer package (version 0.4.9).



Results

Clustering as a function of total mutational load

We began our study by looking at the patterns of mutational density across the genome in non-inherited mutations from 129 normal individuals (CGI), somatic mutations in 1950 tumors from 14 different tissues (PCAWG) and somatic mutations in 30 cancer cell line samples exposed to different drug conditions (KCCCG). When looking at the distributions of groups of SNVs with inter-SNV distances of ≤ 15 kb (which we deem “tuples”), we observed an enrichment in the number of tuples and an under-representation of singletons (SNVs that do not belong to a tuple, or equivalently a tuple of just one element) for normal and cancer data, indicating the number of proximal SNVs is larger than expected in all cases (Figure 2A). This result was particularly significant for samples with a small number of mutations. At low total number of SNVs only a handful of tuples are expected yet dozens to hundreds were typically observed in cancer samples.

If the mutational process was dependent on genomic location tuple locations would be more frequent across samples. We compared the tuple locations across samples for which tuple enrichment was most obvious (Figure 2A): $N_s = 129$ normal samples and $N_s = 784$ cancer samples with $N_{SNV} < 5100$. We identified all regions in the genome containing tuples in at least 8.8% of the normal samples and 3.5% of the cancer samples (providing confidence that the observation is above the Poisson-counting error statistic in each case). For normal samples we found 128 overlapping regions, some containing tuples in as many as 30% of the samples. Many of these regions are located close together, as shown in Figure 2B, none of them were longer than 30 kb, and about a quarter of them involved and overlap of just single base mutations across the samples. In contrast, cancer samples had few overlaps. We observed only 19 regions grouped into five distinct ranges (Figure 2B): an 117-kb region in chromosome 6 associated with the human leukocyte antigen (HLA) complex, which contained tuples in 7% of the samples; two ~ 1 -kb regions in chromosomes 2 and 3 in 3.5% of the samples; a single point mutation in chromosome 1, associated with the zinc finger protein ZNF678, overlapping in $\sim 4\%$ of the

TABLE 1 Contextual rule motifs for each mutational mechanism. The character "N" indicates a wild-card (i.e., any nucleotide) and characters between parentheses indicate synonymous options (i.e., "(A|T)" means "A" or "T").

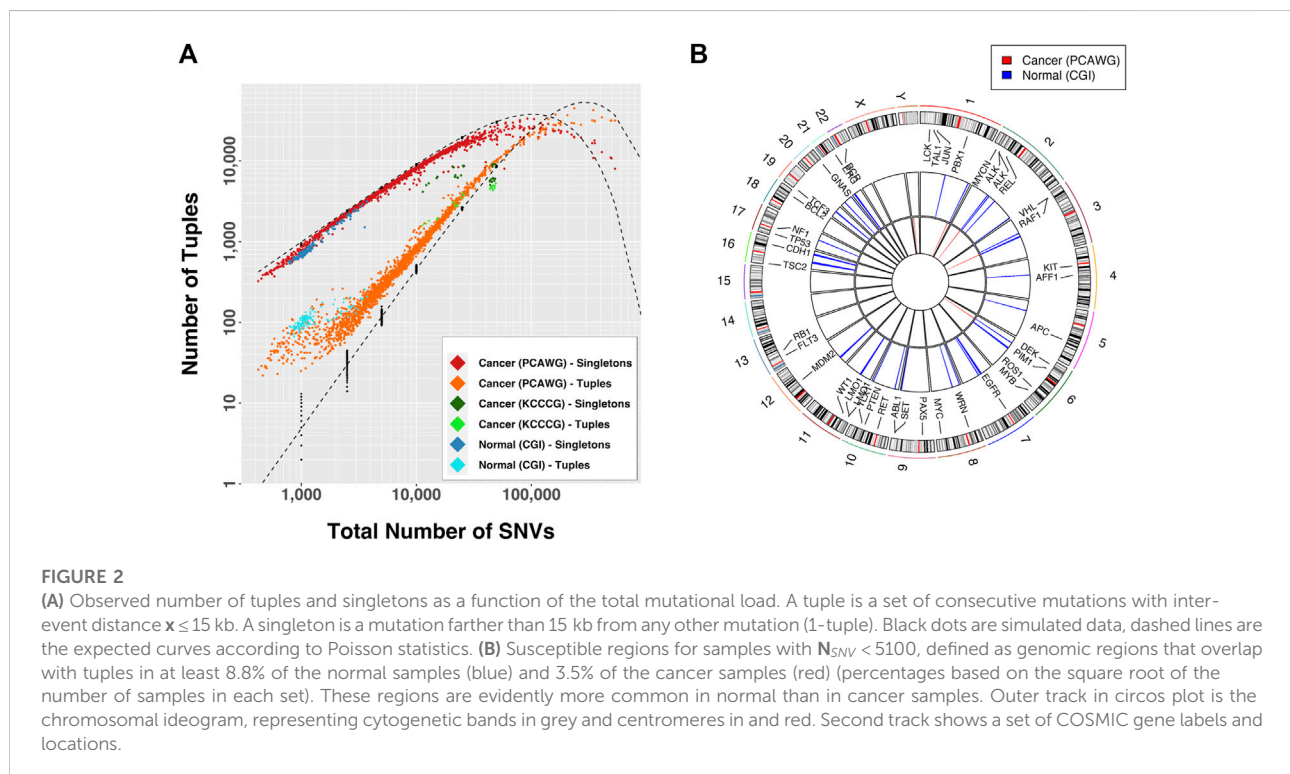
Mechanism	Motif name	Wild match	Mutant match
TLS	REV1_1	(A T G)	C
	REV1_2	(T A C)	G
	POLH_1	T	C
	POLH_2	A	G
	POLH_3	NN	(A G)A
	POLH_4	NN	T (T C)
	POLK_1	G	A
	POLK_2	C	T
	POLQ_3	NNN	AG (C T)
	POLQ_4	NNN	(G A)CT
	POLM_1	G	-
	POLM_2	C	-
	POLI_1	GNN	CAG
	POLI_2	CNN	CTG
APOBEC	APOBEC_1	TC (A T)	TT (A T)
	APOBEC_2	(T A)GA	(T A)AA
AID	AID_1	(A T) (A G)C	(A T) (A G)T
	AID_2	G (T C) (A T)	A (T C) (A T)

samples; and a half-kb region in chromosome Y with 11% overlap in the 479 male samples. This analysis excluded the cell line data because the number of variations in those was large and the number of samples too small.

These observations suggest that a differential mutational rate across the genome is likely a combination of bias in recovery, differential DNA repair efficiencies, genomic location, and interdependency between mutational events, as recognized by others (Martincorena and Campbell, 2015; Supek and Lehner, 2015; Supek and Lehner, 2017; Akdemir et al., 2020b). As the total number of mutations increases the distributions approached the predicted curve, but then departed again. In fact, for large mutational loads the relationship between the proportion of tuples and singletons with respect to the expectation was inverted. In this case the conclusion is that certain regions in the genome are protected from accumulation of mutations, a process that renders sections with fewer than the expected number events.

SlTH scores in normal tissue, cancer, and cancer cell lines

Most algorithms for finding mutational signatures linked to mutational mechanisms look for patterns in the actual sequence changes. However, SIM's predicted pattern is not one of specific



sequence change, but rather SNV distribution geometry. Previous work demonstrates that SNVs cluster together in both normal tissues and cancer, and the sequence contexts of both the reference and mutant calls can be used to infer mechanism (Roberts et al., 2012). The association of APOBEC cytosine deaminases with clusters is well established (Lada et al., 2012; Burns et al., 2013; Roberts et al., 2013; Taylor et al., 2013), but it only accounts for at most 50% of the clusters observed (Roberts et al., 2013). Furthermore, nothing in the mechanism of APOBEC suggests a characteristic shape of the mutational clusters. In contrast, the SIM response of bacteria, mediated by the Pol IV polymerase (encoded by *DinB*), leads to a clustering pattern in which more SNVs are found at the center of the cluster than at the edges (Shee et al., 2012). Therefore, we developed a score that measure how SNVs are distributed within a cluster. Called the Stress-Introduced Heterogeneity (SiTH) score (see Methods), it discriminates between clusters where SNVs are uniformly distributed, those where the density of SNVs is toward the edges of the cluster (negative scores) and those where the density of SNVs is concentrated in the middle of the cluster (positive scores, Figure 1). SIM is predicted to result in clusters with positive scores. SiTH scores can be computed as single score based on all clusters found in a sample and representing the average cluster “shape,” or can be computed on a cluster-by-cluster basis to define how variable cluster shape is within a sample. In this situation, we compute the inter-quartile range as we have no reason to believe that SiTH scores will be normally distributed.

