Effect sizes of somatic mutations in cancer

Vincent L. Cannataro¹, Stephen G. Gaffney¹, Jeffrey P. Townsend^{1,2,3†}

¹Department of Biostatistics, Yale University, New Haven, CT,

²Program in Computational Biology and Bioinformatics and

³Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT

[†]To whom correspondence should be addressed.

Conflict of interest disclosure statement: The authors declare no conflicts of interests.

ABSTRACT

A major goal of cancer biology is determination of the relative importance of the genomic alterations that confer selective advantage to cancer cells. Tumor sequence surveys have frequently ranked the importance of substitutions to cancer growth by *P* value or a false-discovery conversion thereof. However, *P* values are thresholds for belief, not metrics of effect. Their frequent misuse as metrics of effect has often been vociferously decried. Here, we estimate the effect sizes of all recurrent single nucleotide variants in 23 cancer types, quantifying relative importance within and between driver genes. Some of the variants with the highest effect size, such as EGFR L858R in lung adenocarcinoma and BRAF V600E in primary skin cutaneous melanoma, have yielded remarkable therapeutic responses. Quantification of cancer effect sizes has immediate importance to the prioritization of clinical decision-making by tumor boards, selection and design of clinical trials, pharmacological targeting, and basic research prioritization.

STATEMENT OF SIGNIFICANCE

We calculated the effect size for all recurrent single-nucleotide variants in 23 cancer types, quantifying the relative importance of the mutations driving cancer cell replication and survival. This quantification provides a means to prioritize basic research, inform decisions by tumor boards, and inform design of clinical trials.

INTRODUCTION

Since the advent of whole-exome and whole-genome sequencing of tumor tissues, studies of somatic mutations have revealed the underlying genetic architecture of cancer (1), producing ordered lists of significantly mutated genes whose ordering implies their relative importance to tumorigenesis and cancer development. Typically, differentiation of selected mutations from neutral mutations is performed by quantifying the over-representation of mutations within specific genes in tumor tissue relative to normal tissue, and disproportionate prevalence of somatic mutations in a gene has been taken as *prima facie* evidence of a causative role for that gene. Two quantifications have implicitly ordered the importance of discovered cancer "driver" genes: the prevalence of the mutation among tumor tissues sequenced from that tumor type (2,3), the statistical significance (P value) of the disproportionality of mutation frequency (4), or both (1). Versions of these metrics have shifted from simple ranks by mutation prevalence in a tumor population (5-7) to calculation of statistical significance of mutation prevalence over genome-wide context-specific background mutation rates (8–10), to ratios of the prevalence of nonsynonymous and synonymous mutations (11), to P values based on a gene-specific mutation rate and a diversity of genomic data (12). Although the approaches used to calculate P values have become more sophisticated, P values are not an appropriate metric for quantifying the vital role of genes or their mutations to tumorigenesis and cancer development, as P values are thresholds for belief (13) and not metrics of effect (14,15). Failure to report effect sizes is a persistent issues in the scientific and biomedical literature that can massively misdirect research and health care decision-making (16-20).

While prevalence of a somatic mutation in a cancer type has important consequences for biomarker studies (21) and identification of therapeutic population for a targeted therapy (22), there is only a correlative—rather than causal—link between prevalence of mutation and its contribution to tumorigenesis and cancer development. The lack of causal linkage is easily seen by considering the mutated genes that, in spite of their high prevalence in tumor populations, are universally regarded as false positives. For example, the gene *TTN* is a structural protein of striated muscle. Because it is long, because it replicates relatively late in the synthesis phase of mitosis, and because it is inaccessible to transcription-mediated repair in non-muscle tissues, it has a high mutation rate and is frequently mutated in non-sarcoma cancer tissues. While *TTN* is an extreme example—sometimes showing up at the top of lists ordered by prevalence of genic mutation (11,12)—it exemplifies the problem with using mutation prevalence as a proxy for importance. Any consideration of whether mutated genes are contributing to tumorigenesis and cancer development—or of the degree to which they are contributing—must address the issue of their underlying mutation rate.

The appearance of such biologically implausible genes in ranked mutation lists prompted the development of increasingly sophisticated statistical approaches designed to "weed out" false-positives via calculation of a *P* value that accounted for gene length and background mutation rate. The classical evolutionary biology approach is to use the frequency of synonymous site mutations in each gene as a proxy for mutation rate. As in the divergence of species, synonymous site mutations are presumably neutral (or nearly so) to the success of cancer lineages during the divergence from normal to resectable tumor. As the number of mutations observed in a given gene is typically much smaller in the somatic evolution of cancer than is observed in the divergence between most species, use of synonymous sites within a single gene leads to many genes with zero synonymous mutations in most cancers, and an ineffective calculation of P value. Alternative approaches currently in use obtain a reasonably robust estimate of genic mutation rate using correlates such as gene expression levels, chromatin states, and replication timing, and are largely successful at excluding known false positives (12).

In genomic tumor surveys, the sample size of tumors varies among studies, posing a problem for comparison of *P* values within or between cancer types (14,15). An even more serious issue with using *P* values for ranking genes or mutations arises from the same source that obviates use of genic mutation prevalence: the confounding effect of mutation rate. Because the "sample size" of overall mutations in a gene is dictated by the genic mutation rate, it is much easier for genes with high mutation rates not only to reach high genic prevalence, but also to reach statistical significance despite small effect sizes. While approaches accounting for genic mutation rates will eliminate false positives (12), and the *P* value will serve to exclude genes like *TTN* that have no role in tumorigenesis and cancer development, the rank order by *P* value of genes that do have a role in tumorigenesis and cancer development will remain highly affected by mutation rate. Genes with higher mutation rate will (correctly) be more likely to achieve statistical significance, and thus will appear deceptively high on a ranked list whose ordering suggests importance in tumorigenesis and cancer development.

