Non-hypermutator cancers access driver mutations through reversals in germline mutational bias

Marwa Z. Tuffaha,[1] David Castellano,[2] Claudia Serrano Colome,[3] Ryan N. Gutenkunst,[2] Lindi M. Wahl[1,*]

[1]Department of Mathematics, Western University,
London, Ontario N6A 5B7, Canada
[2]Department of Molecular and Cellular Biology, University of Arizona,
Tucson, Arizona 85721, USA
[3]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology,
Dr. Aiguader 88, Barcelona 08003, Spain.

[*]Corresponding author: lwahl@uwo.ca.

Cancer is an evolutionary disease driven by mutations in asexually-reproducing somatic cells. In asexual microbes, bias reversals in the mutation spectrum can speed adaptation by increasing access to previously undersampled beneficial mutations. By analyzing tumors from 20 tissues, along with normal tissue and the germline, we demonstrate this effect in cancer. Non-hypermutated tumors reverse the germline mutation bias and have consistent spectra across tissues. These spectra changes carry the signature of hypoxia, and they facilitate positive selection in cancer genes. Hypermutated and non-hypermutated tumors thus acquire driver mutations differently: hypermutated tumors by higher mutation rates and non-hypermutated tumors by changing the mutation spectrum to reverse the germline mutation bias.

Introduction

Cancer is an evolutionary disease arising from DNA mutations that allow cells to proliferate abnormally and invade surrounding tissues and other organs (*1*). During the asexual reproduction of normal somatic cells, *de novo* mutations accumulate through time, leading to an increased risk of cancer with age (*2*). Cancer develops when mutations in specific genes or combinations of genes (so-called "cancer genes") impair normal cell function by, for example, deactivating cell cycle checkpoints or by activating cell signaling pathways that favour cell division (*3*).

The supply of mutations is a key factor underpinning evolutionary trajectories (*4*), and is especially important in asexual reproduction, where recombination is reduced or absent (*5*). In microbial evolution, lineages with an increased mutation rate (mutators) frequently emerge due to their improved access to rare beneficial mutations (*6*). Along with increases in mutation rate, changes in mutation spectra also affect mutational supply (*7–9*). We previously demonstrated that the interaction between mutation rate and spectrum powerfully influences the evolutionary trajectory of asexual populations (*10*). Simply put, previously undersampled mutations may be reached by either making more mutations (mutation rate elevation, fig. 1A), or by making different mutations (mutation spectrum change, fig. 1B). If mutations have historically occurred with some bias (undersampling some classes of mutations while oversampling others), reversing this bias affords access to mutations that were previously unlikely to have occurred, such as undersampled beneficial mutations (*9*).

Cancer is a set of diseases that have similar hallmarks, but there are many differences between cancer types, including different mutational processes and rates (*11*). Broadly speaking, cancer samples can be classified as hypermutated (HM) or non-hypermutated (NHM), where HM tumors are commonly defined as tumors with more than 10 somatic mutations per megabase (*12, 13*). In addition, the underlying mutational processes vary across cancer types,

for example ultraviolet (UV) light is responsible for many mutations in melanoma (*14*). Dysfunction in the mismatch repair system is a leading cause of mutations in several gastrointestinal cancers, such as colorectal, gastric, and hepato-pancreato-biliary cancer (*12*). Whether and to what extent these diverse mutational processes exhibit mutational biases, and how these biases compare with the mutation bias of the germline has not yet been determined.

We investigate the interaction between mutation rate and mutation spectrum changes in cancer. In particular, we compare the mutation spectrum of HM and NHM cancers, demonstrating a highly conserved spectrum across NHM samples that consistently reverses the mutation bias of the germline. Cancer driver mutations thus occur through distinct mechanisms in HM and NHM cancers. In HM cancers the supply of cancer driver mutations is increased through higher mutation rates. In NHM cancers, the supply of driver mutations is increased through a change in mutation spectrum that reverses the mutational biases in the germline.

Results

Reduced transition bias in non-hypermutated samples We define the mutation bias of any mutational type to be reversed, compared to the germline bias, if it differs from the germline bias in the direction of the unbiased state (*10*) (Fig. 1C). Using family-based datasets (*15*), we find that the germline oversamples transitions (Ti) with a relative mutation rate (RMR) of 0.853 in coding mutations (Fig. 1D).

The Ti RMR in 9 normal tissue samples (Fig. 1D) is not significantly different from the germline (t-test, $p = 0.074$). The liver is the only outlier, with a Ti RMR of 0.55, presumably related to the transversion-rich mutational process induced by aristolochic acid exposure (SBS22 in COSMIC database), as observed in female donors (*16*). Similarly, the Ti RMR in the pooled data from HM cancers does not differ significantly from the Ti RMR in the germline (t-test, p=0.57). In contrast, NHM samples show a significantly reduced Ti RMR

(t-test, $p = 9.1 \times 10^{-9}$).

