The whole-genome panorama of cancer drivers

Radhakrishnan Sabarinathan[1,2,*], Oriol Pich[1,2,*], Iñigo Martincorena[3], Carlota Rubio-Perez[1,2], Malene Juul[4], Jeremiah Wala[5], Steven Schumacher[5], Ofer Shapira[5], Nikos Sidiropoulos[6], Sebastian M. Waszak[7], David Tamborero[1,2], Loris Mularoni[1,2], Esther Rheinbay[5,8], Henrik Hornshøj[4], Jordi Deu-Pons[1,2], Ferran Muiños[1,2], Johanna Bertl[4], Qianyun Guo[9], Joachim Weischenfeldt[6,7], Jan O. Korbel[7,10], Gad Getz[5,8,11,12], Peter J. Campbell[13], Jakob S. Pedersen[4], Rameen Beroukhim[5], Abel Gonzalez-Perez[1,2], Núria López-Bigas[1,2,14,†], on behalf of the PCAWG Drivers and Functional Interpretation Group and the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Network

Affiliations:

1. Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac, 10, 08028 Barcelona, Spain.
2. Research Program on Biomedical Informatics, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain.
3. Wellcome Trust Sanger Institute, Hinxton CB10 1SA, Cambridgeshire, UK.
4. Department of Molecular Medicine (MOMA), Aarhus University Hospital, Aarhus, DK-8000, Denmark.
5. Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA.
6. BRIC/Finsen Laboratory, Rigshospitalet, Copenhagen, 2200, Denmark.
7. European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany.
8. Massachusetts General Hospital, Center for Cancer Research, Charlestown, Massachusetts 02129, USA.
9. Bioinformatics Research Centre (BiRC), Aarhus University, Aarhus, Denmark.
10. European Molecular Biology Laboratory (EMBL), EMBL-European Bioinformatics Institute, Hinxton, UK.
11. Massachusetts General Hospital, Department of Pathology, Boston, Massachusetts 02114, USA
12. Harvard Medical School, Boston, 02215, USA.
13. Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA UK.
14. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

* These authors contributed equally to the work.

†Corresponding author. E-mail: nuria.lopez@irbbarcelona.org

SUMMARY

Somatic genetic mutations have been linked to tumorigenesis for more than 30 years. However, only partial insights into their contribution to tumor development have been possible to date. Using a comprehensive approach, here we obtained a whole-genome panorama of driver events across more than 2,500 tumors from 37 types of cancer. This panorama includes coding and non-coding point mutations, copy number alterations and other genomic rearrangements of somatic origin, and potentially predisposing germline variants. We demonstrate that genomic events are at the root of virtually all tumors, and each carry on average 4.6 driver events. Despite several orders of magnitude of variation of tumor mutation burden, the number of driver point mutations is strikingly stable. Twenty-two percent of tumors contain driver non-coding mutations and half of all cancer genes undergo various types of mutations. In summary, the panorama described here provides answers to fundamental questions in cancer genomics.

Tumors arise from genomic mutations, often referred to as 'drivers', that confer proliferative advantage to a cell[1]. Various classes of mutation, such as point mutations (single nucleotide variants, or SNVs, multinucleotide variants, or MNVs, and indels), balanced genomic rearrangements (GRs), and somatic copy number alterations (SCNAs)[2], can drive tumorigenesis. Although some driver mutations are inherited in the germline, most are acquired by somatic cells during tumor development. Identifying drivers from the myriad of 'passenger' mutations present in tumor genomes has been a major focus of cancer genomics over the last 20 years. Recently, methods aimed at identifying signals of positive selection in the pattern of tumor mutations, with the assumption that cancer development follows a Darwinian evolutionary process, have succeed in detecting genes affected by driver mutations[3–7].

Despite enormous advances in cancer genomics in recent years, only partial views of the contribution of genomic mutations to tumorigenesis have been obtained, leaving fundamental questions unanswered. It is unclear, for instance, whether tumors of all cancer types are caused primarily by genomic driver events, and the relative contributions of point mutations and structural variants (SVs) to the emergence of cancer remain unknown. Another open question concerns the number of driver mutations required to make a somatic cell become malignant and how this number differs across tumors. Moreover, beyond point mutations in the promoters of *TERT*[8,9] and a few other genes[10] and some instances of promoter and enhancer hijacking[11–13], we have only patchy insight into the contribution of non-coding mutations to tumorigenesis.