The overall SiTH scores ranged from 0.145 to 0.999 (Figure 3A) and varied significantly by organ site and whether the tumor was one of multiple tumors from a single donor (ANOVA, organ site, $F = 136.70$, $p < 2.2 \times 10^{-16}$; multiple tumor, $F = 3.07$, $p = 0.0799$; maximum SiTH Score, $F = 16.14$, $p = 6.098 \times 10^{-5}$). We observed SiTH scores declined as a function of the total number of SNVs (Figure 3A) in tumors. This decay is not really observed in normal samples because the range of change of mutational burden is too narrow, but it follows the lower end of the general trend of the cancer samples. Cancer cell lines have a different behavior all together, evidently having higher values of SiTH at intermediate mutational burdens. In contrast, in cancer samples the SiTH IQR increased as the total number of SNVs decreased (Figure 3B) while normal samples seem to show abnormally large IQR values compared with tumors with the same number of mutations. On the other hand, cell line samples have smaller IQR values than tumors with the same number of mutations. We should notice that the largest difference with Overall SiTH values calculated from simulated data happens for both low and large mutational burdens, while IQR values are larger across the board, except for tumor cell lines, which seem to generally be more consistently peaked (high Overall SiTH and low SiTH IQR). All this likely reflects the increased contribution of additional mutational processes, beyond SIM, that contribute to increased

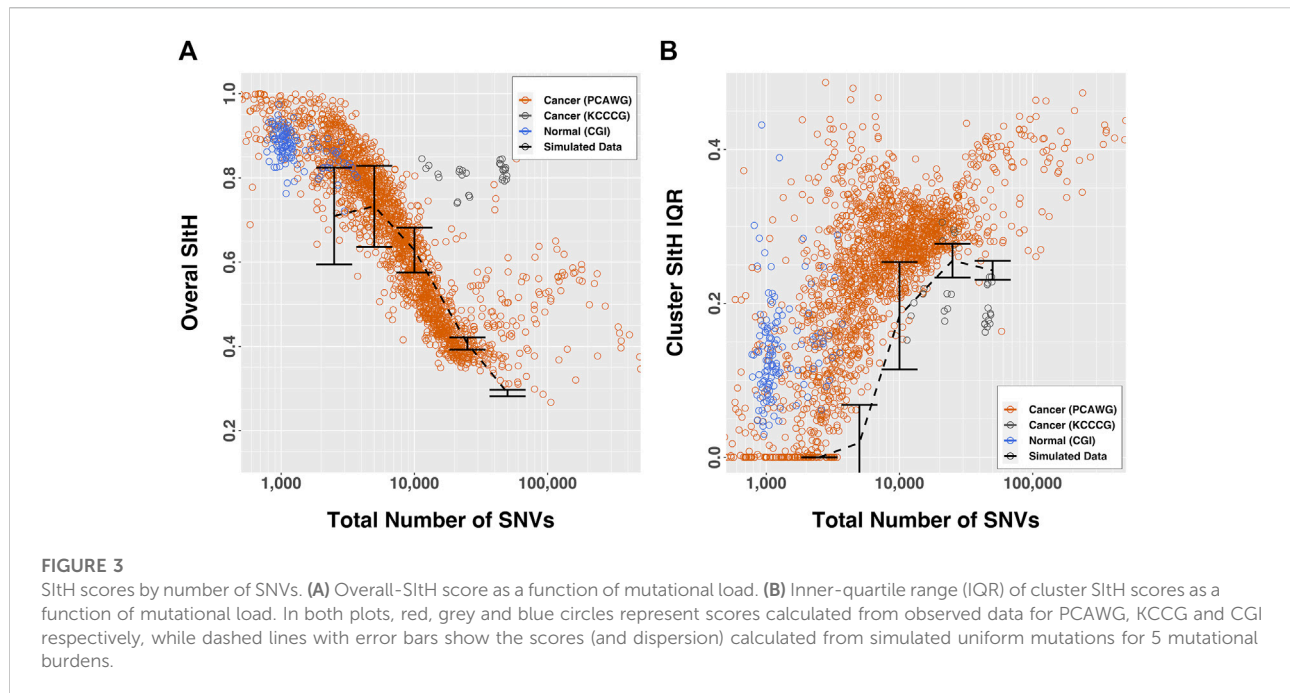
mutational burdens, in addition to reflecting greater intratumor heterogeneity. SiTH score varies by tissue type, with cancer that are characterized by high mutational burdens, like melanoma, showing lower SiTH scores (Figure 4 and Figure 5A).

In the study by Cipponi et al. (2020), cells from three cell-lines were put under strong selective pressure - tunicamycin resulting in endoplasmic reticulum stress and an unfolded-protein response, vemurafenib which inhibits BRAF signaling, and FRAP1 knockdown, which affects mTOR signaling. Selection was carried out under continuous pharmacological stress. Single-cell clones grown out from surviving cells, along with parental controls that spent the same time in culture but were not subject to selection, were expanded and sequenced by WGS (Cipponi et al., 2020). We analyzed the SNVs that were unique to each condition, representing those that arose as either a result of selection or ongoing instability in tissue culture, for SIM through computing SiTH scores and SiTH IQR. Like the data from tumors, we observed clustering among SNVs that resulted in positive SiTH scores in both parental and selected lines. The selected lines were characterized by SiTH scores that were similar to those found in the parental lines event, though the number of SNVs in the selected lines was approximately half those present in the parental lines (Figure 6A). SiTH IQR was generally increased in the selected lines, and the diversity of SiTH IQR between replicates was also increased (Figure 6E). Selection followed by single-cell cloning resulted in clones with different mutational histories after vemurafenib and tunicamycin induced SIM. This is reflected in the SiTH IQR becoming larger for both these conditions and more diverse across clones, as is expected with new rounds of SIM induced through selection. In contrast, the FRAP1 knockdown prohibits the induction of SIM, and therefore individual clones are similar in their cluster formation because no new clusters are introduced during selection, which is reflected in the SiTH IQR becoming smaller (Figure 6E).

TLS drives sharply peaked clusters

Many of the TLS polymerases, as well as APOBEC and AID, have both sequence specific context as well as characteristic mutational profiles (Table 1). We looked at the sequence context of all SNVs identified in both normal and cancer samples to tease out potential mutational mechanisms that contribute to the shape of the clusters we observed. We must note that this analysis is conservative because we only attributed those contexts for which the given enzyme is the only one to fulfill both the wild-type sequence and the resulting mutation; therefore, this analysis under-estimates the role of each mechanism in the generation of clusters.

We computed overall SiTH scores using just those SNVs assigned to a particular mechanism. We then assessed how well



these scores recapitulated the overall SiTH score. As can be seen in [Figures 5B–D](#) and [Figures 6B–D](#), in both the data from the PCAWG tumors as well as the cell lines, SNVs that can be attributed uniquely to the activity of TLS polymerases show a near perfect linear relationship with the overall SiTH score, while those attributable to APOBEC or AID do not. This consistent with the proposed mechanism of SIM driven by TLS activity.

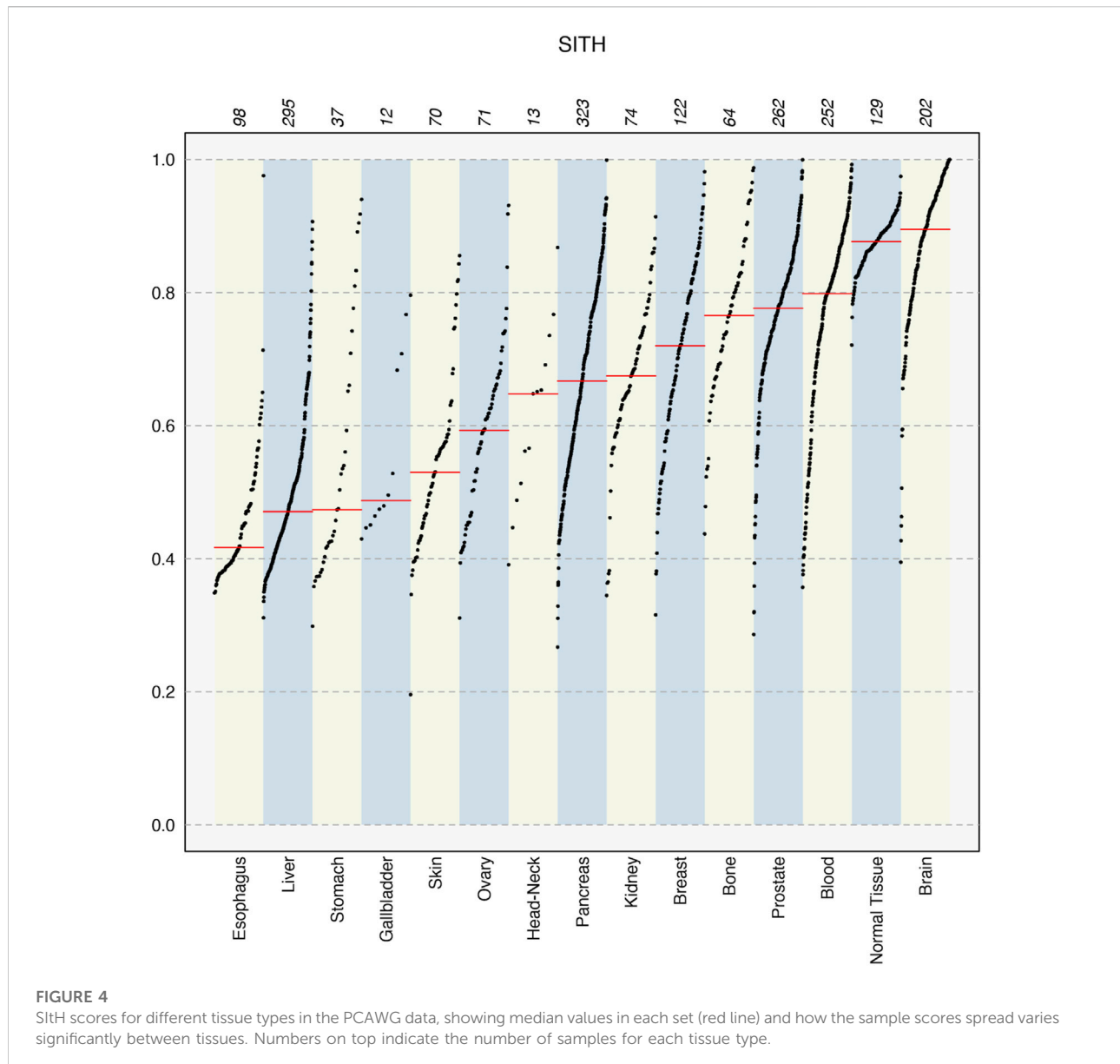
We analyzed the SNV calls from clusters for nucleotide substitution pattern enrichment in clusters ([Figure 7](#)). Normal non-inherited mutations were enriched in A>C, C>G, G>T, T > C, A>G, and G>T and depleted in G>A, G>C, and T>G. These changes are consistent with the activity of the TLS polymerases, Pol- η , Pol- ι , and Pol- θ . In contrast, clusters in cancer samples were enriched for C>T and G>A but excluded all other changes. This is consistent with Pol- κ activity driving the G>A, while the already recognized activity of APOBEC, AID and Pol- η likely drive the C>T enrichment.