These results are supported when the overall Ti RMR is decomposed at the 1-mer level (Fig. 1E). Note that the 1-mer RMRs of the germline are higher than the uniform spectrum level (1/6) for transitions and lower for transversions. Even when the germline RMR is very close to the uniform level (e.g. C>A mutations), non-hypermutators show strong evidence for a reversed bias; for both transitions they have a significantly reduced RMR compared to the germline, while two of four transversions show a significantly elevated RMR (stars; Bonferroni-corrrected $p < 0.05/18 = 0.0028$). In contrast, the hypermutator 1-mer mutation spectra differ significantly from the germline only for C>G transversions, where the Tv bias is not reduced but is in fact significantly reinforced. Note that none of the 1-mer mutation distributions for normal tissues have significantly different means from the germline. Analogous results hold for whole-genome mutations (Fig. S5).

To extend our analysis to the 3-mer level, we define a new metric, the bias reversal measure, to summarize spectrum changes across mutation types. For any spectrum, the bias reversal measure sums the degree to which the spectrum reverses the bias observed in the germline (see Methods). For passenger-gene mutations in HM and NHM cancers and at the Ti:Tv, 1-mer, and 3-mer levels, non-hypermutators show distributions of this measure across tissues that are significantly higher than zero (Fig. 1F; t-test, $p < 2.8 \times 10^{-9}$ in all three cases), whereas normal tissues and hypermutators are not significantly different from zero (t-test, $p > 0.065$ in all six cases), meaning that the bias is significantly reversed in NHM, but no such pattern is observed in normal tissues or HM. Similar results hold for the whole genome (Fig. S6).

We restricted our analyses of coding spectra to passenger (non-cancer) genes to reduce possible confounding effects of positive selection (see Methods). Returning to cancer genes, we find that in normal tissue, HM, or NHM cancers, mutations in cancer driver genes tend to have a higher bias reversal than mutations in passenger genes ('x' in Fig. 1F). This higher

bias reversal for cancer genes relative to passenger genes is not driven by differences in the expected counts of 3-mer mutations, as these are highly correlated between these gene classes ($R > 0.99$).