The characterization of more than 2,500 tumors by the International Cancer Genomics Consortium's (ICGC) Pan-Cancer Analysis of Whole Genomes (PCAWG) provides an unprecedented opportunity to obtain a comprehensive whole-genome view of driver events in cancer. Here, we set the goal of identifying all driver

events (point mutations, SCNAs, GRs, and potential predisposing germline variants) across the whole genome of each tumor in the PCAWG cohort. We call this list of driver events the per-patient panorama of driver mutations (or simply panorama). We designed an incremental approach to disclose the panorama of driver mutations (Fig. 1; Methods and Suppl. Notes 1 and 2) that exploits the power of this cohort to discover novel drivers, both coding and non-coding, and the knowledge accumulated from decades of research into cancer genetics[2]. We found driver mutations in virtually all tumors, thereby providing definitive evidence of the oft-quoted maxim that cancer is fundamentally a genomic disease. We also demonstrated the presence of a small number of driver mutations in each tumor –4.6 on average, a number that is relatively stable regardless of huge variations in mutational burden. While the contribution of point mutations and SVs to tumorigenesis across cancer types differs, their relative proportions across the entire cohort are very similar. Our analysis also reveals that up to 22% of tumors contain non-coding driver point mutations, with large differences between tumor types.

Patient-centric catalog of driver point mutations

To obtain a whole-genome panorama of driver mutations, we first focused on identifying driver point mutations –SNVs, MNVs, and short indels– in each individual tumor. This is still an unsolved problem, despite the development in recent years of methods to detect cancer genes[4–7] and their application to cohorts of most major tumor types. We start by focusing on point mutations affecting coding and non-coding genomic elements (GEs; protein-coding genes, promoters, 5'UTRs, 3'UTRs, and enhancers) with signals of positive selection across the PCAWG cohort (Discovery of Drivers Elements, Fig. 1a, left panel) –or discovery GEs. While this reduces the problem, two important challenges persist.

First, not all point mutations in these discovery GEs are driver mutations. This is revealed from the comparison of the computed expected rate of passenger point mutations (background) and the observed excess of point mutations above it –henceforth referred to as excess[14,15] (Fig. 1b; Extended Data Fig. 1a). This excess constitutes the fraction of driver point mutations of each GE in the cohort. For example, while virtually all point mutations of *TP53* across breast adenocarcinomas (Breast-AdenoCA) are likely to be drivers (~100% computed excess), the excess computed for *NF1* point mutations in malignant melanomas (Skin-Melanoma) indicates that about two thirds are passenger point mutations. To solve this problem, we designed a rank-and-cut approach to identify the driver point mutations in each driver GE. Briefly, point mutations observed across the cohort were sorted on the basis of a number of features that distinguish known tumorigenic mutations from passenger events, and top-ranking instances –up to the number computed from the excess– were nominated as drivers (Details in Suppl. Note 1 and Extended Data Figs. 1b-e).

Second, the statistical power of the PCAWG cohort to discover driver GEs is limited; some *bona fide* cancer genes could be mutated at frequencies below the detection threshold of current driver discovery methods[16]. To overcome this problem, we first collected GEs missed by driver discovery methods in PCAWG, but detected on other cohorts of tumor whole-exomes[17,18], or validated through clinical or experimental evidence

in recent decades[2] (Fig. 1a, left panel). (We call the integrated list of discovery and prior knowledge GEs, the Compendium of Mutational Driver GEs; Suppl. Table 1.) Then, we designed a set of stringent rules to nominate driver point mutations affecting these GEs on the basis of prior knowledge about their mode of action (Suppl. Note 1, Extended Data Fig. 1b and f). Driver point mutations in those elements are assumed to be few, due to their lack of signal of positive selection in the cohort. (Of note, point mutations affecting the intronic splice sites of tumor suppressor genes in the Compendium undergo the rule-based approach.) The decision-making process integrated by the rank-and-cut and the rule-based approaches accurately distinguishes between known driver and benign point mutations (Suppl. Note 1; Extended Data Fig. 1g)

Note that we may still underestimate the total numbers of driver point mutations, because some driver GEs may remain undiscovered. The underestimation is probably more acute for non-coding GEs, as our capability to complement non-coding driver GEs with prior knowledge is minimal due to the scarcity of previous whole-genome driver identification efforts.

After applying the rank-and-cut and rule-based approaches to point mutations across all tumors in PCAWG, we obtained a patient-centric catalog composed of 5,802 driver point mutations (Fig. 1c,d; Extended Data Fig. 2a; Suppl. Table 2). Interestingly, the single most recurrent driver point mutation in the cohort affects the promoter of *TERT*, followed by coding SNVs affecting *KRAS*, *BRAF*, and *PIK3CA* (Fig. 1e). The coding sequence of *TP53* is the GE affected by the largest number of driver point mutations in the cohort (n=859), with the promoter region of *TERT* ranking fourth (Fig. 1f). Interestingly, it is the most recurrent driver GE in some cohorts, such as malignant melanomas and glioblastomas (Extended Data Fig. 2b).

The role of point mutations in tumorigenesis

We then used the catalog of driver point mutations to answer open questions about the role of point mutations in tumorigenesis. First, leveraging the unique opportunity provided by the comprehensive whole-genome view of drivers provided by PCAWG, we asked what is the relative contribution of non-coding mutations to cancer.