In the cell lines, as a general result, both tunicamycin and vemurafenib induced less depletion or enrichment when compared to their respective parental lines. This effect was not found in the FRAP1 knockdown ([Table 2](#); [Supplementary Figures S5–S7](#)). Tunicamycin selection resulted in clusters that were enriched in G>A, A>G, T>C, and C>T and depleted in A>C, A>T, C>G, G>T, C>A, and T>A. There was no enrichment in G>C and T>G. The parental line showed similar enrichment patterns except it was enriched in C>G and G>C. Vemurafenib selection demonstrated clusters enriched in G>A, T>C, A>G, and C>T. SNVs in clusters were depleted in A>C, A>T, G>T, C>A, and T>A variants. The parental line showed a similar enrichment and depletion patterns except for C>G and G>C changes. Finally, in the FRAP1 knockdown, SNVs in

clusters were enriched in G>A, T>C, and C>T changes. They were depleted in A>C, A>T, C>G, G>T, C>A, T>A, and T>G. The pattern the parental line was extremely similar except for the C>G, which showed enrichment in the parental cell line but marginal depletion in the knockdown line.

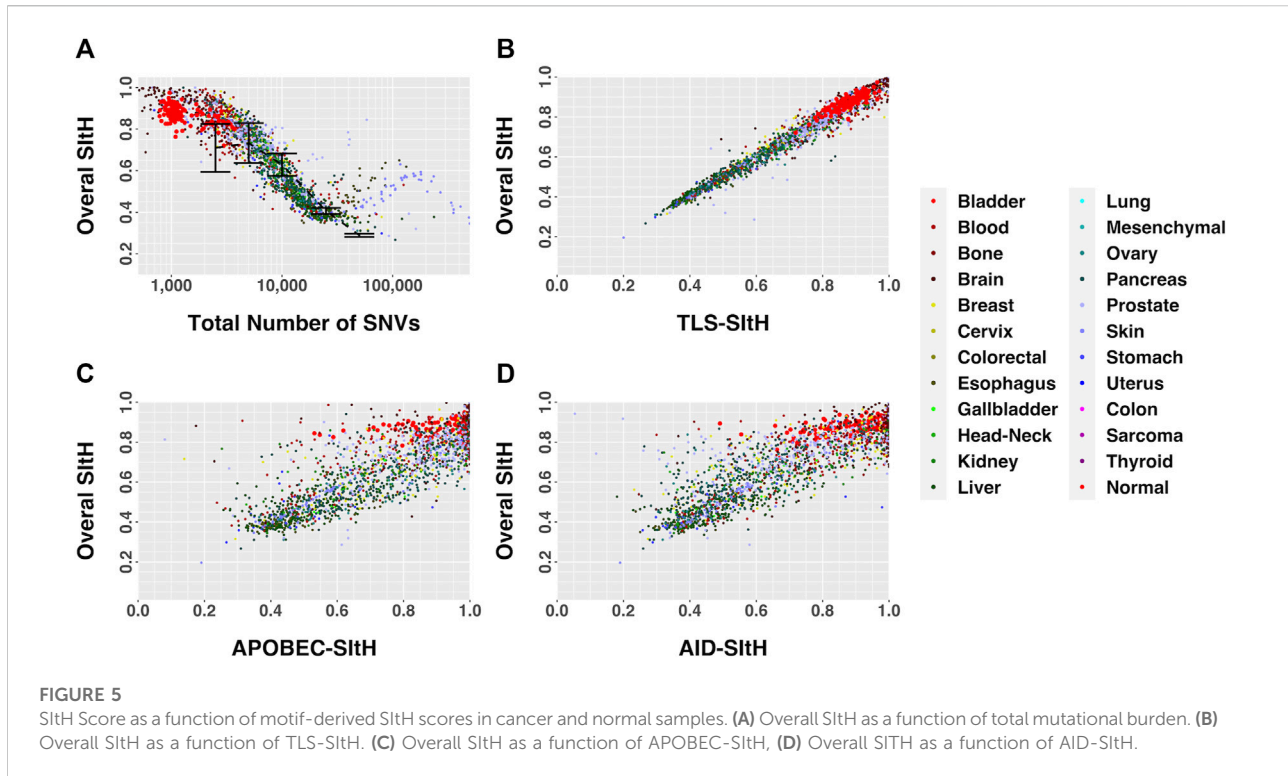
Mutational signature analysis and SIM

To further establish mutational mechanisms underlying SIM, we performed a mutational signature analysis of the SNVs found in clusters and used in our SiTH score calculations. In the PCAWG data, we initially extracted 30 signatures to compare to COSMIC v2 signatures, but discovered that 4 signatures were redundant in the information they contained. We repeated the analysis and identified 26 mutational signatures from clustered mutations. These were compared to both the COSMIC v2 signatures as well as the 82 reference signatures from the SIGNAL project ([Figure 8](#)) ([Alexandrov et al., 2013](#); [Degasperi et al., 2022](#)). Six signatures had a cosine similarity of 0.85 or greater to the COSMICv2 pattern: signature 1-like, signature 7-like, signature 13-like, signature 17-like, signature 26-like, and signature 28 like. In comparison with the 82 reference signatures, an additional 8 signatures have cosine similarities of greater than 0.8 ([Supplementary Table S5](#)). Most of the signatures have unknown etiologies and are characterized by T>C and C>T changes. The activity of Pol- κ and Pol- η results in G>A and A>G alterations. Because of the way mutational signatures are handled computationally, calls that are G>A or A>G on the



reference strand would be assigned to C>T and T>C respectively. Thus, the activity of TLS polymerases, particularly Pol- κ , would be expected to lead to an abundance of T>C and C>T that could be teased out by looking at whether most of the calls contributing to those signatures come from the reference call (e.g., originally C>T) or from a reference call converted to its complement (e.g., originally G>A). We investigated this by running a transcriptional strand bias analysis, setting the annotation up such that G and A reference calls would be annotated as being on the “untranscribed” strand. We see that the abundance of C>T changes present in clustered SNVs are derived from G>A changes and likely represent the activity of Pol- κ as part of SIM (Figure 9, Supplementary Figures S8, S9).

We further explored the relationship between SITH score, SITH IQR, and the mutational signatures we identified in the clustered SNVs to determine whether both types of scores capture the same information. To do this, we ran a Spearman correlation on the signature contribution across samples with either SITH score or SITH IQR (Table 3). SITH score is inversely correlated with all identified signatures, confirming that SITH score is measuring a different aspect, geometric distribution, of the mutational processes taking place in the tumor that the mutational signatures are not detecting. In contrast, the SITH IQR was positively correlated with mutational signatures, further supporting that this score measures intratumor heterogeneity generated by multiple mutational processes.



Survival analysis

A key characteristic of SNV clusters that result from SIM mechanisms is a decay in the frequency of incidental SNVs as a function of distance from the DSB that triggered error-prone repair response (Shee et al., 2012). We postulated that a more positive overall SItH score reflects a greater contribution of SIM to the mutational landscape of the tumor. Therefore, SItH provides a measure of the evolutionary response, or the adaptive capacity, of a tumor to a source of stress, such as chemotherapy.