Because genic mutation prevalence and *P* value inadequately capture importance to tumorigenesis and cancer development, another metric must be appropriate. To provide an evaluation of the relative importance of mutations in diverse cancer types to tumorigenesis and cancer development, we called on an understanding of the development of cancer as an

6

evolutionary process (23,24), and adapted some straightforward insights from classical evolutionary theory. The cognate metric in evolutionary theory for quantifying importance to tumorigenesis and cancer development is the selective effect of the mutation on the cancer lineage. The appropriateness of this metric is fairly easy to recognize. While mutations are the ultimate source of variation contributing to tumorigenesis, we do not conduct genomic tumor sequence surveys to discover neutral mutation rates. We conduct them to determine which mutations spread within cancer tissues because of the effects of mutations on proliferation and survival. Mutation rate is a confounding phenomenon: when it is high, it also increases prevalence of mutations. Because silent site substitutions and other correlates of baseline mutation rate provide a means to independently differentiate silent mutation rate from the impact of natural selection within the tumor, selection intensities can be estimated, providing the effect sizes of each mutation.

RESULTS

We calculated cancer effect sizes by comparing the rate of observed substitutions to the rate that substitutions would be expected to arise in the absence of selection (25). In accordance with population genetic theory, we specify that the rate neutral mutations arise and the rate that they fix as substitutions within tumors are equivalent, and that non-neutral mutations arise at a consistent rate. Thus any increase in the flux of substitutions among tumors of a particular context above the baseline silent rate would be the appropriate estimate of the intensity of selection on that mutation within the tumor population (Methods). These selection intensities quantify the survival and proliferative advantage conferred by variants, facilitating comparisons

of the relative importance of drivers among and within tumor types. For instance, even though lung tissues encounter similar mutagens and hence mutations arise through similar mutational signatures, LUAD and LUSC variants conferring the largest effect sizes are markedly different among the two cancers (Fig. 1). Moreover, within a tumor type, the selection intensity decouples the effects of mutation rate on frequency: even though KRAS G12C variants are over twice as prevalent as EGFR L858R variants in LUAD tumors, EGFR L858R is estimated to have a higher effect size, due to its much lower baseline mutation rate. Furthermore, the relative selection intensity among variants within a single gene reveals whether 'hotspots' of somatic variation are purely driven by mutational processes or rather under differential (or similar) selective pressure. For instance, the V157F, R158L, R175G, G245C, and E298* variants in TP53 in LUSC tumors exhibit markedly different prevalences attributable to different trinucleotide mutation contexts, yet resulting in remarkably similar estimated effect sizes.

This calculation yielded cancer effect sizes for all fixed substitutions (Methods, Figure S1, Table S1) that quantify contribution to the cancer phenotype within 23 tumor types. Their relative rank corresponds to their relative importance within the respective tumor types. Several common known oncogenic substitutions, such as BRAF V600E in SKCMP and EGFR L858R in LUAD (26,27), and substitutions in known tumor suppressor genes, such as APC in READ and TP53 in HPV-negative HNSCC, are highly selected, and those genes are also typically determined as significantly mutated (e.g., by MutSigCV(12)). However, genes determined to be statistically significantly altered in cancer are also well-dispersed within a large range of site-specific cancer effect sizes (Fig. 2), illustrating how discrepant gene-specific *Q* values are with site-specific cancer effect sizes. Several substitutions within genes that are not statistically

significantly over-mutated are interspersed among more prevalent substitutions within genes that are estimated to be significantly mutated, for instance Mastermind-like3 (MAML3) G1069A in READ, a protein that binds to and stabilizes the DNA-binding complex of the Notch intracellular domain (28) and Nuclear factor (erythroid derived 2)-like 2 (NFE2L2) R34G in UCEC, a protein that is believed to play a causative role in squamous cell lung cancer (29,30). Indeed, substitutions in this gene comprise two of the top three most selected substitutions within our analysis of lung squamous cell carcinoma (Fig. 2).

DISCUSSION

Here, we have argued that frequencies of mutation and *P* values are not an appropriate metric for the importance of somatic variants in tumor growth. We derive the appropriate metric—the cancer effect size of mutations. This effect size, quantifying the intensity of selection on mutations in cancer cells in patients, conveys the replicative and survival benefit conferred by genetic variants, and therefore is a direct metric of the contribution of a variant to the cancer phenotype. Using our approach, we estimated the effect size of all recurrent SNVs in 23 cancer types, re-evaluating their importance across SNV effect size to cancer in disparate tumor types projected to account for approximately 82% of all new cancer cases within the United States in 2017 (31).

Current approaches using conservative *P* values are particularly underpowered to detect genes that are of high importance to tumorigenesis and cancer development in some cancer cases because the site or sites conveying the relevant phenotype are mutated at low rates. For instance, FBXW7 R505G was estimated to have the highest selection intensity in HPV⁺ HNSCC, and

9

BRAF V600E was estimated to have the fifth highest selection intensity in LGG and the tenth highest selection intensity in LUAD, but both of these genes were classified as not significantly mutated at the gene-level within these three cancer types. Mutations within these two well-known oncogenes were estimated to confer large effect sizes, and these genes were determined to be significantly mutated in other cancer types, yet their importance in patients who have these rarer somatic variants within HPV⁺ HNSCC, LGG, and LUAD has been poorly illuminated by gene-wide analyses of statistically significant mutation burden across patients. Thus, calculating the cancer effect size at the level single nucleotide variants identifies drivers that have low prevalence but large effect in different tumor types.

Identification of these low-prevalence high-effect drivers is increasingly important as precision targeting of therapeutics becomes increasingly integrated into clinical trial design. Targeted therapies are often developed for, and necessarily tested in, a single tumor type with the targeted variant at high frequency. However, the same variant often exists at lower prevalence in other tumor types. Quantifications of the cancer effect size can guide the selection and design of clinical trials to target small populations that can benefit from targeted therapeutics developed for other cancer types. Targets for therapies that were originally developed for variants at high prevalence in one tumor type are expected to be effective when treating the same variant in a secondary tumor type if the variant has a similarly high selection intensity in the second tumor type, even if it is at low prevalence in the newly considered tumor type.

The effect sizes of somatic mutations in cancer should play key roles in clinical decision-making, providing crucial insight into the relative upper limits of the efficacy of precision-targeted treatments. When a therapy abrogates novel oncogenic function primarily by

10

disabling a gain-of-function mutation, the upper limit of the efficacy of the precision-targeted treatment will be dictated by the selection intensity that somatic variant imparted to the cancer cell lineage. Therefore, while incorporating other knowledge regarding the pharmacokinetics, efficacy, and side effects, and the evolution of resistance to therapies, precision treatments can be selected in clinical decision-making on tumor boards to target mutations with the greatest cancer effect size. That is, the effectiveness of a targeted therapy is expected to scale with the selection intensity of the target for any therapy that specifically inhibits the selective advantage conferred by the mutant allele. Similarly, cancer effect size can be used in the same manner to prioritize targeted development of new therapeutics, indicating the upper limit of effect for a perfect therapeutic ameliorating an oncogenic mutation.