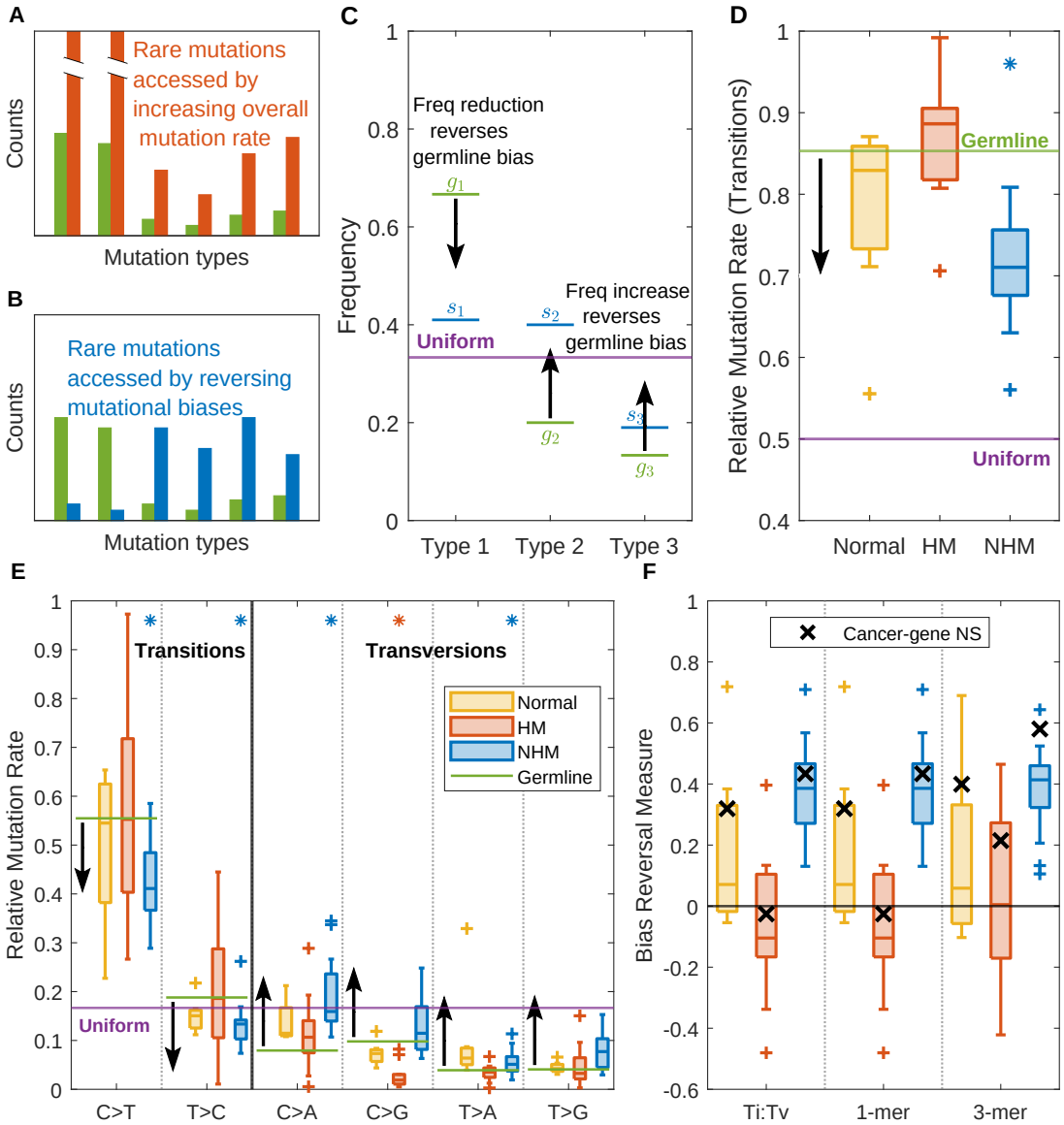

Fig. 1: Caption on a separate page

Fig. 1. (A,B) Schematic: a germline spectrum (green) under-represents some classes of mutations. Access to these classes increases by either an overall mutation rate increase (A) or reversing the bias (B). (C) Example of a spectrum with 3 types of mutations having the same uniform frequency 1/3. Type 1 is over-represented in the germline (g_1, green) while the other two are under-represented. Arrows show the direction that reverses the bias for each mutational type. An example of a bias-reversing spectrum is shown in blue (s_i). (D) Distributions of the RMR for transitions in normal tissues (yellow), hypermutator tumors (HM, red) and non-hypermutator tumors (NHM, blue) in different tissues, compared to the uniform (purple) and germline (green) levels. (E) The corresponding RMRs for each 1-mer mutation type. Stars indicate distribution means that are significantly different from the germline. (F) Bias reversals (y-axis) for passenger gene spectra in different tissues (boxplots) are significantly higher in NHM than in HM or normal tissues on all three levels of analysis (x-axis). The bias reversal measure for non-synonymous mutations in cancer genes is also plotted ('x'). Horizontal bars within boxplots indicate medians; whiskers indicate the 95% confidence interval (CI); '+' signs represent outliers.

Non-hypermutator 3-mer spectrum is highly similar across tissues. The 3-mer mutation rate spectra for HM and NHM samples, summed across tissues, are significantly positively correlated (Fig. 2A, Spearman rank correlation $R = 0.54$, p=1.89$\times 10^{-08}$; also see Fig. S7 and Fig. S8 for the full 3-mer spectra). To further test for similarity among spectra, we computed the rank correlation coefficient between each tissue spectrum and all other spectra (pooled) in the same class. For instance, we correlate the spectrum from each tissue's NHM samples with the spectrum computed by pooling NHM samples from all other tissues. NHM spectra are significantly more strongly correlated with one another (mean $R = 0.86 \pm$ sd $= 0.052$), than HM spectra are with one another (mean $R = 0.62 \pm$ sd $= 0.16$; Fig. 2B t-test, $p = 7.8 \times 10^{-4}$).

Considering unpooled pairs of tissues, the correlation between spectra of NHM cancers is higher than HM cancers (Fig. 2C).

What mutational processes characterize non-hypermutator tumors? To examine the mutational processes underlying mutation spectra changes, we conducted a signature decomposition analysis (*17*)(Fig. 3A). For the pooled NHM spectrum, 52% of the mutations are attributed to mutation signature SBS5, 16.8% are related to the AID/APOBEC family of cytidine deaminases (SBS2 and SBS13), and 14.4% of the mutations are attributed to SBS40, while other signatures have lower weights. All of the signatures composing the NHM spectrum have positive bias reversal measures (Fig. 3B), excluding SBS1 and SBS2.

Since the spectra of HM tumors vary substantially across tissues (Fig. 2C), we aimed to perform tissue-specific signature decompositions of HM samples, but only skin and colon had sufficient mutations (Fig. S1, also see Fig. S8 for the full 3-mer spectra). As expected, UV light signatures (SBS7a through SBS7d) dominate the skin spectrum (Fig. 3A); the two dominant signatures, SBS7a and SBS7b, have a reinforced bias (negative bias reversal measure; Fig. 3B). In contrast, mutations derived from a defective polymerase epsilon dominate the colon HM spectrum (SBS10a and SBS10b, and the associated SBS28; Fig. 3A). All signatures underlying the colon HM spectrum have a positive bias reversal measure (Fig. 3B), explaining why colon cancer was the outlier with strong transition bias reversal in HM cancers (Fig. 1D).

Positive selection in cancer genes anticorrelated with the germline spectrum. In the absence of positive selection, both cancer and passenger genes should have the same genomic mutation rates. To examine positive selection in both HM and NHM tumors, we define an excess measure to quantify the degree to which the number of mutations of a particular type that are observed in cancer genes exceeds the number that would be expected, given the mutation rates observed in passenger genes (Eq. S2, Methods). In both HM and NHM categories,