To determine the relationship between SItH scores and clinical outcome, we conducted Cox proportional hazard analysis of the overall SItH score as well as the IQR of the cluster SItH. We used the tissue of origin to account for differences in mutational load, age of onset, and general overall survival. The model for overall SItH is specified as follows:

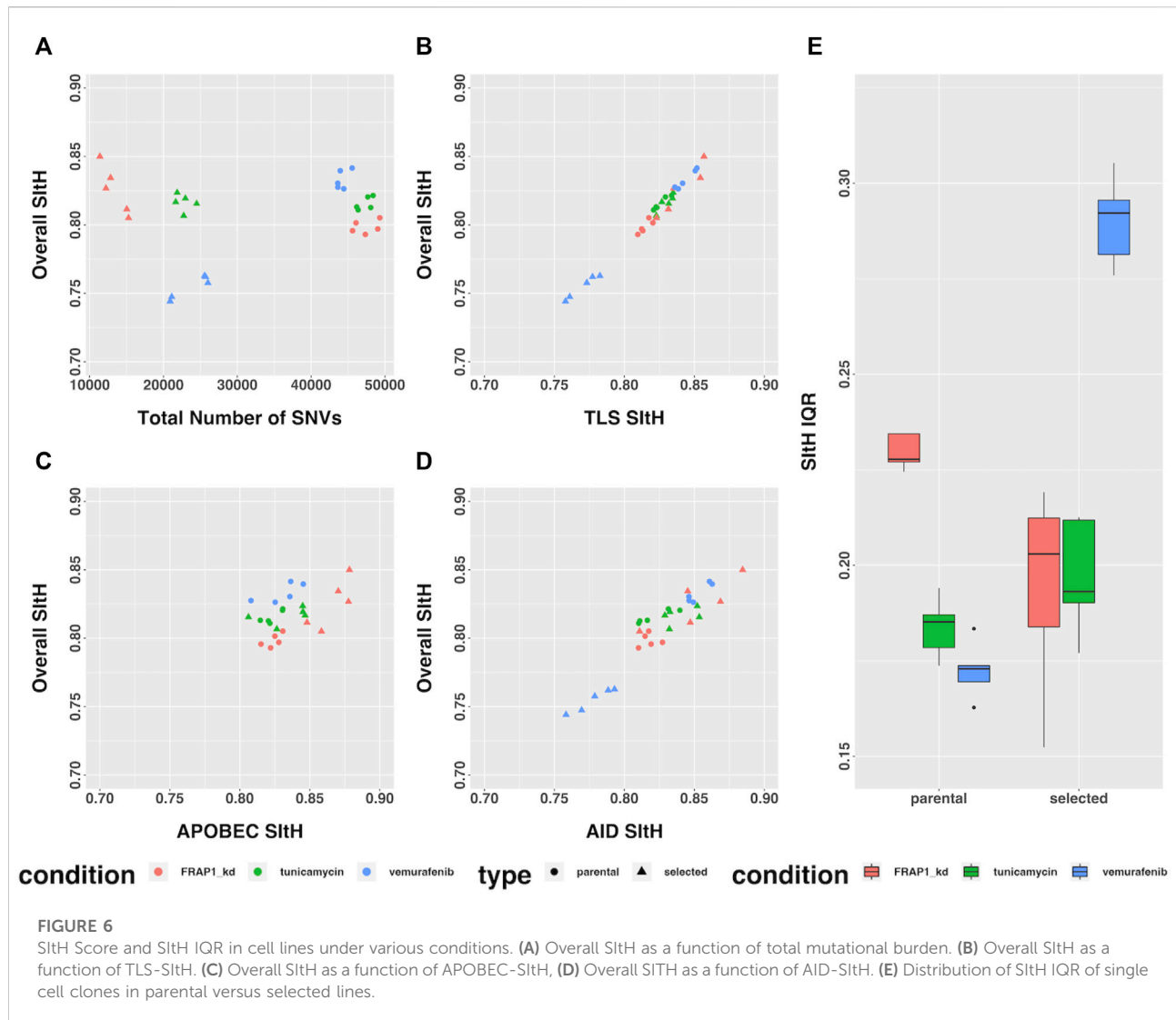
$$\text{Overall Survival} \sim \text{SItH} + \text{multiple.tumor} + \text{is.Max.SItH} + \text{strata(Organ)}$$

where the data analyzed were either primary tumors or the group of metastases and recurrences. Similarly, for IQR of the cluster SItH the model is:

$$\text{Overall Survival} \sim \text{SItH IQR} + \text{multiple.tumor} + \text{is.Max.SItH} + \text{strata(Organ)}$$

After controlling for organ site and multiple tumor status, we found that overall SItH scores predict patient survival, with different effects depending on whether the sample was a primary tumor or from a metastasis or recurrence. For primary tumors, more positive overall SItH scores predicted better patient survival (Cox Proportional Hazard Regression (CPHR), Hazard Ratio (HR) = 0.4516, 95% CI: 0.2274–0.8968, $p = 0.0231$). However, when the recurrences and metastatic tumors were considered as a group, more positive overall SItH scores predicted worse survival, with a HR of 14.84 (CPHR, 95% CI: 1.934–113.876, $p = 0.00947$). These results suggest that in the context of a primary tumors SIM can lead to tumors being too mutable and evolving in ways that do not promote survival. In contrast, in metastases and recurrent tumors, the strong selective pressure of therapy or distant organ location selects for SIM, leading to higher SItH scores, indicative of a stronger contribution of SIM to the mutational landscape, being associated with poorer survival.

In looking at the diversity of SItH scores on a cluster basis, the type of tumor sample was no longer relevant. Wider IQR of cluster-level SItH scores was associated with worse survival, with a HR of 5.744 (CPHR, 95% CI: 1.824–18.09, $p = 0.00283$). We then examined whether there was a difference in survival between patients with SItH IQRs above or below the median SItH IQR, as clinical translation will likely require creating a