The 667 driver mutations identified in non-coding regions constitute 11% of the catalog (Fig. 1d and Fig. 2a; table below the barplot). While this contribution may be underestimated because our detection of driver point mutations is biased towards coding genes due to the use of prior knowledge, we obtain a similar result (10%) if the calculation is done only on discovery GEs (Suppl. Note 1; Extended Data Fig. 2c). Up to 22% of all PCAWG tumors contain at least one non-coding driver point mutation (Fig. 2a). Interestingly, more than 50% of samples of certain types of tumors, such as bladder transitional cell carcinomas (Bladder-TCC) and glioblastomas (CNS-GBM), have at least one non-coding driver point mutation. At the other extreme of the spectrum, in tumor types such as acute myeloid leukemias (Myeloid-AML) and pilocytic astrocytomas (CNS-PiloAstro), non-coding driver point mutations appear to contribute little to tumorigenesis.

Next, we asked how many driver point mutations appear in each tumor, and how this number varies with the

mutational burden. We found that 78% of PCAWG tumors have at least one driver point mutation, with the mean reaching 2.9 per tumor (Figs. 2b and 3b). Strikingly, even heavily mutated tumors, such as malignant melanomas (Skin-Melanoma) and colorectal adenocarcinomas (ColoRect-AdenoCA), carry a mean of 3.1 and 7.3 driver point mutations, respectively. Overall, if the entire range of variation of the mutational burden across the cohort is considered (i.e., between just 41 and 2.5 million point mutations per genome), the number of driver point mutations remains remarkably stable, with the interquartile range between 1 and 4 (Fig. 2b). The number of driver point mutations increases slower than the total number of point mutations in PCAWG tumors (Extended Data Fig. 3a). As a result, the proportion of driver point mutations is anticorrelated with the mutational burden of tumors (Fig. 2c). Across most cancer types, driver point mutations show a higher cancer cell fraction than passenger point mutations (Extended Data Fig. 3b). This observation is more apparent for certain GEs in the Compendium, which exhibit a significant accumulation of highly clonal point mutations, thereby indicating that they carry potential founder or potent driver mutations (Fig. 2d).

In at least 4% (22 out of 560) of the tumors where we fail to identify driver point mutations the cause may be an insufficient coverage in the sequencing (Extended Data Fig. 3c). This affects in particular some cohorts, such as prostate adenocarcinomas (Prost-AdenoCA) and stomach adenocarcinomas (Stomach-AdenoCA). The failure in the tumors of other cohorts may be caused by the incompleteness of the Compendium or the prevalence of other types of genomic alterations in their malignization.

In summary, despite huge differences in mutational burden, tumors carry a stable low number of driver point mutations –2.9 as average across PCAWG. While non-coding point mutations represent only a small fraction (3%) amongst all point mutations in GEs of the Compendium, they play a role in the malignization of 22% of the cohort, and constitute 11% of all driver point mutations in the PCAWG cohort.

Integrated panorama of whole-genome driver mutations

In addition to point mutations, tumors may be driven by SVs, such as somatic copy number alteration (SCNA) gains or losses, and balanced genomic rearrangements (GRs), such as translocations and inversions. As with point mutations, we first created a Compendium of driver SVs (Fig. 1a, right panel), and then used strict rules to identify which SCNAs (Supp. Note 2; Extended Data Fig. 4a) and GRs (Extended Data Fig. 4b, Methods) most likely act as drivers in each PCAWG tumor. The integration of the catalogs of driver point mutations and SVs across the cohort thus produced the first pan-cancer whole-genome panorama of driver events (Fig. 1a, right panel; Fig. 3a and b; Suppl. Table 3).

We asked whether the panorama gave support to the widespread notion that all tumors have a genomic origin. We found driver mutations in 90% of the tumors of the cohort, with variable fractions across tumor

types (Fig. 3b). Our failure to detect genomic drivers in some cohorts, such as kidney chromophobe adenocarcinomas (Kidney-ChRCC) and medulloblastomas (CNS-Medullo) may be attributed to our incomplete knowledge of cancer genes in these malignancies or to a prevalence of epigenomic alterations. Nevertheless, our results demonstrate, to the extent allowed by current analyses, that genomic mutations drive at least 9 out of 10 tumors.

We then turned to the longstanding question of the minimum number of driver mutations needed to turn a cell malignant. We found on average 4.6 driver mutations across tumors of the cohort, again with large differences across cancer types (table below Fig. 3b). Even tumors of intensely studied types of cancer –such as breast adenocarcinomas (Breast-AdenoCa)– for which accumulated prior knowledge probably encompasses most possible driver events, contain on average fewer than 7 driver mutations. If we assume that driver mutations accumulate over time during the development of cancer[19], the number required to initiate a tumor is probably even lower. In agreement with this idea, we found that tumors diagnosed at older ages tend to carry more driver mutations (Extended Data Fig. 5a), although this general trend masks important differences among malignancies (Extended Data Fig. 5b and c). However, the fact that tumors across all cohorts possess, on average more that one driver mutation suggests that the combination of different driver events is relevant in the path towards tumorigenesis. In this regard the panorama is a great resource to explore the combinations of driver mutations in tumors. For example, we found that BCL2 cis-activating GRs and CREBBP mutations in mature B-cell lymphomas tend to occur in the same tumors, like TP53 mutations and MYC activating mutations. On the other hand, SPOP point mutations occur in mutual exclusivity with ERG activating mutations across prostate adenocarcinomas. The same is true for mutations affecting genes of the cell cycle (CDKN2A/B and CCNE1/D1) across malignant melanomas and esophageal adenocarcinomas (Extended Data Fig. 9b).