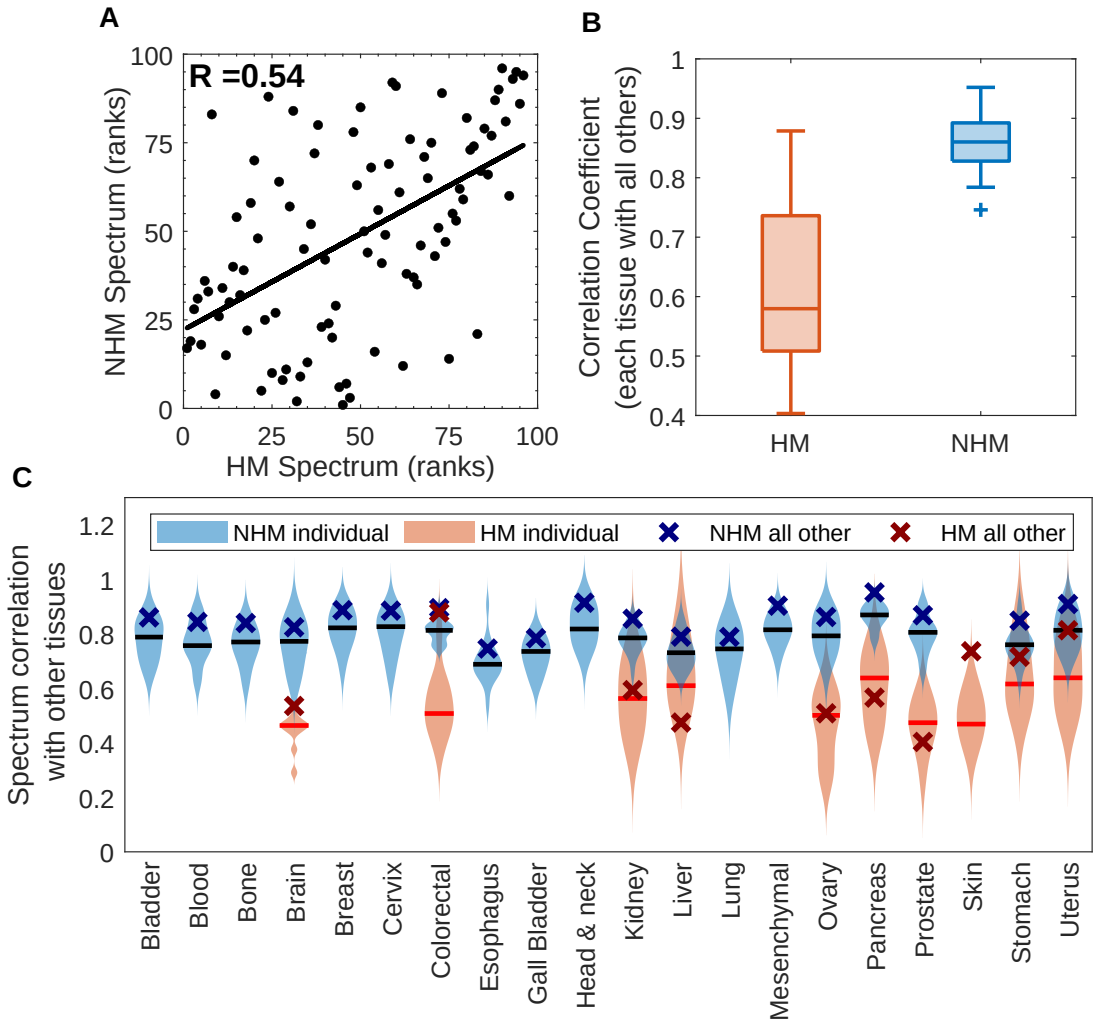

Fig. 2: Non-hypermutator spectra are highly similar across tissues, unlike hypermutator spectra. (A) Correlation of the 3-mer spectra from NHM samples and HM samples; the Spearman correlation coefficient is $R = 0.54$. (B) Correlating the spectrum of each tissue with the pooled spectrum from all other tissues in the same class (HM or NHM) shows high Spearman correlation coefficients for NHM (blue) and lower values for HM (red). Centre bar indicates distribution median and whiskers show 95% CI; '+' signs represent outliers. (C) Spearman correlation coefficients of the spectrum of each tissue with each other tissue in the same class (violin plots); horizontal lines represent the medians of the distributions. Correlation coefficients between each tissue and pooled samples from all other tissues in the same class ('x' symbols) are also shown for comparison; these are the values summarized in panel (B).

