Effect sizes of somatic mutations in cancer

Vincent L. Cannataro[1], Stephen G. Gaffney[1], Jeffrey P. Townsend[1,2,3†]

[1]Department of Biostatistics, Yale University, New Haven, CT,
[2]Program in Computational Biology and Bioinformatics and
[3]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT

[†]To whom correspondence should be addressed.

Cancer growth is fueled by genomic alterations that confer selective advantage to somatic cells [1]. A major goal of cancer biology is determining the relative importance of these alterations. Genomic tumor sequence surveys have frequently ranked the importance of genetic substitutions to cancer growth by *P* value or a false-discovery conversion thereof [2,3]. However, *P* values are thresholds for belief [4], not metrics of effect [5,6]. Their frequent misuse as metrics of effect has often and ineffectively been vociferously decried [5,7–9], even in cases when the only attributable mistake was omission of effect sizes [10,11]. Here, we draw upon an understanding of the development of cancer as an evolutionary process [12,13] to estimate the effect size of somatic variants. We estimate the effect size of all recurrent single nucleotide variants in 23 cancer types, ranking their relative importance within and between driver genes. Many of the variants with the highest effect size per tumor, such as EGFR L858R in lung adenocarcinoma and BRAF V600E in colon adenocarcinoma, are within genes deemed significantly mutated by existing whole-gene metrics. Quantifying the effect sizes of somatic mutations underlying cancer has immediate significance to the prioritization of clinical decision-making by tumor boards, selection and design of clinical trials, pharmacological targeting, and basic research prioritization.

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 [14], 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 [15,16], the statistical significance (*P* value) of the disproportionality of mutation frequency [2], or both [14]. Versions of these metrics have shifted from simple ranks by mutation prevalence in a tumor population [17–19] to calculation of statistical significance of mutation prevalence over genome-wide context-specific background mutation rates [20–22], to ratios of the prevalence of nonsynonymous and synonymous mutations [1,23], to *P* values based on a gene-specific mutation rate and a diversity of genomic data [3]. Although the approaches used to calculate *P* values have become more sophisticated, neither prevalence nor *P* value is an appropriate metric for quantifying the vital role of genes or their mutations to tumorigenesis and cancer development.

While prevalence of a somatic mutation in a cancer type has important consequences for biomarker studies [24] and identification of therapeutic population for a targeted therapy [25], 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 [1,3]—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 [3].

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 [5,6]. 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 [3], 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 evolutionary process [12,13], 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.

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 [26]. 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 [27], 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).

This calculation yielded cancer effect sizes for all fixed substitutions (Methods, Supplemental Figure 1, Supplemental Table 1) 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 COAD and EGFR L858R in LUAD, 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 determined as significantly mutated by MutSigCV[3]. However, genes determined to be significantly altered in cancer by MutSigCV are well-dispersed within a large range of site-specific cancer effect sizes (Figure 1), illustrating how discrepant *Q* values are with cancer effect size. Several substitutions within genes that are not determined to be significantly over-mutated via MutSigCV 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 (Figure 1).