The relative effect sizes of cancer mutations can inform nearly every aspect of basic research related to oncology, including which genes and pathways deserve greater attention in basic research. While network interactions, epigenetic and tumor microenvironment interactions, and aspects of cellular differentiation and cellular plasticity mean that quantification of the selective effect of mutations does not provide an upper bound on the importance of a gene or pathway in the molecular biology of cancer, its quantification does provide a lower bound, as the potential role of genes with a somatic variant can be no lower than the selection intensity the variant imparts. It does not escape our notice that here we only calculate the effect size of single nucleotide variants. Importantly, effect sizes of other mutational processes, such as copy number alterations or epigenetic modifications, could be similarly calculated once estimates of intrinsic mutation rate in the absence of selection are identifiable for these processes.

METHODS:

Data acquisition and processing:

Data were either obtained either from The Cancer Genome Atlas (TCGA) projects downloaded from the National Cancer Institute's Genomic Data Commons (32), or derived from whole exome sequencing of tumors as part of collaboration with Gilead Sciences (Table S2). All TCGA data used in this analysis were from GDC version gdc-1.0.0 and relevant UUID are found in Table S3.

All TCGA data were first converted to hg19 coordinates using the liftOver function of the rtracklayer R package(33). To obtain consistency between gene symbols used in MutSigCV and in mutation data, symbols in the MutSigCV default covariates files were mapped to HUGO symbols. Non-HUGO symbols were mapped using substitutions from the NCBI Gene database using unambiguous 'synonym' matches; unambiguous 'previous symbol' matches; and manual lookups. CDS coordinates for each gene were obtained from UCSC's hg19 annotation database. The MutSigCV covariates files, and gene annotation files used in our analyses can be found at https://github.com/Townsend-Lab-Yale/SNV selection intensity. Nucleotide variants one or two positions apart in the same tumor sample were removed from the analysis to ensure that we analyzed only single nucleotide variants. Head and neck squamous cell tumors were designated as HPV positive if they both contained greater than 100 HPV RNA viral transcript reads per hundred million (RPHM) (34) and were designated positive in clinical data obtained from The Broad GDAC Firehose (35), and tumors were designated HPV negative if they contained less than 100 HPV RNA viral transcript RPHM and were designated negative in clinical data obtained from The Broad GDAC Firehose.

Calculating mutation rate and selection intensity:

We define the selection intensity of a single nucleotide variant as the ratio of the flux of fixed mutations to the expected flux of fixed mutations in the absence of selection. We estimate the expected rate of fixed mutations in the absence of selection by using MutSigCV to calculate the silent mutation rate for each gene. We use the median gene expression from cell lines derived from each analyzed tissue type (Table S3) as the MutSigCV expression covariate for that tissue type. We then normalized this gene-level rate at every site for every mutation across each gene given the specific trinucleotide mutation profile of the tissue, such that the average rate of mutations among all trinucleotide combinations in each gene is equal to the average rate calculated by MutSigCV. The trinucleotide profile of a tissue is calculated as the average of trinucleotide COSMIC signatures among non-recurrent mutations calculated by the deconstructSigs R package (36) for all tumors with over 50 single nucleotide variants.

Formally, we define the rate of non-neutral substitutions, λ , as $N\mu \times u(s)$, where *N* is the effective population size of the tumor, μ is the mutation rate of a substitution, and u(s) is the probability the mutation fixes within the population, which is a function of the selection coefficient of the mutation, *s*. The probability that a silent mutation fixes is the inverse of the population size within which it fixes (37), and thus rate of non-neutral substitutions, divided by the rate of neutral substitutions, is $\frac{\lambda}{\mu} = \frac{N\mu \times u(s)}{N\mu \times \frac{1}{N}}$, or $N \times u(s)$, the selection intensity and effect size of the mutation. The term selection intensity is used as a direct parallel to the classic derivation of "scaled selection coefficient" or "selection intensity" in the population genetics literature (38–41). Because a tumor sequence is a single snapshot in time, we can only detect one substitution event per site per tumor. To correct for this issue of detection, we define the rate of substitution, λ , as the Poisson rate of occurrence of one *or more* observed fixation events, i.e. the value of λ that maximizes $(e^{-\lambda})^{n_0} \times (1 - e^{-\lambda})^{n_1}$, where n_0 is the number of tumors without any substitution in that gene and n_1 is the number of tumors with the specified substitution at that site. We define n_0 as the number of tumors without any substitutions in that gene, instead of in that gene at that specific site, because doing so accounts for the reduced or eliminated selective pressure on additional mutation for that gene when other sites are already mutated. We define substitutions as single nucleotide variants that are observed in tumors from more than one patient (recurrent) within our dataset, and we calculate the selection intensity of the recurrently mutated substitutions to minimize the probability of analyzing passenger mutations.

Scripts used to perform this analysis are available online at https://github.com/Townsend-Lab-Yale/SNV selection intensity

Supplementary Information available online.

Acknowledgements: This project was supported by Gilead Sciences, Inc. We also thank the

Yale High Performance Computing Center for providing computational resources.

Author contributions: VLC and JPT designed the analyses. VLC performed the analyses with assistance from SGG. JPT, VLC, and SGG wrote the manuscript.