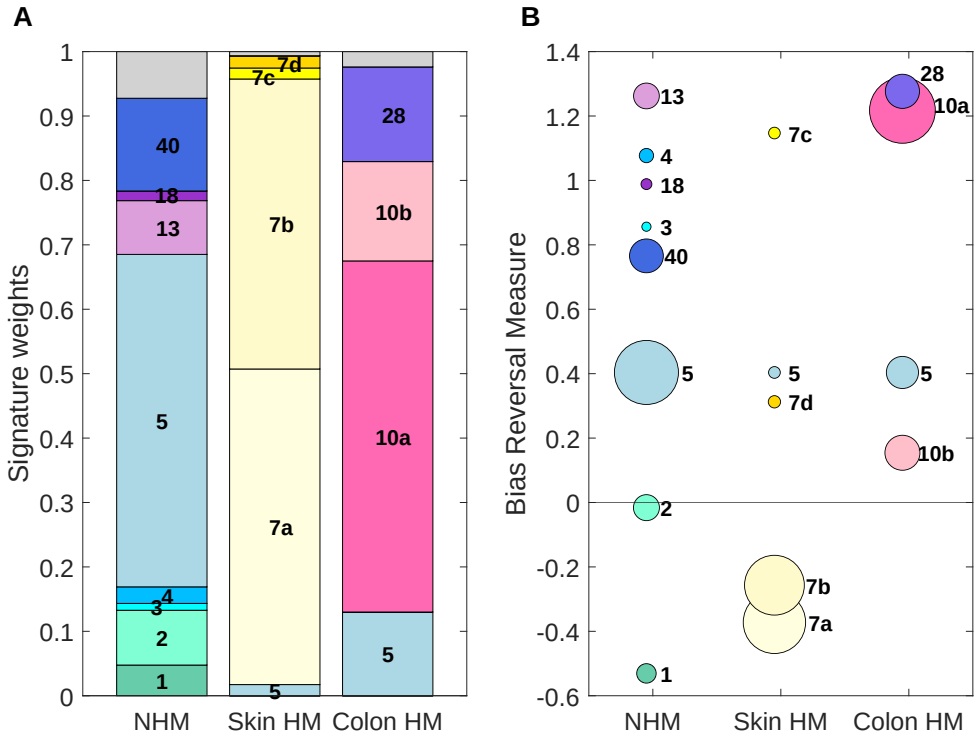

Fig. 3: Most signatures composing the NHM and colon HM spectra reverse the bias, while mutational processes tend to reinforce the bias in skin HM samples. (A) The weights for each mutational signature identified as contributing to the three spectra are shown; grey proportions at the top represent mutations from unknown mutational processes. (B) Bias reversal measures for each signature; circles have an area proportional to the weight of that signature within the respective spectrum.

we observe positive and negative excess measures for different 3-mer mutation types, but non-synonymous mutations in NHM show the highest number of positive values, with a distribution mean that is significantly different from zero (Fig. 4A; t-test, $p = 2.02 \times 10^{-21}$). The non-synonymous distribution in NHM is significantly different from each of the other three distributions (Wilcoxon rank sum test, $p < 1.1 \times 10^{-8}$). In agreement with previous work (*18, 19*), we thus find stronger evidence for positive selection acting on non-synonymous mutations in NHM than in HM tumors. Note that the distribution mean for non-synonymous mutations in HM is also significantly different from zero (t-test, p=0.0104), but the distribution means for synonymous mutations are not significantly different from zero (t-test, Bonferroni-corrected, $p = 0.35$ and $p = 0.04$ respectively).

To detect selection on individual 3-mer mutation types (and to account for the variance in excess measure for rare mutation types), we compared the numbers of mutations observed in cancer genes with bootstrapped samples from all coding mutations. For almost all 3-mer contexts in HM and NHM, the number of synonymous mutations in cancer genes falls within the 99.95% confidence interval of the bootstrapped samples, showing no sign of selection (Fig. 4B). In contrast, non-synonymous mutations show 24 positively-selected 3-mer mutation types in NHM (out of 96), but only 2 in HM (fig. 4C). Non-synonymous excess measures in these 24 3-mer mutation types are anti-correlated with the corresponding mutation count frequencies in the non-synonymous germline in both NHM (Fig. 4D) and HM (Fig. 4E) cancers, demonstrating that among mutations under positive selection, those least likely to occur in the germline show the strongest selective effect. There is no significant correlation, however, between the germline mutation spectrum and the measure of positive selection across all 3-mer mutation types.

Negative selection is typically detectable only in tumors with a very low mutation burden, as expected under strong clonal interference (*18*); consistent with this expectation, we found no 3-mer mutation type with an observed value that falls below the confidence interval, in HM or

in NHM (Fig. 4B).

Discussion

We examined the mutational spectra of thousands of human tumor samples, computing spectra for hypermutator (HM) and non-hypermutator (NHM) cancers across twenty tissue types. We demonstrate that the mutational spectra in NHM tumors is highly correlated across tissues, an effect not observed in HM tumors (Fig. 2). Thus while HM tumors show a dramatic increase in mutation rate, NHM tumors show only modest increases in mutation rate but have a distinct mutation spectrum that is repeated across diverse tissues and donors.

Reversals of mutation bias offer access to undersampled classes of mutations (*10*), which we hypothesized may include the mutations that drive cancer. Consistent with this prediction, whether analyzed at the Ts:Tv, 1-mer or 3-mer levels, the distinct mutation spectrum of NHM (Fig. 2) tumors significantly reverses the germline bias (Fig. 1), while HM tumors and normal tissues show no such effect. These results suggest that while HM tumors access cancer driver mutations through elevated mutation rate, NHM tumors access driver mutations through changes in mutation spectrum that correspond to reversals in germline mutational biases (Fig. 5).

We further confirm this hypothesis by demonstrating that driver mutations are indeed stronger, and/or more common, for those types of mutations that are less likely to occur in the germline or in healthy tissues (Fig. 4D,E). This suggests that indirect selection on 3-mer mutation modifiers (*20*) may tune the human germline and healthy somatic mutation spectrum such that the greater the effect of a cancer driver mutation, the less likely it is to occur, and if it does occur, the less likely it is to remain unrepaired.