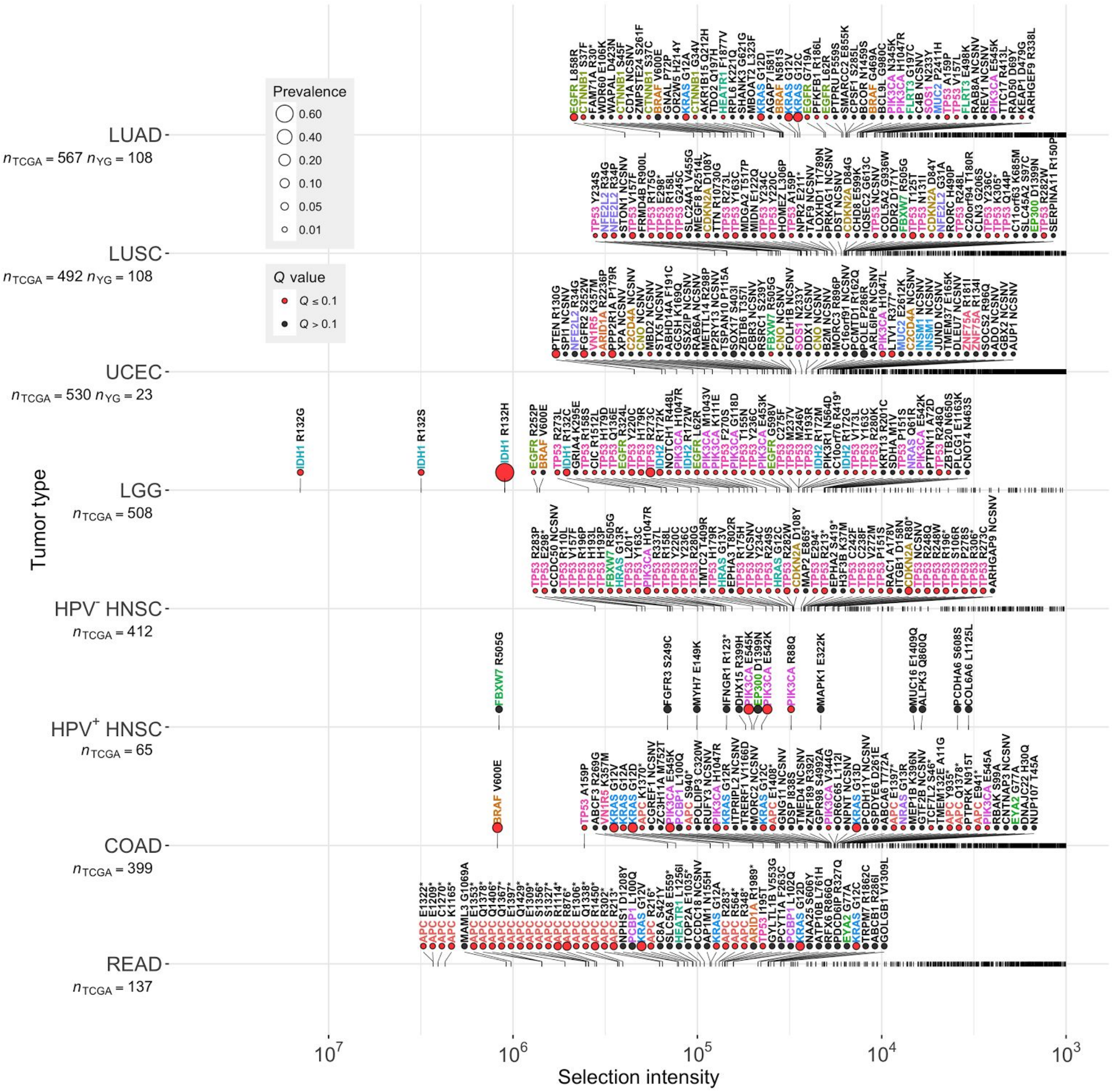


Figure 1. Cancer effect sizes of recurrent somatic substitutions in eight of the 23 cancer types analyzed. Effect sizes greater than 1 × 10^3 are indicated by ticks along the tumor-type axes. The highest 50 effect sizes are labeled within each tumor. Names of genes that have more than one mutation within or between tumors are uniquely colored. Genes deemed significantly burdened with mutation [3] 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); COAD: Colon adenocarcinoma; 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 [31], thus they are presented separately.