References cited

- Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014;505:495–501.
- 2. Miki Y, Swensen J, Shattuck-Eidens D, Futreal P, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science. 1994;266:66–71.
- 3. Futreal P, Liu Q, Shattuck-Eidens D, Cochran C, Harshman K, Tavtigian S, et al. BRCA1 mutations in primary breast and ovarian carcinomas. Science. 1994;266:120–2.
- 4. Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D, Stern HM, et al. Diverse somatic mutation patterns and pathway alterations in human cancers. Nature. 2010;466:869–73.
- 5. Zang ZJ, Cutcutache I, Poon SL, Zhang SL, McPherson JR, Tao J, et al. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. Nat Genet. 2012;44:570–4.
- 6. Kumar A, White TA, MacKenzie AP, Clegg N, Lee C, Dumpit RF, et al. Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. Proc Natl Acad Sci U S A. 2011;108:17087–92.
- Zhao S, Choi M, Overton JD, Bellone S, Roque DM, Cocco E, et al. Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. Proc Natl Acad Sci U S A. 2013;110:2916–21.
- 8. Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, et al. Initial genome sequencing and analysis of multiple myeloma. Nature. 2011;471:467–72.
- 9. Wood LD, Parsons DW, Jones S, Lin J, Sjöblom T, Leary RJ, et al. The genomic landscapes of human breast and colorectal cancers. Science. 2007;318:1108–13.
- 10. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science. 2006;314:268–74.
- 11. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. Nature. 2007;446:153–8.
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499:214–8.
- 13. Evans DM, Purcell S. Power Calculations in Genetic Studies. Cold Spring Harb Protoc.

Cold Spring Harbor Laboratory Press; 2012;2012:db.top069559.

- 14. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev Camb Philos Soc. 2007;82:591–605.
- Hojat M, Xu G. A Visitor's Guide to Effect Sizes Statistical Significance Versus Practical (Clinical) Importance of Research Findings. Adv Health Sci Educ Theory Pract. Kluwer Academic Publishers; 2004;9:241–9.
- 16. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. Br Med J . British Medical Journal Publishing Group; 1986;292:746–50.
- 17. Goodman S. A dirty dozen: twelve p-value misconceptions. Semin Hematol. 2008;45:135–40.
- 18. Hayat MJ. Understanding statistical significance. Nurs Res. 2010;59:219-23.
- 19. Sullivan GM, Feinn R. Using Effect Size-or Why the P Value Is Not Enough. J Grad Med Educ. 2012;4:279–82.
- Chavalarias D, Wallach JD, Li AHT, Ioannidis JPA. Evolution of Reporting P Values in the Biomedical Literature, 1990-2015. JAMA. American Medical Association; 2016;315:1141–8.
- 21. Lui VWY, Hedberg ML, Li H, Vangara BS, Pendleton K, Zeng Y, et al. Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. Cancer Discov. 2013;3:761–9.
- 22. Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, et al. High frequency of mutations of the PIK3CA gene in human cancers. Science. 2004;304:554.
- 23. Nowell P. The clonal evolution of tumor cell populations. Science. 1976;194:23-8.
- 24. Crespi B, Summers K. Evolutionary biology of cancer. Trends Ecol Evol. 2005;20:545-52.
- 25. Cannataro VL, Gaffney SG, Stender C, Zhao Z, Philips M, Greenstein AE, et al. Heterogeneity and mutation in KRAS and associated oncogenes: evaluating the potential for the evolution of resistance to targeting of KRAS G12C. Oncogene. 2017;in press.
- Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, et al. Activating Mutations in the Epidermal Growth Factor Receptor Underlying Responsiveness of Non–Small-Cell Lung Cancer to Gefitinib. N Engl J Med. Massachusetts Medical Society; 2004;350:2129–39.
- 27. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med.

2011;364:2507–16.

- Oyama T, Harigaya K, Sasaki N, Okamura Y, Kokubo H, Saga Y, et al. Mastermind-like 1 (MamL1) and mastermind-like 3 (MamL3) are essential for Notch signaling in vivo. Development. 2011;138:5235–46.
- 29. Sasaki H, Hikosaka Y, Okuda K, Kawano O, Moriyama S, Yano M, et al. NFE2L2 gene mutation in male Japanese squamous cell carcinoma of the lung. J Thorac Oncol. 2010;5:786–9.
- 30. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 2012;489:519–25.
- 31. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin. 2017;67:7-30.
- 32. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a Shared Vision for Cancer Genomic Data. N Engl J Med. 2016;375:1109–12.
- 33. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for interfacing with genome browsers. Bioinformatics. 2009;25:1841–2.
- 34. Cao S, Wendl MC, Wyczalkowski MA, Wylie K, Ye K, Jayasinghe R, et al. Divergent viral presentation among human tumors and adjacent normal tissues. Sci Rep. 2016;6:28294.
- 35. Broad Institute. Broad Institute TCGA Genome Data Analysis Center (2016): Firehose stddata_2016_01_28 run. Broad Institute of MIT and Harvard. doi:10.7908/C11G0KM9. 2016.
- 36. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. 2016;17:31.
- 37. Gillespie JH. Population genetics: a concise guide. 2004. Baltimore, Maryland, US: The John Hopkins University Press;
- 38. Sawyer SA, Hartl DL. Population genetics of polymorphism and divergence. Genetics. 1992;132:1161–76.
- 39. Bustamante CD. Population Genetics of Molecular Evolution. Statistical Methods in Molecular Evolution. Springer New York; 2005. page 63–99.
- 40. Innan H, Kim Y. Pattern of polymorphism after strong artificial selection in a domestication event. Proc Natl Acad Sci U S A. 2004;101:10667–72.
- 41. Parsons TL, Quince C. Fixation in haploid populations exhibiting density dependence I: The non-neutral case. Theor Popul Biol. 2007;72:121–35.

- 42. Mutation Annotation Format (MAF) Specification TCGA National Cancer Institute -Confluence Wiki [Internet]. [cited 2018 Feb 7]. Available from: https://wiki.nci.nih.gov/display/TCGA/Mutation+Annotation+Format+(MAF)+Specification
- 43. Seiwert TY, Zuo Z, Keck MK, Khattri A, Pedamallu CS, Stricker T, et al. Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. Clin Cancer Res. 2015;21:632–41.

TABLES

Tables S1, S2, and S3 available as a supplement to the manuscript.

TABLE LEGENDS

Table S1: Tab delimited text file containing the selection intensity of all recurrent single nucleotide variants in our analysis. Informative columns also include the amino acid position of the variant (AA_pos), the amino acid reference (AA_Ref), the amino acid change (AA_Change), and prevalence of the variant among tumors (freq), the silent substitution rate of the variant (mu), and the MutSigCV *P* and *Q* values for that gene (MutSigCV_p and MutSigCV_q). When the variant is outside of an exon, the position is given in terms of the nucleotide position on the chromosome (Nucleotide_position), and the reference is the nucleotide reference (Nuc_Ref) and the change is the nucleotide change (Nuc_Change).