Our signature decomposition analysis (Fig. 3) suggests SBS40 as a mutational process that plays a key role in the NHM mutation spectrum and reverses the germline bias. While SBS5

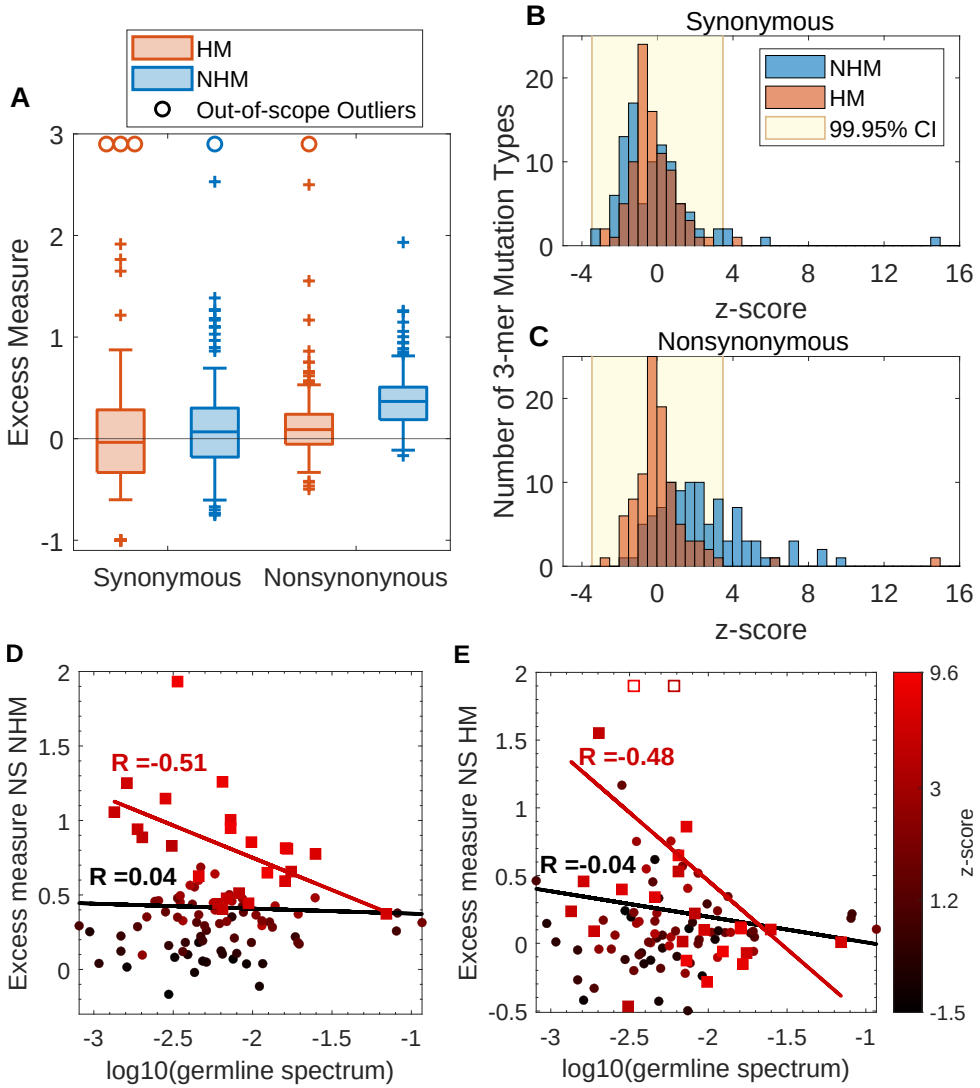

Fig. 4: Evidence of positive selection in cancer genes is stronger in non-hypermutators than hypermutators. (A) Excess measures comparing cancer and non-cancer genes for all 3-mer mutation types are shown for synonymous and non-synonymous mutations in both HM (red) and NHM (blue). Only the non-synonymous distributions are significantly higher than zero. Boxplot whiskers include 95% CI. The positive y-axis is truncated for clarity; '+' signs represent outliers. (B) Histograms of z-scores obtained by comparing the numbers of synonymous mutations of each 3-mer mutation type in cancer genes with results in 50,000 bootstrapped samples from all synonymous mutations. The Bonferroni-corrected CI from the boostrapped dataset is shown in yellow. (C) Analogous results for non-synonymous mutations. (D,E) Excess measure for non-synonymous mutations in NHM and HM is correlated with the germline non-synonymous passenger-gene spectrum (log-scale). The 3-mer mutations that show significant positive selection for NS NHM (out of CI in panel (C), squares) show a significant anti-correlation in both cases (red lines, (D):p=0.012, (E):p=0.017), while all 3-mer mutations taken together (squares and circles) do not show any significant correlations (black lines, (D):p=0.7, (E):p=0.67). The color of each point corresponds to the NHM z-score from panel (C). Open circles/squares (in panels A and E) indicate out-of-scope outliers.