Since there have been contradictory reports of the relative contribution of SVs and point mutations to tumorigenesis[20,21], we next explored the relative prevalence of both types of alterations in the PCAWG cohort. We discovered that both appear to play a role of similar preponderance, with a mean of 2.9 driver point mutations in 78% of samples with at least one driver point mutation, and 3.4 driver SVs in 78% of samples with at least one driver SV (Fig. 3c). However, these figures hide important differences across cancer types. For example, a higher contribution to tumorigenesis is made by SVs in breast adenocarcinomas (Breast-AdenoCa; mean of 6.1 SVs vs. 2.2 point mutations) and ovary adenocarcinomas (Ovary-AdenoCA; mean of 5.3 SVs vs. 1.9 point mutations, respectively). On the other hand point mutations make a larger contribution in colorectal adenocarcinomas (ColoRect-AdenoCA; mean of 2.4 SVs vs. 7.3 point mutations, respectively) and mature B-cell lymphomas (Lymph-BNHL; mean of 2 SVs vs. 6 point mutations, respectively). A major proportion of the samples with driver mutations (1,725/2,577) harbor both types of drivers, whereas a small (and comparable) proportion of samples harbor either driver point mutations

(n=292) or driver SVs (n=284) (Extended Data Fig. 4c). Furthermore, as with point mutations, the number of driver SCNAs in tumors is stable, irrespective of the fraction of the genome involved in focal amplifications or deletions (Extended Data Fig. 6).

We also examined the contribution of each type of driver mutation to the landscape of therapeutic actionability of the cohort (Fig. 3b, bottom panel). We matched drugs employed in the clinic –both a-approved or under clinical trials– to the panorama of driver mutations in PCAWG, using the Cancer Biomarkers Database[22]. We considered only mutations that constitute biomarkers of response to an anti-cancer drug in tumors bearing no concomitant biomarker of resistance to the same drug (Methods). Overall, we found that driver point mutations contribute the largest number of tumors with actionable targets (785), followed by driver SCNAs (780), whereas only 90 tumors have actionable driver GRs. Overall, 1,242 tumors in the cohort bear mutations that constitute biomarkers of response to drugs that are either FDA-approved or in clinical trials. If opportunities for repurposing these drugs –to different mutations affecting the same driver gene or the same mutation in a different tumor type– and mutations that could be targeted by molecules in pre-clinical stages are also considered, this number increases to 1,943 tumors (Extended Data Fig. 7). Interestingly, several elements with non-coding driver mutations contribute repurposing opportunities to this actionability landscape. These include mutations causing over-expression of *TERT*, which are susceptible to targeting via Eribulin, currently in pre-clinical trials for gliomas[23]. Moreover, considerable fractions of the tumors of some cohorts (181 over all PCAWG) could potentially benefit from immune checkpoint inhibitors[24].

In summary, our results contribute to answering three longstanding questions in cancer genomics. First, to the extent allowed by current analyses and bioinformatics methods, we showed that essentially all tumors contain one or more driver genomic mutations. Second, we demonstrated that both point mutations and SVs contribute to tumorigenesis in similar proportions, albeit with differences across cancer types. Finally, we found that the number of driver mutations in a tumor is small, and virtually independent of the tumor's burden of alterations.

Contribution of other genomic events

To complement the panorama of driver mutations, we identified germline SNVs that may contribute to tumorigenesis. Specifically, we searched for i) truncating (stop gain, frameshift, splice site) mutations and rare SV deletions affecting loss-of-function driver genes and cancer predisposition genes; and ii) loss-of-function missense SNVs of *TP53*. We found that 15% of the tumors contain one or more germline variants potentially contributing to tumorigenesis (Fig. 3d).

Then, motivated by the idea that tumor suppressor genes in principle require two loss-of-function mutations

to inactivate both alleles[25], we asked how frequent these biallelic hits are across tumors. We considered both germline and somatic mutations affecting both alleles of loss-of-function driver genes across tumors. We found instances of biallelic inactivation of tumor suppressor genes in more than half (60%) of PCAWG tumors (Fig. 3d). In some cohorts, such as bladder transitional cell carcinomas (Bladder-TCC) and lung squamous cell adenocarcinomas (Lung-SCC), more than 90% of tumors carry at least one biallelic hit, comprising either somatic/somatic or somatic/germline events.

We also found that chromosomal arm-level gains and losses and whole genome duplications appear to be a pervasive type of mutation in the cohort (table below Fig. 3b; Extended Data Fig. 8). Almost one third (28%) of the tumors carry whole genome duplications, ranging between 0% of some cohorts (thyroid adenocarcinomas (Thy-AdenoCA), acute myeloid leukemia (Myeloid-AML), and chronic lymphocytic leukemia (Lymph-CLL)) and 67% of bone osteosarcoma (Bone-Osteosarc). Although it is difficult to assess the role of chromosomal-level and whole-genome duplication events in tumorigenesis, their pervasiveness suggests that they contribute to this process.