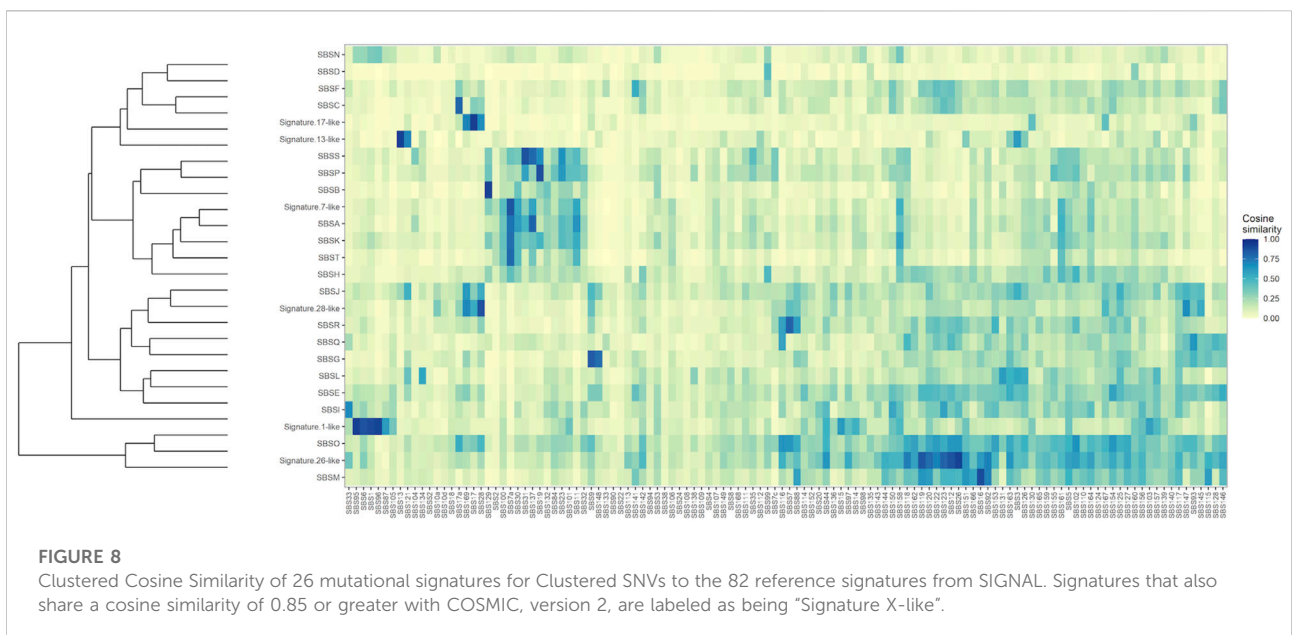
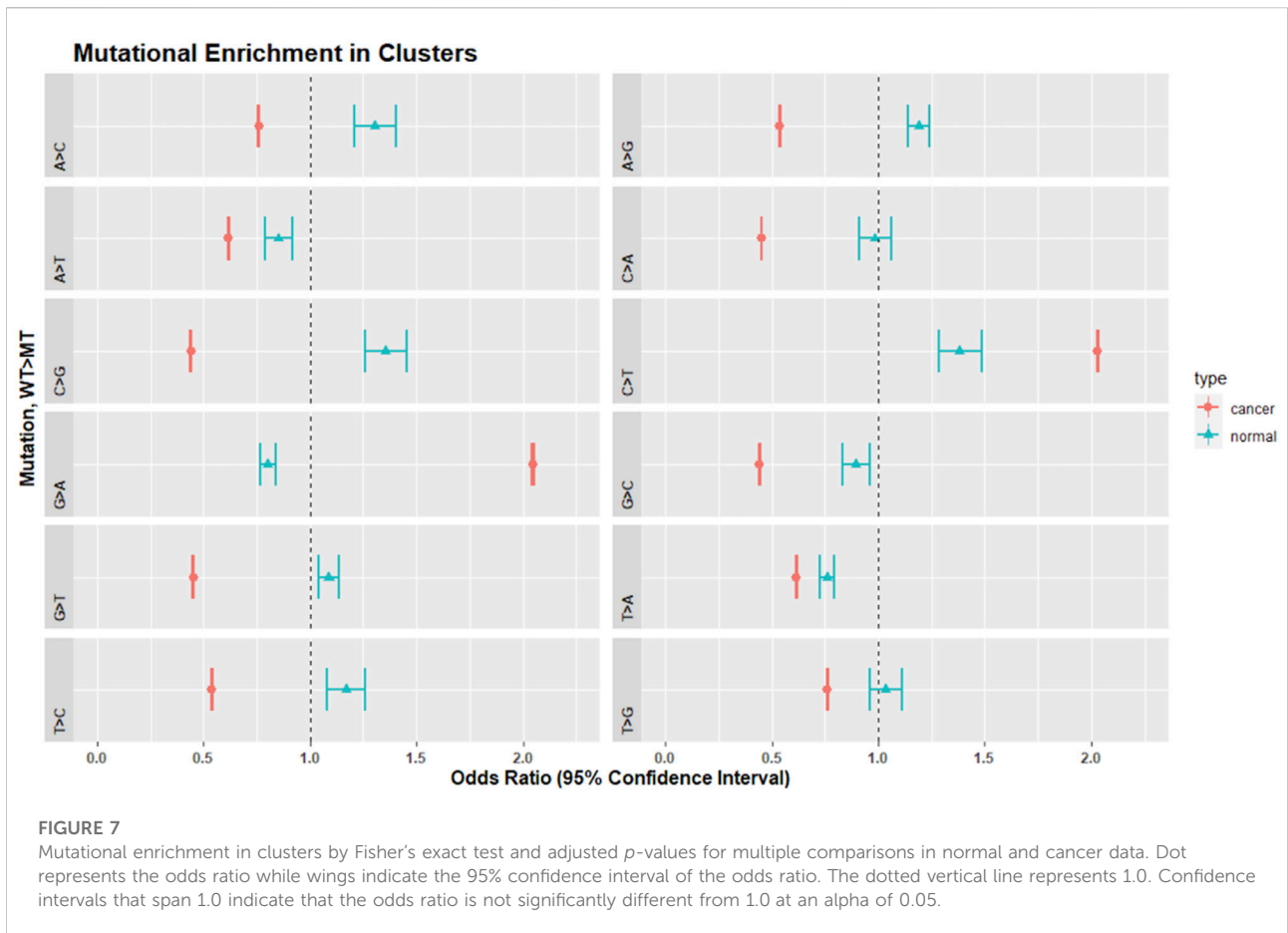
cut-off value above which one would predict poor prognosis. As seen in Figure 10, there is a significant difference in survival, even after accounting for the baseline differences in survival by tissue of origin (CPHR, HR = 1.37, 95% CI: 1.10–1.7, $p = 0.006$).

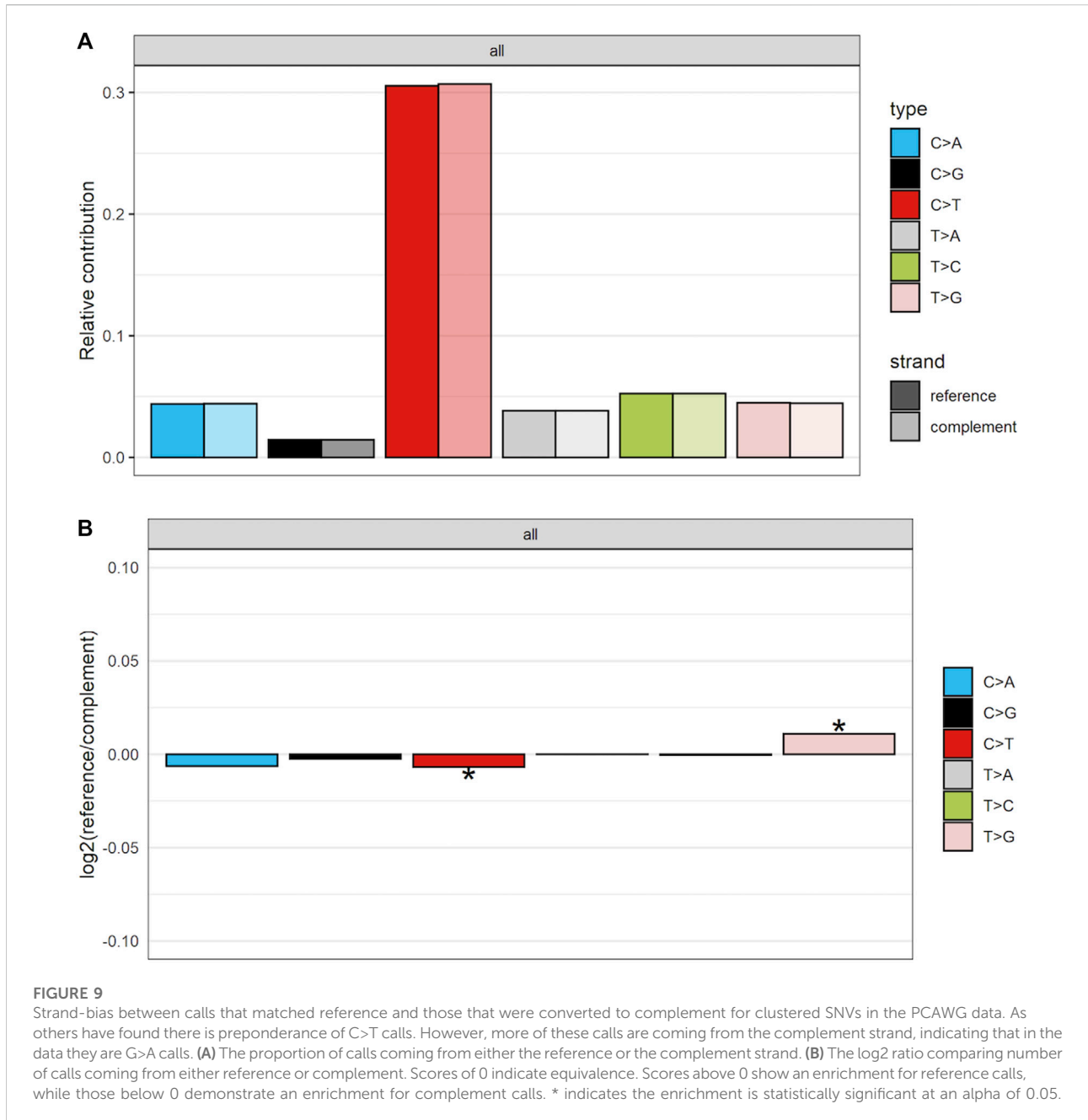
Discussion

Our study provides evidence that a signature of stress-induced mutagenesis, characterized by clustering of SNVs with a defined cluster geometry, is widespread across multiple cancer types. Both the strength of SIM and the diversity of mutational processes within a tumor are expected to impact disease outcome (Andor et al., 2016). Our results show an association of both overall cluster shape (overall SiTH) and increased cluster shape diversity (SiTH IQR) with patient survival. We submit that SiTH IQR predominantly represents the amount of time SIM has been active during carcinogenesis

and clonal diversification, while overall SiTH represents the ratio of the intensity of SIM relative to other mutational processes. This is supported by the behavior of the score in experimental models of stress-induced mutation. Our work shows that an increase in mutational load leads to increases in both cluster size and the percentage of SNVs involved in clusters, but only up to a point. In tumors with high mutational burdens, the number of clusters, the genomic distance covered by clusters, and the number of SNVs contained within a cluster all level out. This implies that under high mutational burden the variations in mutation density across the genome flatten out, likely due to alterations in DNA repair pathways, such as a loss of mismatch repair (Campbell et al., 2017; Supek and Lehner, 2017) that obscures the detection of clusters.

The influence of intra-tumor diversity on clinical outcome is an area of active investigation. Evidence from measures of clonal diversity and copy number diversity are associated with both worse outcome and therapeutic response (Andor et al., 2016;





Dagogo-Jack and Shaw, 2017; Davoli et al., 2017; Roh et al., 2017; Ben-David and Amon, 2019; Turajlic et al., 2019). Cancer must balance the introduction of genomic rearrangements that contribute to cellular diversity with a sufficient level of genome stability to avoid a genomic error catastrophe. Our results are consistent with this idea, in that large positive overall SItH scores in primary tumor samples are associated with better patient survival. The SItH IQR represents a measure of mutational heterogeneity that ties intra-tumor diversity to an underlying evolutionarily conserved process in response to cellular stress. In other words, the SItH IQR is a measure of

the heterogeneity of adaptive strategies within a patient. This diversity manifests as a broad ensemble of mutational cluster shapes within a tumor, driven by the heterogeneity in mutational processes to generate genomic diversification. This in turn increases the substrates available for broad phenotypic plasticity, including transcriptional responses. Such responses have been shown to be important in the rapid acquisition of resistance to Doxorubicin (Wu et al., 2015). In this case high diversity results in a direct survival advantage for the tumor, allowing it to respond to a wider range of stresses and leading to poorer outcomes for patients.