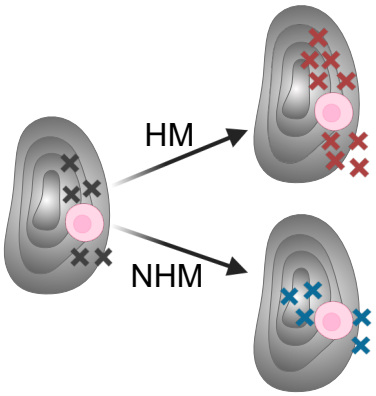

Fig. 5: Cancer driver mutations are rarely accessed but "beneficial" in the fitness landscape (grey surface) of a cellular lineage. These can be accessed by two possible changes in mutational processes: an increase in mutation rates as seen in hypermutators (HM), or a reversal in the mutational bias as seen in non-hypermutators (NHM).

makes the largest contribution to the NHM spectrum, it is a clock-like signature (prevalence increases with patient age) that shows a modest bias reversal and is found in the germline and healthy tissues as well. In contrast, SBS40 is a mutational signature detectable in many cancers (see Fig. 3 in (*11*)); because it mimics a uniform spectrum, it strongly reverses the germline transition bias (Fig. 3). There is a significant positive correlation between SBS40 and hypoxia (*17*), and the presence of transient or chronic hypoxia, characterized by critically low tissue oxygen, is a defining characteristic of cancer (*21*). This hypoxic state is associated with a significant increase in mutation rate, thought to result from a reduction in the effectiveness of various DNA repair mechanisms (*22, 23*). Further studies are needed to determine whether the distinct and remarkably stable mutation spectrum observed in NHM tumors is due to alterations in DNA repair enzymes resulting from common hypoxic conditions among patients, from other unknown causes shared among patients, or from multifactorial convergent evolution at the mutation spectrum level.

In contrast with the NHM spectrum, we observed a wide variability in the spectra of HM tumors. While skin cancers show no evidence of reversing the germline bias, HM samples

from colon show a strong bias reversal. Simultaneously reversing mutation bias and increasing mutation rate provides more access to beneficial mutations than either alone (*10*). We speculate that the high incidence of colon cancer could be in part due to this powerful combination of an increased mutation rate along with a reversal of the germline bias.

By comparing observed and expected mutation counts, we show that NHM tumors have an excess of non-synonymous mutations in cancer genes in several 3-mer contexts, which we interpret as evidence for positive selection (*19, 24, 25*). This excess is significant not only for non-synonymous mutations, but for synonymous mutations in a few contexts, which is not completely unexpected since several synonymous mutations have been previously identified to act as cancer driver mutations due to their impact on splicing, RNA secondary structure and expression levels (*26, 27*).

On the other hand, we find no strong signal for positive selection in HM samples. This could be due to the combined effects of clonal interference (*18, 28*) and deleterious load in HM tumors, such that only a small fraction of mutations will successfully spread through the population. This effect could dilute signals of positive selection when mutation rates are very high. Since the small fraction of beneficial mutations that spread are also expected to have very large selective effects, negative epistasis could also reduce the signals of positive selection in HM tumors. This can occur when a single, large-effect mutation eliminates the need for further mutations that would otherwise be beneficial, for example when cancer driver genes "break" the same regulatory pathway. Moreover, HM tumors are known to have longer sequence context dependencies than those accounted for by our 3-mer model (*25, 29*). This may also make our detection of positive selection across 3-mer mutations less reliable in HM tumors than in NHM tumors. Nevertheless, we find that the excess of non-synonymous mutations in the 24 3-mer mutation types that show positive selection in NHM is also anticorrelated with the germline mutation spectrum in HM tumors (Fig. 4D,E). This suggests that the underlying

fitness landscape may be quite similar for HM and NHM, but the inference of positive selection is simply more difficult in HM.

Our work putatively identified 24 3-mer mutation types that show positive selection in NHM. Classifying the contribution of these mutations to oncogenes and tumor suppressor genes is a clear avenue for future work, as is characterizing the distribution of fitness effects of these mutation types in the germline. More generally, we hope that characterizing the distinct 3-mer spectrum observed in NHM cancers may shed further light on cancer driver mutations and their critical role in early oncogenesis.

References

1. J. S. Brown, *et al.*, Updating the definition of cancer. *Molecular Cancer Research* **21**, 1142 (2023).

2. Cancer Research UK, Age and cancer, `https://www.cancerresearchuk.org/about-cancer/causes-of-cancer/age-and-cancer` (2024). Accessed: 2024-02-13.

3. National Cancer Institute, What is Cancer?, `https://www.cancer.gov/about-cancer/understanding/what-is-cancer` (2021). Accessed: 2024-01-23.

4. A. Stoltzfus, L. Y. Yampolsky, Climbing mount probable: Mutation as a cause of nonrandomness in evolution. *Journal of Heredity* **100**, 637 (2009).

5. T. Johnson, Beneficial mutations, hitchhiking and the evolution of mutation rates in sexual populations. *Genetics* **151**, 1621 (1999).

6. Y. Raynes, P. D. Sniegowski, Experimental evolution and the dynamics of genomic mutation rate modifiers. *Heredity* **113**, 375 (2014).

7. J. F. Storz, *et al.*, The role of mutation bias in adaptive molecular evolution: insights from convergent changes in protein function. *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**, 20180238 (2019).

8. K. Gomez, J. Bertram, J. Masel, Mutation bias can shape adaptation in large asexual populations experiencing clonal interference. *Proceedings of the Royal Society B* **287**, 20201503 (2020).

9. M. Sane, G. D. Diwan, B. A. Bhat, L. M. Wahl, D. Agashe, Shifts in mutation spectra enhance access to beneficial mutations. *Proceedings of the National Academy of Sciences* (2023).

10. M. Z. Tuffaha, S. Varakunan, D. Castellano, R. N. Gutenkunst, L. M. Wahl, Shifts in mutation bias promote mutators by altering the distribution of fitness effects. *The American Naturalist* **202**, 503 (2023).