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 BRAF V600E was estimated to have the fifth highest selection intensity in LGG, but both of these genes were classified as not significantly mutated by `MutSigCV` within these two 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 degree of influence within HPV+ HNSCC and LGG is obscured in cancer-wide significance analyses. The relative effect sizes of cancer mutations can inform nearly every aspect of basic research related to oncology and should play key roles in clinical decision-making. They should be integrated into clinical decision-making in tumor boards, where they would provide crucial insight into the relative upper limits of the efficacy of precision-targeted treatments. They should guide the selection of clinical trials to target small populations that can benefit from targeted therapeutics developed for other cancer types. They should guide targeted development of new therapeutics, indicating the upper limit of effect for a perfect therapeutic ameliorating an oncogenic mutation. Lastly, they should guide the selection of important areas of basic research that have potential to lead to therapies and cures for cancer. 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:

All data were obtained either from The Cancer Genome Atlas (TCGA) projects downloaded from the National Cancer Institute's Genomic Data Commons [32], or from projects part of the Yale-Gilead collaboration (Supplemental Table 2). All TCGA data used in this analysis were from GDC version gdc-1.0.0 and relevant UUID are found in Supplemental Table 3.

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 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 (Supplemental Table 3) 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, 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 [37–40]. 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 $\left(e^{-\lambda}\right)^{n_0} \times \left(1 - e^{-\lambda}\right)^{n_1}$, where n_0 is the number of tumors without any substitution at that site and n_1 is the number of tumors with the specified substitution at that site. 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

References cited

1. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153–158 (2007).
2. Kan, Z. *et al.* Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466,** 869–873 (2010).
3. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
4. Evans, D. M. & Purcell, S. Power Calculations in Genetic Studies. *Cold Spring Harb. Protoc.* **2012,** db.top069559 (2012).
5. Nakagawa, S. & Cuthill, I. C. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol. Rev. Camb. Philos. Soc.* **82,** 591–605 (2007).
6. 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.* **9,** 241–249 (2004).
7. Gardner, M. J. & Altman, D. G. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br. Med. J.* **292,** 746–750 (1986).
8. Goodman, S. A dirty dozen: twelve p-value misconceptions. *Semin. Hematol.* **45,** 135–140 (2008).
9. Hayat, M. J. Understanding statistical significance. *Nurs. Res.* **59,** 219–223 (2010).
10. Sullivan, G. M. & Feinn, R. Using Effect Size-or Why the P Value Is Not Enough. *J. Grad. Med. Educ.* **4,** 279–282 (2012).

11. Chavalarias, D., Wallach, J. D., Li, A. H. T. & Ioannidis, J. P. A. Evolution of Reporting P Values in the Biomedical Literature, 1990-2015. *JAMA* **315,** 1141–1148 (2016).
12. Nowell, P. The clonal evolution of tumor cell populations. *Science* **194,** 23–28 (1976).
13. Crespi, B. & Summers, K. Evolutionary biology of cancer. *Trends Ecol. Evol.* **20,** 545–552 (2005).
14. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505,** 495–501 (2014).
15. Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266,** 66–71 (1994).
16. Futreal, P. *et al.* BRCA1 mutations in primary breast and ovarian carcinomas. *Science* **266,** 120–122 (1994).
17. Zang, Z. J. *et al.* Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat. Genet.* **44,** 570–574 (2012).
18. Kumar, A. *et al.* Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 17087–17092 (2011).
19. Zhao, S. *et al.* Landscape of somatic single-nucleotide and copy-number mutations in uterine serous carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 2916–2921 (2013).
20. Chapman, M. A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471,** 467–472 (2011).
21. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318,** 1108–1113 (2007).
22. Sjöblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314,** 268–274 (2006).
23. Krauthammer, M. *et al.* Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat. Genet.* **44,** 1006–1014 (2012).

24. Lui, V. W. Y. *et al.* Frequent mutation of the PI3K pathway in head and neck cancer defines predictive biomarkers. *Cancer Discov.* **3,** 761–769 (2013).

25. Samuels, Y. *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304,** 554 (2004).

26. Cannataro, V. L. *et al.* Heterogeneity and mutation in KRAS and associated oncogenes: evaluating the potential for the evolution of resistance to targeting of KRAS G12C. *Oncogene* **in press,** (2017).