GEs affected by different types of driver mutations

Finally, we used the whole-genome panorama of cancer driver mutations to explore different ways of alterations of cancer genes that result in tumorigenesis. Overall, 53% (193/367) of genes bearing driver point mutations in the coding sequence are susceptible to other types of driver mutations (Fig. 4a and b). *TP53*, for instance, carries the highest number of driver point mutations in the cohort (941 samples), with more than 50% of the samples affected in certain tumor types, such as pancreatic (Panc-AdenoCA), breast (Breast-AdenoCa), and ovarian adenocarcinomas (Ovary-AdenoCA). It is also affected by SCNA losses, truncations due to GRs, and intronic splicing point mutations. The expression of *TP53* is significantly decreased (Mann-Whitney $P<0.0001$) in tumors with intronic splice-site point mutations, SCNA loss, truncation or coding truncating point mutations (via nonsense-mediated decay[26]), compared to those with the wild-type form (Fig. 4c). On the other hand, *CDKN2B*, *PIK3CA*, *CTNNB1*, *ERG*, *MCL1* and *CCND1* are mostly affected by only one type of mutation across tumor types (Fig. 4b).

TERT is the gene most frequently affected by non-coding driver point mutations (Fig. 4b). It appears mutated in 181 samples through multiple routes, such as promoter SNVs and MNVs, balanced genomic rearrangements affecting cis-regulatory elements, and SCNA gains. *TERT* promoter point mutations are particularly frequent in glioblastomas (26/39), oligodendrogliomas (11/18), bladder transitional cell carcinomas (15/23), thyroid adenocarcinomas (11/48), and malignant melanomas (53/107). *TERT* SCNA gains are observed across malignant melanomas (n=12), hepatocellular carcinomas (n=5), lung squamous (n=3) and adenocarcinomas (n=5), whereas enhancer hijacking events affecting TERT are observed in

malignant melanomas (n=4), leiomyosarcomas (n=3), kidney renal clear cell (n=1), kidney chromophobe renal clear cell (n=1) and hepatocellular carcinomas (n=1). Notably, each of these routes of driver mutation leads to the over-expression of *TERT* (Fig. 4c).

Finally, we studied the biallelic inactivation of tumor suppressor genes, that comprise combinations of germline and/or somatic driver events. Genes recurrently targeted by biallelic inactivation events (Fig. 4d) are dominated by somatic/somatic events (96%) over germline/somatic events (18%). The most frequent biallelically mutated gene, *TP53,* is inactivated across 733 tumors of 27 cancer types, most of which (705) harbor a somatic driver point mutation affecting one allele and the somatic loss of the other allele. Overall, we were able to detect the biallelic inactivation of *TP53* in 78% of the tumors with *TP53* driver mutations; in the remaining fraction, other mechanisms such as promoter hypermethylation may play a role. Well-known tumor suppressor genes, such as *CDKN2A* and *CDKN2B*, are rarely inactivated by point mutations, but rather appear homozygously lost (281 and 257 tumors, respectively). On the other hand, biallelic inactivation of *BRCA1* and *BRCA2* appears primarily due to one germline SNV, followed by the somatic loss of the second allele in 63% (22/35 samples) and 44% (24/54 samples) of tumors with *BRCA1* and *BRCA2* driver mutations, respectively. The number of genes with biallelic inactivation in individual tumors differ within and across cancer types (Extended Data Fig. 9a). For example, in both bladder transitional cell carcinomas (18/23) and pancreatic adenocarcinomas (178/232), around 77% of samples carry more than one TSG with biallelic inactivation. Known instances of co-occurring biallelic inactivations were also observed, such as that of *MEN1* and *ATRX/DAXX*[27] in 16% (13/81) of pancreatic endocrine, and that of *VHL* and *SETD2/PBRM1*[28] in 30% (43/143) of kidney renal clear cell carcinomas.

Discussion

In the present study, we implemented a systematic approach to comprehensively identify driver mutations in each tumor of the PCAWG cohort, with the final aim of exploring the genomic roots of tumorigenesis. Although thorough, this panorama of driver mutations is not complete because the approaches used to identify driver events possess limitations. Technical problems of the sequencing and the calling of variants may result in missing driver point mutations. The discovery of driver GEs is limited to regions of the genome with known functional elements, basically coding and non-coding genes and their cis-regulatory elements. Mutations affecting other regions of the genome and perturbing, for example, the higher-order structure of the chromatin may also contribute to tumorigenesis but will be missed by the methods designed to detect signals of positive selection in well-defined GEs. Also, some mutations affecting driver GEs which may still contribute to tumorigenesis at low frequency, may be overlooked by the drivers discovery until larger cohorts of whole-genomes become available[29]. The catalog of driver point mutations is incomplete with respect to non-coding driver mutations, due to the scarcity of prior knowledge of non-coding driver GEs.