Table S2: Exome sequencing data from the Yale-Gilead collaboration displayed in accordance with Mutation Annotation Format (MAF) Specifications (42) in a tab-delimited text file.

Table S3: Metadata on the TCGA data used in this analysis in a tab-delimited text file. Includes NCI UUID for all files, and the gene expression tissue files used for each tumor type.

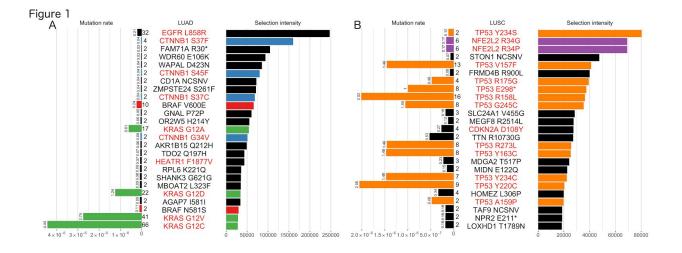
FIGURE LEGENDS

Figure 1: The 25 single-nucleotide variants with the highest selection intensity in LUAD and LUSC tumors, ranked by selection intensity, which corresponds to the length of the bars to the right of the variant names. Length of bars to the left of variant names corresponds to mutation rate, and mutation rate ($\times 10^{-6}$) is listed at the top of the bars. Prevalence of variants among 675 LUAD and 600 LUSC tumors are listed to the left of the variants. Red variant labels refer to genes deemed significantly mutated by MutSigCV (Q < 0.1) and variants occurring in genes present twice in the top 25 have uniquely colored bars. Asterisks denote nonsense mutations.

Figure 2: Cancer effect sizes of recurrent somatic substitutions in eight of the 23 cancer types analyzed. Effect sizes greater than 1×10^3 are indicated by ticks along the tumor-type axes. The highest 50 effect sizes are labeled within each tumor. Names of genes that have more than one mutation within or between tumors are uniquely colored. Genes deemed significantly burdened with mutation (12) are depicted by a red circle next to mutation labels, and the prevalence of each substitution is represented by the size of this circle. LUAD: Lung adenocarcinoma; LUSC: Lung squamous cell carcinoma. UCEC: Uterine corpus endometrial carcinoma; LGG: Brain Lower Grade Glioma; HNSC: Head and neck squamous cell carcinoma, broken into HPV positive and HPV negative tumor samples (methods); SKCMP: Skin cutaneous melanoma (primary); READ: Rectum adenocarcinoma. NCSNV refers to a non-coding single nucleotide variant outside an exon (e.g. 5' or 3' UTRs). HPV⁺ and HPV⁻ HNSCC have been demonstrated to have significantly different genetic architectures (43); thus, they are presented separately. 108 LUAD, 108 LUSC, 23 UCEC, and 47 SKCMP tumors from the Yale-Gilead collaboration are included in the plot.

Figure S1: Selection intensities of recurrent somatic substitutions in 23 cancers. BLCA: Bladder Urothelial Carcinoma; BRCA: Breast invasive carcinoma; Cervical squamous cell carcinoma and endocervical adenocarcinoma; COAD: Colon adenocarcinoma; ESCA: Esophageal carcinoma; GBM: Glioblastoma multiforme; HNSC: Head and neck squamous cell carcinoma, broken into HPV⁺ and HPV⁻ tumor samples using criteria described within the Methods section; KIRC: Kidney renal clear cell carcinoma; LAML: Acute Myeloid Leukemia; LGG: Brain Lower Grade Glioma; LIHC: Liver hepatocellular carcinoma; LUAD: Lung adenocarcinoma; LUSC: Lung squamous cell carcinoma; OV: Ovarian serous cystadenocarcinoma; PAAD: Pancreatic adenocarcinoma; PRAD: Prostate adenocarcinoma; READ: Rectum adenocarcinoma; SKCM: Skin Curaneous Melanoma, broken into primary skin tumors and metastatic skin tumors; STAD: Stomach adenocarcinoma; THCA: Thyroid carcinoma; UCEC: Uterine corpus endometrial carcinoma. 108 LUAD, 108 LUSC, 23 UCEC, 47 SKCMP, 156 SKCMM, 22 THCA, 41 LIHC, 56 CESC, and 50 PAAD tumors from the Yale-Gilead collaboration are included in the plot.