TABLE 2 Patterns of base-change enrichment observed in cell-lines.

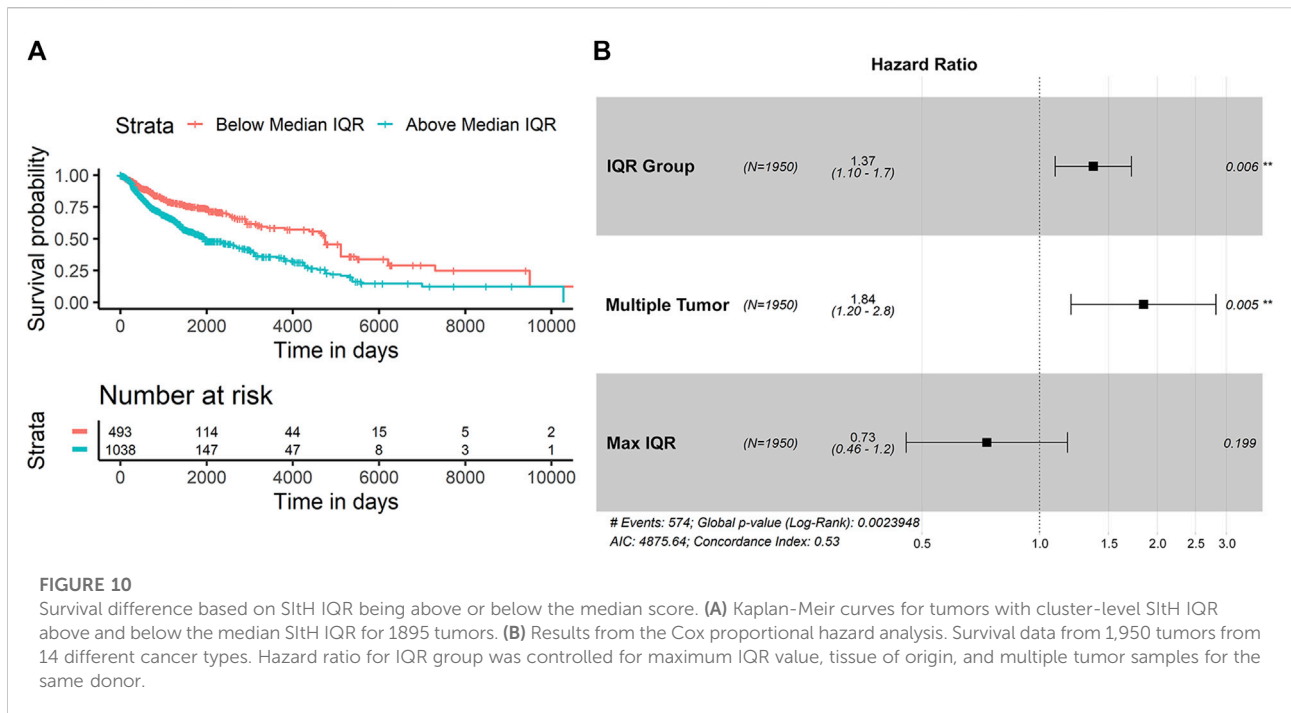
Base change	Tunicamycin	Parental	Vermurafenib	Parental	FRAP1	Parental
A>C	depleted	depleted	depleted	depleted	depleted	depleted
A>G	enriched	enriched	enriched	enriched		
A>T	depleted	depleted	depleted	depleted	depleted	depleted
C>A	depleted	depleted	depleted	depleted	depleted	depleted
C>G	depleted	enriched			depleted	enriched
C>T	enriched	enriched	enriched	enriched	enriched	enriched
G>A	enriched	enriched	enriched	enriched	enriched	enriched
G>C		enriched				
G>T	depleted	depleted	depleted	depleted	depleted	depleted
T>A	depleted	depleted	depleted	depleted	depleted	depleted
T>C	enriched	enriched	enriched	enriched	enriched	enriched
T>G					depleted	depleted

TABLE 3 Spearman correlation between signature contributions across samples and corresponding SITH scores and SITH IQR scores. Correlation factor (ρ), p -value and adjusted p -value are shown in each case.

Signature	SITH_rho	SITH_p	SITH_adj_p	IQR_rho	IQR_p	IQR_adj_p
SBSA	-0.46	9.39E-101	3.62E-100	0.33	6.60E-49	2.55E-48
SBSB	-0.74	0.00E+00	0.00E+00	0.53	1.47E-138	7.35E-137
Signature.1-like	-0.66	1.24E-242	1.25E-241	0.49	2.97E-115	4.95E-114
SBSC	-0.67	1.31E-258	1.88E-257	0.44	2.55E-92	1.97E-91
Signature.28-like	-0.53	3.61E-143	1.64E-142	0.37	1.46E-61	6.08E-61
SBSD	-0.65	7.96E-237	7.25E-236	0.46	6.25E-101	5.70E-100
SBSE	-0.53	9.94E-141	4.33E-140	0.38	4.55E-65	2.07E-64
Signature.26-like	-0.64	3.43E-221	2.65E-220	0.39	6.99E-71	3.50E-70
SBSF	-0.64	1.79E-221	1.50E-220	0.42	5.86E-84	3.67E-83
SBSG	-0.66	8.22E-248	9.16E-247	0.48	4.76E-108	5.97E-107
Signature.13-like	-0.69	4.57E-274	7.63E-273	0.58	2.49E-170	2.50E-168
SBSH	-0.58	1.48E-176	8.25E-176	0.43	1.25E-86	8.91E-86
SBSI	-0.52	1.82E-133	7.61E-133	0.33	5.24E-49	2.10E-48
Signature.17-like	-0.74	0.00E+00	0.00E+00	0.52	2.99E-134	9.99E-133
SBSJ	-0.56	6.92E-163	3.47E-162	0.41	4.95E-80	2.92E-79
SBSK	-0.61	2.48E-199	1.78E-198	0.45	4.09E-96	3.41E-95
SBSL	-0.73	0.00E+00	0.00E+00	0.49	7.95E-116	1.59E-114
SBSM	-0.73	0.00E+00	0.00E+00	0.47	2.79E-106	3.11E-105
SBSN	-0.60	7.60E-192	5.08E-191	0.43	4.66E-85	3.11E-84
SBSO	-0.57	1.72E-167	9.05E-167	0.40	1.14E-75	6.03E-75
S BSP	-0.58	6.32E-178	3.96E-177	0.37	3.54E-63	1.54E-62
Signature.7-like	-0.58	4.22E-177	2.49E-176	0.47	1.01E-103	1.01E-102
SBSQ	-0.71	4.34E-300	8.69E-299	0.48	4.77E-110	6.82E-109
SBSR	-0.67	8.21E-256	1.03E-254	0.49	1.13E-117	2.82E-116
SBS S	-0.55	5.15E-155	2.46E-154	0.41	3.99E-79	2.22E-78
SBST	-0.46	1.75E-104	7.01E-104	0.39	1.21E-70	5.76E-70

Others have proposed mechanisms for clustered mutations in cancer (Lada et al., 2012; Roberts et al., 2012; Burns et al., 2013; Roberts et al., 2013; Taylor et al., 2013; Supek and Lehner, 2017).

In particular, Supek and Lehner showed that Pol- η , a TLS polymerase closely related to Pol IV, is involved in the generation of clustered mutations that preferentially locate at



the 3'-end of active genes (Supek and Lehner, 2015). Although our method uses a broader definition of mutational clusters than previously proposed (Shee et al., 2012; Fitzgerald et al., 2017) we were able to confirm this key finding.

An open question that remains is whether the clusters we, and others, have detected arose from singular events reflective of bursts of mutational activity, or were accumulated over time. The latter scenario would identify distinct regions of the genome prone to mutation. Measuring allele fraction has been suggested as one way to address this question. However, the limited precision of most allele fraction measurements prevents the accurate discrimination of varying degrees of heterogeneity across a tumor. For example, the 95/95 binomial tolerance interval for a true allele fraction of 0.5 at a read depth of 60x ranges from 0.25 to 0.75 (Supplementary Material). This interval represents the bounds in which we are 95% confident that 95% of the measurements of a true allele fraction of 0.5 would lie. If we have a cluster where the allele fractions of multiple SNVs all fall within this range, we cannot rule out whether these represent a true allele fraction of 0.5 and therefore all come from the same event. Experimental evidence in mammalian systems leading to cluster formation is necessary to answer this question. This is an important study to pursue as the strategies one might propose for influencing mutational patterns with impact on clinical outcomes will depend on whether the target is the mutational process itself or the regions of the genome being acted upon by the mutational process.