11. L. B. Alexandrov, *et al.*, The repertoire of mutational signatures in human cancer. *Nature* **578**, 94 (2020).

12. K. Yuza, M. Nagahashi, S. Watanabe, K. Takabe, T. Wakai, Hypermutation and microsatellite instability in gastrointestinal cancers. *Oncotarget* **8**, 112103 (2017).

13. B. B. Campbell, *et al.*, Comprehensive analysis of hypermutation in human cancer. *Cell* **171**, 1042 (2017).

14. L. Garibyan, D. E. Fisher, How sunlight causes melanoma. *Current oncology reports* **12**, 319 (2010).

15. M. Rodriguez-Galindo, S. Casillas, D. Weghorn, A. Barbadilla, Germline de novo mutation rates on exons versus introns in humans. *Nature communications* **11**, 3304 (2020).

16. R. Li, *et al.*, A body map of somatic mutagenesis in morphologically normal human tissues. *Nature* **597**, 398 (2021).

17. C. Serrano Colome, O. Canal Anton, V. Seplyarskiy, D. Weghorn, Mutational signature decomposition with deep neural networks reveals origins of clock-like processes and hypoxia dependencies. *bioRxiv* (2023).

18. S. Tilk, S. Tkachenko, C. Curtis, D. A. Petrov, C. D. McFarland, Most cancers carry a substantial deleterious load due to hill-robertson interference. *Elife* **11**, e67790 (2022).

19. I. Martincorena, *et al.*, Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029 (2017).

20. W. R. Milligan, G. Amster, G. Sella, The impact of genetic modifiers on variation in germline mutation rates within and among human populations. *Genetics* **221**, iyac087 (2022).

21. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *cell* **100**, 57 (2000).

22. A. R. Kaplan, P. M. Glazer, Impact of hypoxia on dna repair and genome integrity. *Mutagenesis* **35**, 61 (2020).

23. V. Bhandari, C. H. Li, R. G. Bristow, P. C. Boutros, Divergent mutational processes distinguish hypoxic and normoxic tumours. *Nature communications* **11**, 737 (2020).

24. D. Weghorn, S. Sunyaev, Bayesian inference of negative and positive selection in human cancers. *Nature genetics* **49**, 1785 (2017).

25. F. Dietlein, *et al.*, Identification of cancer driver genes based on nucleotide context. *Nature genetics* **52**, 208 (2020).

26. F. Supek, B. Miñana, J. Valcárcel, T. Gabaldón, B. Lehner, Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**, 1324 (2014).

27. Y. Sharma, *et al.*, A pan-cancer analysis of synonymous mutations. *Nature communications* **10**, 2569 (2019).

28. P. J. Gerrish, R. E. Lenski, The fate of competing beneficial mutations in an asexual population. *Genetica* **103**, 127 (1998).

29. E. D. Pleasance, *et al.*, A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191 (2010).

30. L. Moore, *et al.*, The mutational landscape of human somatic and germline cells. *Nature* **597**, 381 (2021).

31. Wellcome Sanger Institute, Cancer Gene Census, `https://cancer.sanger.ac.uk/census` (2024). Accessed: 2024-02-06.

32. J. M. Hess, *et al.*, Passenger hotspot mutations in cancer. *Cancer Cell* **36**, 288 (2019).

33. R. Rosenthal, N. McGranahan, J. Herrero, B. S. Taylor, C. Swanton, deconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biology* **17**, 31 (2016). Publisher: BioMed Central Ltd.

34. S. Li, F. W. Crawford, M. B. Gerstein, Using sigLASSO to optimize cancer mutation signatures jointly with sampling likelihood. *Nature communications* **11**, 3575 (2020). Publisher: Nature Publishing Group UK London.

35. H. Jin, *et al.*, Accurate and sensitive mutational signature analysis with MuSiCal. *Nature Genetics* pp. 1–12 (2024). Publisher: Nature Publishing Group.

36. P. Cingolani, *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: Snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *fly* **6**, 80 (2012).

Acknowledgments

Funding: This work was supported by the Natural Sciences and Engineering Research Council of Canada grant RGPIN-2019-06294, by the National Institute of General Medical Sciences of the National Institutes of Health through grants R01GM127348 and R35GM149235, and by the Spanish Ministry of Science and Innovation through the Centro de Excelencia Severo Ochoa (CEX2020-001049-S, MCIN/AEI /10.13039/501100011033), the Generalitat de Catalunya through the CERCA programme, and the European Union's H2020 research and innovation programme under Marie Sklodowska-Curie grant agreement No.754422.

Authors' contributions: MZT, DC, RNG and LMW designed the study. MZT and DC performed data analysis; CSC performed spectral decompositions. All authors contributed to data interpretation. MZT, LMW and RNG created the figures. MZT and DC drafted the manuscript. All authors edited and finalized the manuscript.

Competing interests: none.

Data and materials availability: Data are reported in the supplementary tables. Tables S1 to S4 provide data shown in figures 1 to 4 in the main text, respectively. Table S5 provides the 3-mer spectra plotted in figures S7 and S8.

Supplementary materials

Materials and Methods

Figures S1 to S8

Tables S1 to S5

References (*30–35*)