FIGURES



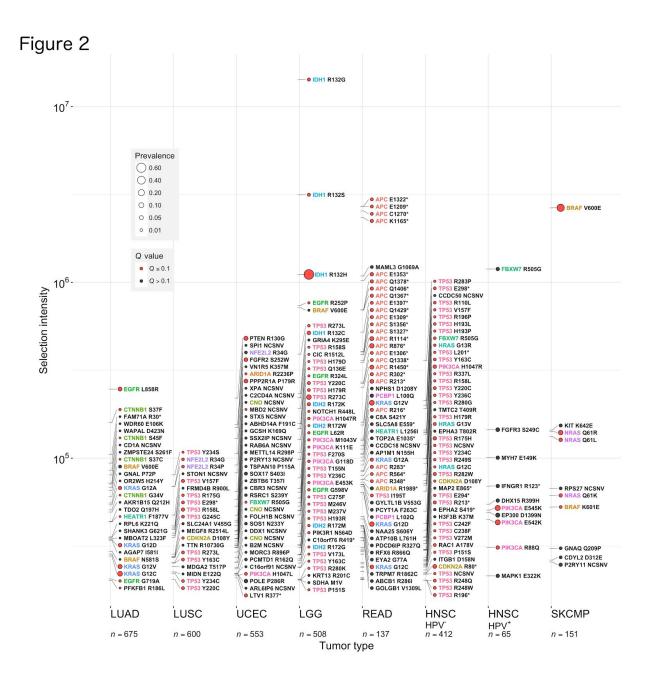


Figure S1

BRCA-	
n _{TCGA} = 986	
- LUAD n _{TCGA} = 567 n _{YG} = 108	
LUSC - <i>n</i> _{TCGA} = 492 <i>n</i> _{YG} = 108	
UCEC - <i>n</i> _{TCGA} = 530 <i>n</i> _{YG} = 23	
- SKCMP n _{TCGA} = 104 n _{YG} = 47	
SKCMM - <i>n</i> _{TCGA} = 363 <i>n</i> _{YG} = 156	AF V600E
THCA - <i>n</i> _{TCGA} = 492 <i>n</i> _{YG} = 22	
LGG - n _{TCGA} = 508	
PRAD - <i>n</i> _{TCGA} = 495	Q value
STAD-	 Q ≤ 0.1
$n_{\text{TCGA}} = 437$	 Q > 0.1
$\frac{\Theta}{1}$ $n_{TCGA} = 436$	Prevalence
$\frac{1}{100} HPV^{-} HNSC - n_{TCGA} = 412$	0.00
HPV ⁺ HNSC ⁻	
$n_{\text{TCGA}} = 65$	O 0.10 O 0.05
BLCA - <i>n</i> _{TCGA} = 412	o 0.01
LIHC - <i>n</i> _{TCGA} = 364 <i>n</i> _{YG} = 41	
COAD - <i>n</i> _{TCGA} = 399	
GBM - <i>n</i> _{TCGA} = 393	
СЕЅС - n _{тсбА} = 289 n _{YG} = 56	
KIRC - <i>n</i> _{TCGA} = 336	
PAAD-	
n _{TCGA} = 178 n _{YG} = 50 bioRxiv preprint doi: https://doi.org/10.1101/229724; this version posted February 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.	
$n_{\text{TCGA}} = 184$ LAML -	
$n_{TCGA} = 143$ READ-	
n _{TCGA} = 137	