27. Gillespie, J. H. Population genetics: a concise guide. 2004.

28. Oyama, T. *et al.* Mastermind-like 1 (MamL1) and mastermind-like 3 (MamL3) are essential for Notch signaling in vivo. *Development* **138,** 5235–5246 (2011).

29. Sasaki, H. *et al.* NFE2L2 gene mutation in male Japanese squamous cell carcinoma of the lung. *J. Thorac. Oncol.* **5,** 786–789 (2010).

30. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489,** 519–525 (2012).

31. Seiwert, T. Y. *et al.* Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas. *Clin. Cancer Res.* **21,** 632–641 (2015).

32. Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **375,** 1109–1112 (2016).

33. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics* **25,** 1841–1842 (2009).

34. Cao, S. *et al.* Divergent viral presentation among human tumors and adjacent normal tissues. *Sci. Rep.* **6,** 28294 (2016).

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, B. S. & Swanton, C. DeconstructSigs:

delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17,** 31 (2016).

37. Sawyer, S. A. & Hartl, D. L. Population genetics of polymorphism and divergence. *Genetics* **132,** 1161–1176 (1992).

38. Bustamante, C. D. Population Genetics of Molecular Evolution. in *Statistical Methods in Molecular Evolution* 63–99 (Springer New York, 2005).

39. Innan, H. & Kim, Y. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 10667–10672 (2004).

40. Parsons, T. L. & Quince, C. Fixation in haploid populations exhibiting density dependence I: The non-neutral case. *Theor. Popul. Biol.* **72,** 121–135 (2007).

Supplementary Information available online.

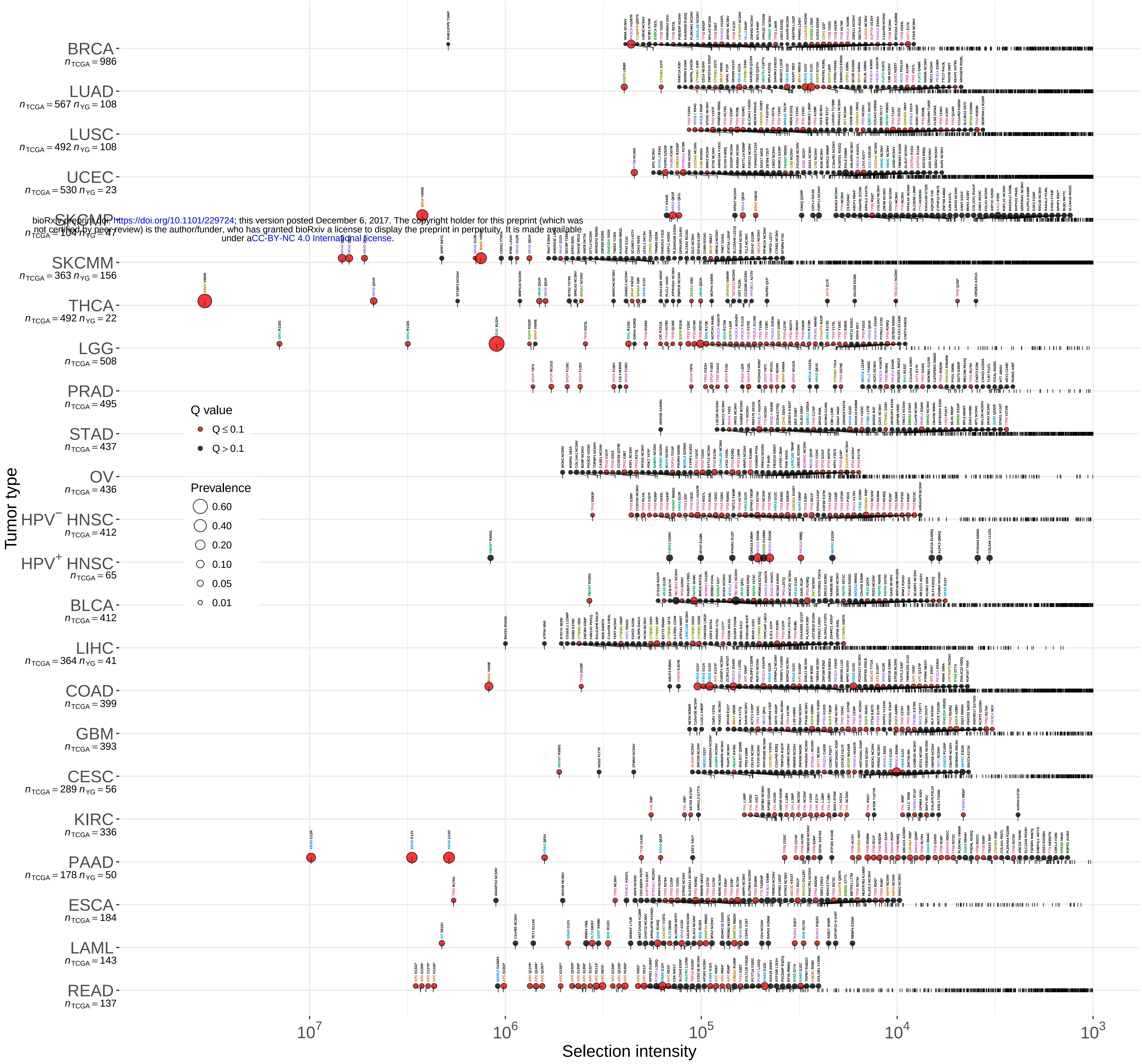
Acknowledgements: This project was supported by Gilead Sciences, Inc. We also thank the Yale High Performance Computing Center for providing computational resources.