In conclusion, cancer is notorious for outsmarting physicians. To make progress, we need to factor in how

cancer cells evolve and adapt in the face of the challenges of medical treatment. A deeper understanding of the mechanisms of mutation and adaptation in cancer is therefore an essential pre-requisite for improving patient outcomes. Stress-induced mutagenesis, an ancient and evolutionarily conserved adaptive mutation mechanism well-characterized in *E. coli*, comprises some part of the genomic instability seen in cancer and contributes to the ability of the tumor to evolve resistance to therapy (Fitzgerald et al., 2017). We have described a way to quantify this biological response and shown that SIM has a strong association with poor prognosis.

Further investigations into the process of SIM in cancer should lead to better patient outcomes by giving clinicians a metric by which they can tailor treatments to regulate tumor progression and minimize the risk of triggering an aggressive evolutionary response.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: 1- PCAWG cancer data: <https://dcc.icgc.org/pcawg> 2- Normal tissue variant data from the Complete Genomics Indices database in the 1000 Genomes Project (release 20130502): ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/cgi_variant_calls/ 3- Cell-line data: NCBI's BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject>) accession number PRJNA623123. Processed data

sets along with R code can be found at <https://github.com/kjbussey/SItH>.

Author contributions

LC and KB contributed to the conception, design, and analysis of the study. LC performed clustering and SItH analyses. KB performed statistics and mutational analyses. CV provided supervision with bioinformatics and funding acquisition. LC and KB wrote the manuscript. All authors contributed to manuscript revisions, read, and approved the submitted version. LC and KB contributed equally to this work.

Funding

This work was initially supported by NIH grant U54CA143862 (<https://projectreporter.nih.gov>) and NantOmics, LLC. The NIH had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript. NantOmics provided support in the form of salaries for authors but did not have any additional role in the study design, data collection, and analysis, decision to publish, or preparation of the manuscript. LC acknowledges support from the Arizona Cancer Evolution Center in the Biodesign Center for Biocomputing, Security, and Society at Arizona State University during the preparation of the manuscript, and by the National Cancer Institute of the National Institutes of Health under Award Number NIH-U54CA217376. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. KB acknowledges funding and support from Midwestern University.

References

- Akdemir, K. C., Le, V. T., Chandran, S., Li, Y., Verhaak, R. G., Beroukhim, R., et al. (2020). Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat. Genet.* 52, 294–305. doi:10.1038/s41588-019-0564-y
- Akdemir, K. C., Le, V. T., Kim, J. M., Killcoyne, S., King, D. A., Lin, Y. P., et al. (2020). Somatic mutation distributions in cancer genomes vary with three-dimensional chromatin structure. *Nat. Genet.* 52 (11), 1178–1188. doi:10.1038/s41588-020-0708-0
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500 (7463), 415–421.
- Andor, N., Graham, T. A., Jansen, M., Xia, L. C., Aktipis, C. A., Petritsch, C., et al. (2016). Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat. Med.* 22, 105–113. doi:10.1038/nm.3984
- Ben-David, U., and Amon, A. (2019). Context is everything: Aneuploidy in cancer. *Nat. Rev. Genet.* 21 (1), 44–62. doi:10.1038/s41576-019-0171-x
- Buisson, R., Langenbucher, A., Bowen, D., Kwan, E. E., Benes, C. H., Zou, L., et al. (2019). Passenger hotspot mutations in cancer driven by APOBEC3A and mesoscale genomic features. *Science* 364 (6447), eaaw2872. doi:10.1126/science.aaw2872
- Burns, M. B., Temiz, N. A., and Harris, R. S. (2013). Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat. Genet.* 45, 977–983. doi:10.1038/ng.2701
- Campbell, B. B., Galati, M. A., Stone, S. C., Riemenschneider, A. N., Edwards, M., Sudhama, S., et al. (2021). Mutations in the RAS/MAPK pathway drive replication repair-deficient hypermutated tumors and confer sensitivity to MEK inhibition. *Cancer Dis.* 11, 1454–1467. doi:10.1158/2159-8290.CD-20-1050
- Campbell, B. B., Light, N., Fabrizio, D., Zatzman, M., Fuligni, F., de Borja, R., et al. (2017). Comprehensive analysis of hypermutation in human cancer. *Cell* 171 (5), 1042–1056. e10. doi:10.1016/j.cell.2017.09.048
- Chen, Z., Wen, W., Bao, J., Kuhs, K. L., Cai, Q., Long, J., et al. (2019). Integrative genomic analyses of APOBEC-mutational signature, expression and germline deletion of APOBEC3 genes, and immunogenicity in multiple cancer types. *BMC Med. Genomics* 12 (1), 131. doi:10.1186/s12920-019-0579-3
- Cipponi, A., Goode, D. L., Bedo, J., McCabe, M. J., Pajic, M., Croucher, D. R., et al. (2020). MTOR signaling orchestrates stress-induced mutagenesis, facilitating adaptive evolution in cancer. *Science* 368 (6495), 1127–1131. doi:10.1126/science.aau8768

Acknowledgments

We thank Paul Davies, Adam Orr, Charles Lineweaver, Susan Rosenberg, Julia A. Bos and Robert Austin for their insightful discussions into the role of genomic instability and DNA repair in cancer. We thank Sabrina Leung, Terry Christenson and Jessica Pirkle Callan for their technical writing support. We also acknowledge the work of the clinical collaborators, data analysis teams, and funders generating the WGS data in the Pan-Cancer Analysis Working Group of the International Cancer Genome Consortium and the Complete Genomics Indices database in the 1000 Genome Project database.

Conflict of interest

LC, CV, and KB were employed by the company NantOmics, LLC.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.932763/full#supplementary-material>