		2				
	MINA NCSNV PIK3CA H1047R C10orf76 Q267K USP40 NCSNV SF3B1 K700E	 EXOC4 S21L TP53 Y220S FAM189A2 I241I TP53 R273L PDE4DIP NCSNV KIAA0528 R181Q PLEKHM1 NCSNV 	 LDHALGB NCSNV TPS3 R342P MYLK3 NCSNV TPS3 1195T PIK3CA H1047L PIK3CA H1047L ZMYM1 NCSNV ZMYM1 NCSNV TP53 F113C CNTNAP3 NCSNV MLL3 E640* MLL3 E640* ZNF692 NCSNV MYL5 R46P MYL5 R46P VPS13C T3732M 	FRMD7 NCSNVTPS3 L194RUSO1 E610QAKAP9 NCSNVHEATR6 L162PPPARG L242VC10nf129 NCSNVC10nf129 NCSNVC10nf129 NCSNVC10nf129 NCSNVC10nf129 NCSNVC10nf129 NCSNVC10nf129 NCSNVTP53 H193RTP53 H193RTP53 H179RPIK3CA N345KZNF611 NCSNV	SEC31A R242L ALDOA NCSNV SUPTGH D130Y PIK3CA E545A C12orf69 NCSNV MYO15A K2535M MYO15A K2535M AKT1 E17K STAR NCSNV STAR NCSNV	
	EGFR L858R	CTNNB1 S37F FAM71A R30* WDR60 E106K WAPAL D423N	CTNNB1 S45F CD1A NCSNV ZMPSTE24 S261F ZMPSTE24 S261F CTNNB1 S37C BRAF V600E GNAL P72P GNAL P72P GNAL P72P GNAL P72P CRNB1 634V KRAS G12A CTNNB1 G34V AKR1B15 Q212H TD02 Q197H TD02 Q197H	RPL6 K221Q SHANK3 G621G MBOAT2 L323F MBOAT2 L323F KRAS G12D AGAP7 I5811 AGAP7 I5811 AGAP7 I5811 BRAF N5815 RAS G12V KRAS G10V KRAS G10V	ARAH SCL9 SCL9 SCL9 SCL9 SCL9 SCL9 SCL9 SCL9	
	Ψ	P53 Y234S			P53 NC SNV P53 NC SNV QSEC2 G613C P53 N131 P53 T125T P53 T125T P53 N131 P53 T125T P53 T125T <td></td>	
	EN R130G		ALA NCSNV NCSNV 2 NCSNV 5 NCSNV 5 NCSNV H K169Q H K169Q CIP NCSNV 6A NCSNV 711 15A AN10 P115A AN10 P115A AN10 P115A AN11 P115A			
		642E S Q61R S Q61L S Q61L	Z7 NCSNV ABH STX SXX SXX ABH ABH ABH CNO CNO CNO CNO CNO CNO CNO CNO CNO CNO	CB CB CB CB CB CB CB CB CB CB CB CB CB C	91 D723N 1L2 S1472L R342* 5A3 NCSNV 9 NCSNV 27 NCSNV NCSNV NCSNV NCSNV NCSNV NCSNV NCSNV NCSNV 1199 G722E 1199 G722E 1199 G722E 1199 G722E 1199 G722E 1199 G722E 1199 G722E 1199 G722E 1199 G722E 118716 118716 118717 1 A238V 1 A238V	AMY1C K293K RAC1 P29S NSFL1C NCSNV MAGEA11 S208L DHTKD1 P842L SERPINA1 NCSNV SERPINA1 NCSNV ALPPL2 G448R GDF2 F336F REG1B NCSNV REG1B NCSNV REG1B NCSNV REG1B NCSNV REG1B NCSNV SYDE1 F703F NPHP1 R347* BUB1 R677C SLC4A10 R410C SLC4A10 R410C
SUCEP 1268A EPRS 1820L DHX32 R51S ANO9 D473A SYTL2 NCSNV ARHGEF2 R609G	ZNF830 K2395 FAM120A 1302K TM2D3 Y136S KIAA0100 R952L PPAT E11D SCUBE3 L427V CDK4 R24S UCKL1 D473H CPSF1 Y1164H	91353M 91353M 51A V31D A435G HG4B D330A HG4B D330A SP1 D145V SP1 D145V Z R216G CSNV		TRPV2 L127V CTRB2 NCSNV MTMR6 F75C GNA GNA P2R' P2R' D2R' D2R' D2R' D2R' D2R' D2R' D2R' D		
					A251A M251A	
 INTS2 Y578N SPACA3 NCSNV EIF1AX NCSNV 	MARCH4 NCS GNB2L1 NCS BRAF K601E				2D 63K 63K 019 0R52E2/	
	GRIA4 K295					1 10 1 11 1 NUTUR 1 1 FT 11 NT 1 11
SPOP F133C	CUL3 M305R CUL3 M305R SPOP F102C	SPOP Y87N		╸	 MED12 L1224F MLL3 Y4440C ACP1 NCSNV PK3CA H1047R PK3CA H1047R PK3CA E542K PK3CA E545K PK5CA F175H PK77 P147L 	
		ARID5B A1009G	 LHCGR NCSNV MACC1 NCSNV MACC1 NCSNV RHOA Y42S VM01 NCSNV GABRE NCSNV GABRE NCSNV TTN NCSNV RNF175 S215G PIK3CA H1047R 	PIK3CA N345K ZC3H4 E779Q ZC3H4 E779Q CDH1 D254Y ZBTB7A K424T DEK S288T OLIG2 S88A OLIG2 S88A IQSEC2 G283A IQSEC2 G283A IQSEC2 G283A IQSEC2 G283A CLIG2 S88A IQSEC2 G283A CLIG2 S88A IQSEC2 G283A CSEC2 G283A CSEC3 C176F CSE L138R CSB L138R CSC CSB L138R CSB L	TPS3 Y220C TGB1 L378 INSIG2 I63F INSIG2 I63F CAV1 NCSNV C1orf9 NCSNV C10FBL5 NCSNV MLH1 R659* RPL5 M208T CHL5 NCSNV NT1 NCSNV NALCN NCSNV NALCN NCSNV NALCN NCSNV	GART Q241R STAG1 A116T TP53 V272M
MGRN1 G64A COL14A1 NCSNV BDNF NCSNV PIK3CD V233G TOPBP1 NCSNV CASD1 NCSNV	 TPS3 V157F TPS3 1251S CCDC66 Q379E CCDC66 Q379E TPS3 1195T RSF1 NCSNV RSF1 NCSNV RPS16 NCSNV PDILT Y476* 	GAE FRM MAF MAF PIK3 MAA CYP	TP53 Y163C TP53 Y220C SYTL4 NCSNV TP53 R175H LDHAL6B NCSNV UXS1 V105L UXS1 V105L TP53 R248Q TP53 R248Q TP53 R248Q NRIP1 NCSNV NRIP1 NCSNV TP53 R248W RP53 R248W RP53 R248W	TF M45I FBXO10 S950T UTP20 L384H PNN S443A LRRC16B T804P UBE3C NCSNV UBE3C NCSNV SORBS1 NCSNV NRAS Q61R NRAS Q61R TP53 Y234C TP53 S241F TP53 S241F TP53 S241F TP53 S241F TP53 S241F TP53 C192* C1orf129 NCSNV	1718 1718	
TP53 R283P	TPS3 E298* CCDC50 NCSNV TPS3 R110L TPS3 V157F	TP53 R 196P TP53 H 193L TP53 H 193P FBXW7 R 505G HRAS G 13R HRAS G 13R TP53 L 201* TP53 Y 163C	PIK3CA H1047R TP53 R337L TP53 R158L TP53 R158L TP53 Y220C TP53 Y220C TP53 Y236C TP53 Y236C TP53 R109R TP53 R179R HRAS G13V EPHA3 T802R TP53 R175H TP53 R175H	TP53 Y234C HRAS G12C TP53 R249S TP53 R249S TP53 R282W CDKN2A D108Y MAP2 E865* TP53 E294* TP53 E294* TP53 E294* TP53 E294* TP53 E294* TP53 C242F H3F3B K37M TP53 C242F TP53 C242F TP53 C238F TP53 C238F	TICB1 D158N	
P		GFR3 S249C	rrth E149K	MARCA E542K		
W7 R425G		G11B R257P AS G13R 3I R77P C1D12 NCSNV 53 D259V BEP2 V352L FR3 S249C	CG R4713L RES1 G133R BB3 V104L DC4 S21* DC4 S21* DC4 S21* 2A NCSNV 2A NCSNV 2A NCSNV 2A NCSNV 22 R34G 22 R34G 21 R3 269Q R83 E69Q R83 E69Q R84 E71Q R84 E71Q R	CAE H1047L CAE K905N S3 L257Q VCR2 NCSNV AS G12D R1 R13P R1 R12P R1 R12P R12P R12P R12P R12P R12P R12P R12P	orf16 E268K AS Q61K D3 NCSNV FR3 G50E FR3 G50E FR3 G370C D2 NCSNV D2 NCSNV MCSNV MCSNV MP1 NCSNV MF1 R341G MF1 R341G	