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

Author Information: The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to jeffrey.townsend@yale.edu.

Supplemental figure 1 legend: 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.

Supplemental Figure 1


Tumor type
BRCA
$n_{TCGA} = 986$
LUAD
$n_{TCGA} = 567$ $n_{YG} = 108$
LUSC
$n_{TCGA} = 492$ $n_{YG} = 108$
UCEC
$n_{TCGA} = 530$ $n_{YG} = 23$
SKCMP
$n_{TCGA} = 104$ $n_{YG} = 47$
SKCMM
$n_{TCGA} = 363$ $n_{YG} = 156$
THCA
$n_{TCGA} = 492$ $n_{YG} = 22$
LGG
$n_{TCGA} = 508$
PRAD
$n_{TCGA} = 495$
STAD
$n_{TCGA} = 437$
OV
$n_{TCGA} = 436$
HPV^{-} HNSC
$n_{TCGA} = 412$
HPV^{+} HNSC
$n_{TCGA} = 65$
BLCA
$n_{TCGA} = 412$
LIHC
$n_{TCGA} = 364$ $n_{YG} = 41$
COAD
$n_{TCGA} = 399$
GBM
$n_{TCGA} = 393$
CESC
$n_{TCGA} = 289$ $n_{YG} = 56$
KIRC
$n_{TCGA} = 336$
PAAD
$n_{TCGA} = 178$ $n_{YG} = 50$
ESCA
$n_{TCGA} = 184$
LAML
$n_{TCGA} = 143$
READ
$n_{TCGA} = 137$
Q value
● Q ≤ 0.1
● Q > 0.1
Prevalence
○ 0.60
○ 0.40
○ 0.20
○ 0.10
○ 0.05
○ 0.01
10^7
10^6
10^5
10^4
10^3
Selection intensity
RAB11FIP5 T246P
BRAF V600E
IDH1 R132H
KRAS G12R
KRAS G12V
KRAS G12D
KRAS Q61H
FBXW7 R505G
NRAS Q61R
EGFR L858R
PIK3CA H1047R
PIK3CA E545K
TP53 R175H
APC R1450*