This probably also affects the identification of driver SVs in some cohorts of yet unexplored cancer types, since only few events –due mostly to the lack of statistical power– were discovered in the PCAWG cohort.

The panorama of driver mutations of PCAWG tumors reveals that almost all tumors contain genomic driver alterations (90% had at least one type of driver mutation), implying that the presence of somatic mutations is a condition necessary –although perhaps not sufficient[30]– for virtually all tumors. One fifth of the PCAWG tumors had at least one driver-balanced rearrangement, 72% at least one driver SCNA, 78% at least one driver point mutation, and 15% carry at least one germline susceptibility variant (Fig. 5). A prototypical glioblastoma, for example, bears a SNV activating the *TERT* promoter, the biallelic loss of *CDKN2A* and *CDKN2B*, the hyperactivation of EGFR via either a CNA gain or a coding point mutation, and/or the double somatic biallelic inactivation of *TP53*. The panorama also showed that the relative contribution of non-coding mutations is about 10%; this contribution surpasses 50% of the samples in some types of cancer. We also discovered that while overall point mutations and SVs represent similar proportions of all driver events, this observation conceals large differences across cancer types, with some preponderantly harboring driver point mutations and others driven primarily by SVs. Another striking finding stemming from the panorama is that the variation in the number of driver mutations required to initiate a tumor is not related to the total burden of mutations of the tumor. Instead, it is fairly constant and tends to be fewer than a dozen events in virtually all tumors.

The recent advances in cancer genomics described in the introduction have reliably identified genes driving tumorigenesis across cohorts of patients of different cancer types. However, the large-term objective of selecting the most effective treatments to target tumors –the personalized cancer medicine paradigm– requires the identification of the specific mutations driving the tumorigenesis in each patient, which remains a largely unsolved problem. The approach that we have described and implemented here to solve this challenge using the PCAWG cohort constitutes a proof-of-principle that bridges the gap between cancer genomics and personalized cancer medicine. We envision that personalized medicine initiatives, via systematic sequencing the tumor of cancer patients will make use of this type of approaches to fine-tune clinical decision-making.

FIGURE LEGENDS

Figure 1. Approach to building a panorama of driver mutations across PCAWG tumors
(a) A schematic representation of the comprehensive incremental approach applied to identify all driver mutations in PCAWG tumors. GEs of different types with signals of positive selection in the cohort, and previously known driver GEs and point mutations obtained from other sources (left panel) integrated the Compendium of Mutational Driver GEs. (The discovery GEs comprise both novel discoveries and rediscoveries of drivers across the cohorts in PCAWG, color coded in the panel.) Previously known and novel driver SVs were collected to build a Compendium of Structural Variant Drivers, and potentially tumorigenic germline variants affecting cancer susceptibility genes or tumor suppressor genes within the Compendium were also identified in the cohort (right panel) using stringent rules. (The list of tumor suppressor genes included in the search for SV drivers and susceptibility germline variants was obtained from the Compendium.) The catalogs of driver point mutations and driver SVs identified across tumors integrate the whole-genome panorama of driver mutations of PCAWG.
(b) Point mutations in discovery GEs may be either drivers or passengers; the fraction of the former is variable across GEs. It is computed as the excess of mutations above the background rate of each GE. The figure shows the distribution of the excess of mutations computed across GEs in the pan-cancer, breast adenocarcinoma and malignant melanoma cohorts. (Colors of GEs as in left panel a).
(c) The catalog comprises 5,802 driver point mutations, including Known Drivers, driver mutations identified via the rank-and-cut approach (Driver by Rank) and mutations in other driver GEs identified via rules (Driver by Rule). The number of passenger mutations (in gray) in driver GEs is also shown,
(d) The number and fraction of drivers among all point mutations identified varied across GEs of different nature. In the case of intronic point mutations affecting splice sites, only those affecting the canonical donor and acceptor sites and predicted to be loss-of-function in the tumor suppressor genes by the LOFTEE[31] are taken into account to compute the fraction of passengers.
(e) Driver point mutations in the cohort include both coding and non-coding events. The heatmap represents the recurrence of driver point mutations across cohorts of GEs with at least 10 driver point mutations in PCAWG, with the color in each cell showing the proportion, and the number, the absolute count of mutated samples. The nature of the GEs affected by them is denoted by circles following the same color code as in Figure 1a. Note that each point mutation is defined by its position and nucleotide change; KRAS hotspot mutations caused by two different nucleotide changes appear at the second and third row. The columns of the heatmap are sorted according to the size of the cohort (high to low).
(f) Identified driver point mutations in GEs with at least 40 across the cohort.