2 E378D 2 E378D 64 E368* 64 E368* 3C P201Q APB R611P APB R611P	D4097H rf39 V463L NCSNV B1 H36P R342G 1 S22W 6 D441H NCSNV	VBI S45P 2 R569H VBI S37A VBI S37A KEL C23W A1 N405T A1 A1 A	(B1 K335) BIR (93A L852F 50 (171L 50 (277* 50 (171L) 50	A37P R249S V157F V157F V157F C1 P417A R248L R248L B10 D120H G3 K305* G3 K305* G3 K305* G1 P417A G3 K305* G1 P417A G3 K305* G1 P417A G1 P41		
159P CTNNI CTNNI CTNNI CTNNI FAM13 FAM13	C146	R269G CTNN R269G CTNN (357M ATP1/ CTNN	112V CTNN 112D CTNN 112D CTNN 112D CTNN 112D CDF2 370 [*] FAM1 12D CTNN 12D CTNN 12D CTNN 12D CTNN 10 [*] FAM1 10 [*] FGD6 FAM1 10 [*] FGD6 FCD7 FCD7 FGD6 FCD7 FCD7 FCD7 FCD7 FCD7 FCD7 FCD7 FCD7	312R GJCf 2 NCSNV GJCf 2 NCSNV TPS 1 V1166D TPS NCSNV SPHK 12C TPS NCSNV SPHK 12C TPS NCSNV SPHK 12C TPS 8S UGT2 NCSNV PLA2 8S UGT2 8S UGT2 8S UGT2 7992A DLA2 7344G USP46 13D USP46 13D CTNN	IY NCSNV 5 D261E 397* 397* 513R K396N K396N K396N S46* 35* 35* 35* 35* 35* 141* 141* 141* 141* 171A 171A 171A 171A 171A 22 H30Q 22 H30Q	
DTP53 A1		ABCF3 F ABCF3 F SR	62P 62P 62P 55 55 55 55 55 55 55 55 55 55 55 55 55		CSNV CSNV CSNV CSNV CSNV CSNV CSNV CSNV	
	Ş	•		1976 GABRAG V 280 SIDT2 A632 280 SIDT2 A632 167P RGAG1 NC SNV FF53 H179F A3K FF53 H179F VCSNV FF53 H179F VCSNV FF53 H179F VCSNV FF53 H179F VCSNV FF53 NCSI VCSNV FF53 NCSI VCSNV FF66 NCSN NV FF66 NCSN ST FF66 NCSN<		Image: Second state Image: Second state
OD2 P177R	STMN3 NCSh	ALDOA	A ZNF335 NCSNV MED12 D23Y ANKRD20A4 NC ANKRD20A4 NC GABPA NCSNV HNRNPR NCSNV HNRNPR NCSNV THAP1 NCSNV FBXW7 R479G SIGLEC7 Q309R TP63 K188M TP63 K188M CELF4 NCSNV TCF20 NCSNV	CACNG7 V19 C12orf43 E28 TRMT2A R16 LMNB2 NCSN RBM39 NCSN RBM39 NCSN ZNF658 N443 ZNF658 N443 FAM155A NC ST6GAL1 NC ST7 ST6GAL1 NC ST6GAL1 NC ST7 ST6GAL1 NC ST6GAL1 NC ST6GAL1 NC ST6GAL1 NC ST6GAL1 NC ST6GAL1 NC ST6GAL1 NC ST7 ST7 ST6GAC ST7 ST7 ST6GAC ST7 ST7 ST7 ST7 ST7 ST7 ST7 ST7 ST7 ST7	HIST 1H2AL R FAT2 NCSNV FRT2 NCSNV PRR22 NCSN PRR22 NCSN FRAS G12V FRAS G12V	
	833	SETD2 W1782	SIP/ SIP/ SIP/ SIP/ SIP/ SIP/ SIP/ SIP/	SPSB3 G241R VHL H115N HNF1B N302K VHL L128H VHL L128H VHL L128H VHL NCSNV VHL NCSNV VHL L130F VHL L130F VHL L130F VHL L158V BAK1 R76W VHL N131K VHL N131K VHL N131K VHL N131K	 VHL R161* MTOR T1977R ML W88* ALLC T83M ALLC T83M BAP1 R710* GPM6A A50V BAP1 M11 GALNT6 P311P EDIL3 T343M PBRM1 R850* 	O NXPH4 K272K
	TP53 V143E	EEF2 Y457*		 TPS3 Y220C TPS3 C275F TPS3 H179R TPS3 H179R TPS3 H179R TPS3 E294* IGFN1 S1976S ATP1B1 E154E ATP1B1 E154E 	 CDKNZA H83Y TP53 R248W TP53 R248W TP53 R248W SMAD4 S144* SMAD4 S144* SMAD4 S144* SMAD4 S144* SMAD4 S144* SMAD4 C289* TP53 R175H ABCA13 A2368V CDKNZA R80* SMAD4 C289* TP53 R175H CDKNZA R80* SMAD4 C289* TP53 R175H CDKNZA R80* SMAD4 C289* TP53 R175H CDKNZA R80* SMAD4 R361C TP53 R196* SMAD4 R361C TP53 R196* SMAD4 R361C TP53 R196* SMAD4 R361C TP53 R196* TP53 R337C TP53 R337C TP53 R306* 	COL5A1 P657L COL5A1 P657L PLEKHA4 R225W TP53 R273H ABCG4 T494M SLC12A6 R319H TGFBR1 R487Q EHBP1L1 A977A ODZ3 R2228H TTN N26927N CDKL4 V106I ERBB2 V842I RSPO1 A146A
	TP53 NCSNV PIK3CA H1047L ARF6 NCSNV CDC42BPA P675T- SUPT6H D130Y	ST6GAL1 NCSNV RNF4 NCSNV TP53 P278A TP53 C135F TP53 Y220C STRN3 NCSNV SLC38A11 NCSNV	TP53 R248Q RBBP8 G810V TP53 C275F TP53 C176F RERE NCSNV TP53 E198* TP53 E198* TP53 E198* TP53 E198* TP53 E198* TP53 R175H AMPH NCSNV SLITRK6 NCSNV TP53 R248W TTN T22904P	PIK3CA E545K TRIM13 NCSNV KPNB1 L552F NTRK3 NCSNV ABCB1 K915T TP53 R213* DNAH10 D1128Y PHACTR1 NCSNV PHACTR1 NCSNV CBE1 F551V HDAC4 F746L TP53 R282W GBE1 F551V HDAC4 F746L TP53 R273C C10orf76 Q267K ERBB2 V777L MBTP51 L77M	HEATR7B2 K1496T ELAVL3 NCSNV TP53 R196* ASPH NCSNV SNAI1 NCSNV SNAI1 NCSNV	
RAS G12V MS4 V86L T3 D835Y KRT E948G	HI R132C	RHGAP30 H1042H H2 R140Q ACNG7 V197G LT3 D835E LOX12B N670T AS G13D ALNT8 NCSNV AC2 NCSNV	HI R132H NMT3A R882C TA2 NCSNV TA2 NCSNV DHHC13 S242G RDM2 M1397L VMT3A R882H 3AR1 1116T 3AR1 1116T 3AR1 1116T 3AR1 1116T	APH1 K296M H2 R172K B17 W38R SB17 W38R SB17 W38R SB17 W38R SB17 W38R SBP4 E330K		
APC Q1429* APC E1309* APC S1356* APC S1356* APC S1327* APC S1327* APC S1327* APC S1327* APC S1326* APC S1326* APC S1356* APC S1327* APC S1356* APC S1356* APC S1327* APC S1356* APC S1327* APC S1327* APC S1356* APC S1327* APC S1327* APC S1356* APC S1327* APC S1356* APC S1327* APC S1356* APC S156* APC S16*	K8/6"	P1 L100Q S G12V S G12V S 21F S421Y 5A8 E559* IR1 L1256I A E1035* A EL	C 18 NCSNV M1 N155H S G12A R283* R283* R564* R348* R34	25 S606Y 10B L761H D6IP R327Q 6 R866Q 2 G77A 2 G77A 3 G12C M7 R1862C M7 R1862C M8 R1862C M8 R1862C M7 R1862C M8 R186		
		1	05		10 ⁴	10 ³
selec	ction int	lensitv				

Selection intensity