- Cisneros, L., Bussey, K. J., Orr, A. J., Miočević, M., Lineweaver, C. H., and Davies, P. C. (2017). Ancient genes establish stress-induced mutation as a hallmark of cancer. *PLoS ONE* 12, e0176258. doi:10.1371/journal.pone.0176258
- Cortés-Ciriano, I., Lee, J. J. K., Xi, R., Jain, D., Jung, Y. L., Yang, L., et al. (2020). Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat. Genet.* 52 (3), 331–341. doi:10.1038/s41588-019-0576-7
- Dagogo-Jack, I., and Shaw, A. T. (2017). Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* 15 (2), 81–94. doi:10.1038/nrclinonc.2017.166
- Davoli, T., Uno, H., Wooten, E. C., and Elledge, S. J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* 355, eaaf8399. doi:10.1126/science.aaf8399
- Degasperi, A., Zou, X., Dias Amarante, T., Martinez-Martinez, A., Koh, G. C. C., Dias, J. M. L., et al. (2022). Substitution mutational signatures in whole-genome-sequenced cancers in the UK population. *Science* 376 (6591), ab9283. doi:10.1126/science.ab9283
- Drake, J. W. (2007). Mutations in clusters and showers. *Proc. Natl. Acad. Sci. U. S. A.* 104 (20), 8203–8204. doi:10.1073/pnas.0703089104
- Fitzgerald, D. M., Hastings, P. J., and Sm, R. (2017). Stress-induced mutagenesis: Implications in cancer and drug resistance. *Annu. Rev. Cancer Biol.* 1, 119–140. doi:10.1146/annurev-cancerbio-050216-121919
- Forment, J. V., and O'Connor, M. J. (2018). Targeting the replication stress response in cancer. *Pharmacol. Ther.* 188, 155–167. doi:10.1016/j.pharmthera.2018.03.005
- Foster, P. I. (2007). Stress-induced mutagenesis in bacteria. *Crit. Rev. Biochem. Mol. Biol.* 42, 373–397. doi:10.1080/10409230701648494
- Goodman, M. F. (2002). Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* 71 (1), 17–50. doi:10.1146/annurev.biochem.71.083101.124707
- Goodman, M. F., and Woodgate, R. (2013). Translesion DNA polymerases. *Cold Spring Harb. Perspect. Biol.* 5 (10), a010363. doi:10.1101/cshperspect.a010363
- Janion, C. (2008). Inducible SOS response system of DNA repair and mutagenesis in *Escherichia coli*. *Int. J. Biol. Sci.* 4, 338–344. doi:10.7150/ijbs.4.338
- Kamburov, A., Lawrence, M. S., Polak, P., Leshchiner, I., Lage, K., Golub, T. R., et al. (2015). Comprehensive assessment of cancer missense mutation clustering in protein structures. *Proc. Natl. Acad. Sci.* 112 (40), 1. doi:10.1073/pnas.1516373112
- Lada, A. G., Dhar, A., Boissy, R. J., Hirano, M., Rubel, A. A., Rogozin, I. B., et al. (2012). AID/APOBEC cytosine deaminase induces genome-wide kataegis. *Biol. Direct* 7 (1), 47–7. doi:10.1186/1745-6150-7-47
- Libbrecht, M. W., Ay, F., Hoffman, M. M., Gilbert, D. M., Bilmes, J. A., and Noble, W. S. (2015). Joint annotation of chromatin state and chromatin conformation reveals relationships among domain types and identifies domains of cell-type-specific expression. *Genome Res.* 25 (4), 544–557. doi:10.1101/gr.184341.114
- Livneh, Z. (2001). DNA damage control by novel DNA polymerases: Translesion replication and mutagenesis. *J. Biol. Chem.* 276 (28), 25639–25642. doi:10.1074/jbc.R100019200
- Ma, J., Setton, J., Lee, N. Y., Riaz, N., and Powell, S. N. (2018). The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. *Nat. Commun.* 9 (1), 3292. doi:10.1038/s41467-018-05228-y
- Martincorena, I., and Campbell, P. J. (2015). Somatic mutation in cancer and normal cells. *Science* 349 (6255), 1483–1489. doi:10.1126/science.aab4082
- McKenzie, G. J., Harris, R. S., Lee, P. L., and Sm, R. (2000). The SOS response regulates adaptive mutation. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6646–6651. doi:10.1073/pnas.120161797
- McKenzie, G. J., Lee, P. L., Lombardo, M. J., Hastings, P. J., and Rosenberg, S. M. (2001). SOS mutator DNA polymerase IV functions in adaptive mutation and not adaptive amplification. *Mol. Cell* 7 (3), 571–579. doi:10.1016/s1097-2765(01)00204-0
- Murai, J. (2017). Targeting DNA repair and replication stress in the treatment of ovarian cancer. *Int. J. Clin. Oncol.* 22 (4), 619–628. doi:10.1007/s10147-017-1145-7
- Nik-Zainal, S., Alexandrov, L. B., Wedge, D. C., Van Loo, P., Greenman, C. D., Raine, K., et al. (2012). Mutational processes molding the genomes of 21 breast cancers. *Cell* 149 (5), 979–993. doi:10.1016/j.cell.2012.04.024
- Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., et al. (2016). Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534 (7605), 47–54. doi:10.1038/nature17676
- Ram, Y., and Hadany, L. (2019). Evolution of stress-induced mutagenesis in the presence of horizontal gene transfer. *Am. Nat.* 194 (1), 73–89. doi:10.1086/703457
- Rheinbay, E., Nielsen, M. M., Abascal, F., Wala, J. A., Shapira, O., Tiao, G., et al. (2020). Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* 578 (7793), 102–111. doi:10.1038/s41586-020-1965-x
- Roberts, S. A., Lawrence, M. S., Klimczak, L. J., Grimm, S. A., Fargo, D., Stojanov, P., et al. (2013). An APOBEC cytosine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* 45, 970–976. doi:10.1038/ng.2702
- Roberts, S. A., Sterling, J., Thompson, C., Harris, S., Mav, D., Shah, R., et al. (2012). Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* 46 (4), 424–435. doi:10.1016/j.molcel.2012.03.030
- Roh, W., Chen, P. L., Reuben, A., Spencer, C. N., Prieto, P. A., Miller, J. P., et al. (2017). Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci. Transl. Med.* 9, eaah3560. doi:10.1126/scitranslmed.aah3560
- Roper, N., Gao, S., Maity, T. K., Banday, A. R., Zhang, X., Venugopalan, A., et al. (2019). APOBEC mutagenesis and copy-number alterations are drivers of proteogenomic tumor evolution and heterogeneity in metastatic thoracic tumors. *Cell Rep.* 26 (10), 2651–2666. doi:10.1016/j.celrep.2019.02.028
- Roschke, A. V., Lababidi, S., Tonon, G., Gehlhaus, K. S., Bussey, K., Weinstein, J. N., et al. (2005). Karyotypic “state” as a potential determinant for anticancer drug discovery. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2964–2969. doi:10.1073/pnas.0405578102
- Roschke, A. V., Stover, K., Tonon, G., Schaffer, A. A., and Ir, K. (2002). Stable karyotypes in epithelial cancer cell lines despite high rates of ongoing structural and numerical chromosomal instability. *Neoplasia* 4, 19–31. doi:10.1038/sj.neo.7900197
- Roschke, A. V., Tonon, G., Gehlhaus, K. S., McTye, N., Bussey, K. J., Lababidi, S., et al. (2003). Karyotypic complexity of the NCI-60 drug-screening panel. *Cancer Res.* 63, 8634–8647.
- Rosenberg, S. M., Shee, C., Frisch, R. L., and Hastings, P. J. (2012). Stress-induced mutation via DNA breaks in *Escherichia coli*: A molecular mechanism with implications for evolution and medicine. *Bioessays* 34 (10), 885–892. doi:10.1002/bies.201200050
- Rosenberg, S. M. (2010). Spontaneous mutation: Real-time in living cells. *Curr. Biol.* 20 (18), R810–R811. doi:10.1016/j.cub.2010.07.031
- Russo, M., Crisafulli, G., Sogari, A., Reilly, N. M., Arena, S., Lamba, S., et al. (2019). Adaptive mutability of colorectal cancers in response to targeted therapies. *Science* 366 (6472), 1473–1480. doi:10.1126/science.aav4474
- Sakořský, C. J., Ayyar, S., Deem, A. K., W-h Clra, G., and Malkova, A. (2015). Translesion polymerases drive microhomology-mediated break-induced replication leading to complex chromosomal rearrangements. *Mol. Cell* 60, 860–872. doi:10.1016/j.molcel.2015.10.041
- Saxton, R. A., and Sabatini, D. M. (2017). mTOR signaling in growth, metabolism, and disease. *Cell* 168 (6), 361–371. doi:10.1016/j.cell.2017.03.035
- Shee, C., Gibson, J. L., Darrow, M. C., Gonzalez, C., and Rosenberg, S. M. (2011). Impact of a stress-inducible switch to mutagenic repair of DNA breaks on mutation in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 108 (33), 13659–13664. doi:10.1073/pnas.1104681108
- Shee, C., Gibson, J. L., and Rosenberg, S. M. (2012). Two mechanisms produce mutation hotspots at DNA breaks in *Escherichia coli*. *Cell Rep.* 2 (4), 714–721. doi:10.1016/j.celrep.2012.08.033
- Shi, M. J., Meng, X. Y., Fontugne, J., Chen, C. L., Radvanyi, F., and Bernard-Pierrot, I. (2020). Identification of new driver and passenger mutations within APOBEC-induced hotspot mutations in bladder cancer. *Genome Med.* 12 (1), 85. doi:10.1186/s13073-020-00781-y
- Supek, F., and Lehner, B. (2017). Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. *Cell* 170 (3), 534–547. doi:10.1016/j.cell.2017.07.003

- Supek, F., and Lehner, B. (2015). Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* 521, 81–84. doi:10.1038/nature14173
- Taylor, B. J., Nik-Zainal, S., Wu, Y. L., Stebbings, L. A., Raine, K., Campbell, P. J., et al. (2013). DNA deaminases induce break-associated mutation showers with implication of APOBEC3B and 3A in breast cancer kataegis. *eLife* 2, e00534. doi:10.7554/eLife.00534
- Turajlic, S., Sottoriva, A., Graham, T., and Swanton, C. (2019). Resolving genetic heterogeneity in cancer. *Nat. Rev. Genet.* 20 (7), 404–416. doi:10.1038/s41576-019-0114-6
- The 1000 Genomes Project Consortium, Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi:10.1038/nature15393
- Ubhi, T., and Brown, G. W. (2019). Exploiting DNA replication stress for cancer treatment. *Cancer Res.* 79 (8), 1730–1739. doi:10.1158/0008-5472.CAN-18-3631
- Wang, J., Gonzalez, K. D., Scaringe, W. A., Tsai, K., Liu, N., Gu, D., et al. (2007). Evidence for mutation showers. *Proc. Natl. Acad. Sci. U. S. A.* 104 (20), 8403–8408. doi:10.1073/pnas.0610902104
- Waters, L. S., Minesinger, B. K., Wiltout, M. E., D'Souza, S., Woodruff, R. V., and Walker, G. C. (2009). Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol. Mol. Biol. Rev.* 73, 134–154. doi:10.1128/MMBR.00034-08
- Wu, A., Zhang, Q., Lambert, G., Khin, Z., Gatenby, R. A., Kim, J. H., et al. (2015). Ancient hot and cold genes and chemotherapy resistance emergence. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10467–10472. doi:10.1073/pnas.1512396112
- Xia, J., Chiu, L. Y., Nehring, R. B., Bravo Núñez, M. A., Mei, Q., Perez, M., et al. (2019). Bacteria-to-Human protein networks reveal origins of endogenous DNA damage. *Cell* 176 (1–2), 127–143. e24. doi:10.1016/j.cell.2018.12.008
- Ye, J., Pavlicek, A., Lunney, E. A., Rejto, P. A., and Teng, C. H. (2010). Statistical method on nonrandom clustering with application to somatic mutations in cancer. *BMC Bioinforma.* 11 (1), 11. doi:10.1186/1471-2105-11-11
- Zhu, H., Swami, U., Preet, R., and Zhang, J. (2020). Harnessing DNA replication stress for novel cancer therapy. *Genes* 11 (9), 990. doi:10.3390/genes11090990