Figure 2. Catalog of driver point mutations
(a) Number of tumors in each cohort and across all PCAWG bearing either coding (green), non-coding (blue), or both (purple) types of driver point mutations. Below the graph, we present the number of samples, and the number of coding and non-coding driver point mutations in each cohort.
(b) Tumors in the cohort exhibit mutational burden with a wide range of variation (gray dots and boxplot), whereas the number of driver point mutations remains remarkably stable (red dots and boxplot). (Only tumors with at least one driver point mutation are included in the plot.)
(c) The fraction of driver point mutations in tumors and the total number of point mutations are negatively correlated across the PCAWG cohort.
(d) Normalized (Zscore) distribution of the cancer cell fraction of GEs across tumor types. Filled circle represent GEs with point mutations cancer cell fraction deviating significantly from expectations (two-tail p-value < 0.05). The names of significant coding and non-coding GEs are colored following the convention in panel (a).

Figure 3. Whole-genome panorama of driver mutations
(a) The whole-genome panorama of driver mutations in the PCAWG cohort, represented as a circos plot. Each arc in the circos plot represents a tumor in the cohort. From the periphery to the center of the plot the layers represent: i) the total number of driver alterations in the tumor; ii) whether the sample bears a whole genome duplication; iii) the tumor type, with color codes specified in panel b; iv) the number of SCNAs; v) the number of GRs; vi) coding point mutations; vii)

non-coding point mutations; and viii) germline SNVs and indels.
(b) Row-stacked bar plot of the total number of tumors in each cohort with different counts of driver mutations. The table below the bar plot offers a complementary description of the panorama. The heatmap at the bottom represents the fraction (color) and number of tumors of each cancer type with driver point mutations, SCNAs, and balanced genomic rearrangements that provide biomarkers of response to anti-cancer therapies either approved or under clinical trials.
(c) Contribution of point mutations and SVs to tumorigenesis across cancer types. (Split violin plots represent the smoothed probability distribution of the numbers of each type of mutations in each cohort, which expand beyond the boundaries of the actual distributions.) The halves of each violin plot represent the distribution of the total number of driver point mutations (bottom, orange) and driver SVs (top, green) in the tumors of each cohort.
(d) Fraction of tumors of each cohort affected by germline susceptibility variants (top), and biallelic inactivation events (bottom).

Figure 4. Multiple types of mutations affecting driver GEs
(a) Circos plot representation of the number of GEs in the Compendium affected by different types of driver mutations. Each arc in the circos represents GEs of a different nature; links between them depict GEs susceptible to two types of alterations, with their thickness proportional to the number of elements sharing both types of alterations. The numbers at the border of the plot represent unique counts of GRs suffering different types of mutations.
(b) GEs targeted by different types of mutations in the cohort in more than 65 samples. The heatmap shows the recurrence of alterations experienced by each of them across cancer types (with the color indicating the proportion, and the number the absolute count of mutated samples); the barplot at the right reflects the proportion of each type of alteration affecting each GE.
(c) Different types of driver mutations affecting a GE that result in the same type of change to their mRNA level.
(d) Tumor suppressor genes showing biallelic inactivation in 10 or more samples.

Figure 5. Snapshots of the panorama
Each row summarizes either the pan-cancer cohort or the cohort of a particular tumor type. The horizontal bar plot (at the left of each row) represents the proportion of samples with different types of mutations. The dot plot with whiskers at the right represents the mean number of each type of driver mutations registered across tumors with at least one event (the square dot), and its standard deviation (gray whiskers). For each tumor type, a representative set of genes which receive frequent driver mutations is presented at the right side of each row. For the full list of genes with driver mutations see Supplementary Table 3.

EXTENDED DATA FIGURE LEGENDS

Extended Data Figure 1. The onCohortDrive method
(a) Distribution of the expected proportion of drivers in each type of GE, computed using NBR (Supplementary Note 1) as the excess of mutations of each element above the background rate.
(b) Point mutations in Genomic elements (GEs) follow different processes (either rank-and-cut approach or rule-based) to be nominated as drivers, depending on whether we are able to compute the excess of somatic mutations above the background rate in these GEs. In the rank-based process (right panel), mutations affecting each GE are ranked according to several features, and the rank is then cut at the position equal to the estimated excess of mutations in the element above the background rate. Mutations above the cut position are considered drivers, and those below, passengers. In the rule-based process (left panel), mutations in each tumor are nominated as drivers or passengers using the same features employed in the rank-based approach. (Details in Supplementary Note 1.)
(c) Distribution of CADD scores of groups of variants in cancer genes. Sets of mutations, from left to right: i) known driver point mutations in cancer genes observed in PCAWG samples; ii) all other mutations in cancer genes in PCAWG; iii) all possible variants in cancer genes; iv) all polymorphisms observed in ExAC in cancer genes. The distributions of CADD scores of different sets of mutations were compared using the Kolmogorov-Smirnoff test. (Details in Supplementary Note 1.)
(d) Relative rank of known driver mutations within all mutations observed in the PCAWG cohort. The violinplots represent the distribution, from right to left, of random relative ranking of known driver mutations, functional impact based ranking, and functional impact and clustering based ranking. (Details in Supplementary Note 1.)
(e) Distribution of mutational probability (computed using ncdDetect) of known driver mutations relative to other mutations. Known driver mutations overall have significantly lower probability to occur than other mutations observed in tumors. In the left panel each dot is a known driver mutation in a sample and its probability of occurrence, relative to the probability of all other mutations in the sample is shown. The bars in the right panel show the comparison of the average probability of known driver mutations to the average probability of other sampled mutations (size equal to the number of known driver mutations) in each cohort. The cohorts with significantly ($P<0.01$) lower probability of occurrence of known mutations are highlighted (*). (Details in Supplementary Note 1.)
(f) Distribution of the probability of occurrence of point mutations (ncdDetect) in coding genes (top, green) and promoters (bottom, yellow). The vertical broken lines mark the cutoffs of probability used to make the groups of point mutations considered separately in the rule-based approach.
(g) Results of the benchmark of onCohortDrive on groups of known driver and benign mutations. The three main rows in the table describe the three datasets used in the benchmark (described in Supplementary Note 1), and the columns represent (from left to right) the number of variants in each dataset classified by onCohortDrive as drivers or passengers, and the total number of variants included in each group.

Extended Data Figure 2. Building the catalog of driver point mutations
(a) Fraction of point mutations affecting GEs in different cancer types and in the pan-cancer cohort that are identified as known-driver, driver-by-rank, driver-by-rule in the catalog of driver point mutations. (Cohorts are sorted by the proportion of driver point mutations identified across each of them, and the classes of driver point mutations are color-coded as in Fig. 1c.)
(b) Fraction of known-driver, driver-by-rank, driver-by-rule and passenger mutations in selected GEs across four cohorts in PCAWG.
(c) Unbiased contribution of non-coding point mutations to tumorigenesis. Similar to main Figure 2a, but computed only from the discovery GEs (Supplementary Note 1)

Extended Data Figure 3. Features of the catalog of driver point mutations
(a) Correlation between the number of driver point mutations in tumors and their mutational burden (in logarithmic scale). There is a significant positive correlation between these two quantities, showing that tumors with more mutation burden carry more driver point mutations. Nevertheless, the number of driver point mutations increases slower that the total number of point mutations in PCAWG tumors: the slope of the regression between the number of driver point mutations and the logarithm of the total number of point mutations is approximately 2.
(b) Driver point mutations overall have significantly higher Clonal Cell Fraction (CCF) than other mutations observed in tumors. The left panel shows the comparison of the CCF of each driver mutation to the average CCF of other mutations in the same tumor, and the right panel shows the comparison of the average CCF of driver mutations to the average CCF of driver mutations in each cohort. The cohorts with significantly ($P<0.01$) higher or lower CCF of driver mutations than randomly sampled mutations are highlighted (*).
(c) In some cohorts, samples with no identified driver point mutation are significantly enriched (Fisher's p-value<0.05) for tumors with insufficient coverage (estimated via the number of reads per clonal copy, or NRPCC) to call point mutations reliably (NRPCC<5). In the graph, each cohort is represented as a circle, located according to the odds-ratio

and the p-value of the Fisher's test. The size of eah circle is proportional to the fraction of samples with no driver point mutations in the cohort. Cohorts with no drivers in samples with sub-threshold NRPCC (infinite odds-ratio) appear at the right margin of the graph; cohorts with no sub-threshold NRPCC samples (undetermined odds-ratio) appear at the left margin of the graph.

Extended Data Figure 4. Driver SVs in the panorama
(a) Most recurrent driver CNA gains and losses in the PCAWG cohort. The frequency of each event across cohorts is represented as a row-stacked barplot using the same color legend for tumor types as in previous figures. The names of the events correspond to the driver gene within the event or to the chromosomal band the event overlaps with if it contains no known driver gene. The circles below the bar plot denote the nature of each driver as CNA gain or loss.
(b) Coding genes involved in the most recurrent driver genomic rearrangements in the PCAWG cohort. Note that events involving several partners of translocation may have been added to produce the total count of events for each gene. The frequency of the events involving each gene across the cohort of each cancer type is represented as a row-stacked bar plot using the same color legend for tumor types as in previous figures. The circles below the bar plot characterize each event as truncated tumor suppressor gene, fusion oncogene or cis-activating rearrangement.
(c) Distribution of samples with driver mutations, driver SVs or both types of driver mutations across cohorts and across the whole PCAWG. This graph complements the view in Fig. 3c of the paper on the relative contribution of point mutations and SVs to tumorigenesis. Cohorts are sorted following the same order as the violinplots in Fig. 3c.

Extended Data Figure 5. Correlation between the number of driver mutations in tumors and the patients' age
(a) Correlation between the number of driver point mutations in tumors and patients' age across the PCAWG cohort. Tumors of different cancer types are represented following the same color convention as throughout the manuscript. From left to right the panels represent the correlation between i) the number of driver point mutations; ii) the number of driver SVs; iii) the number of biallelic events; and iv) the total number of driver events and the age of diagnosis of the tumor. The number of all types of mutations correlates with the age of patients at diagnosis, suggesting that driver mutations accumulate with the progression of tumors, and therefore supporting the idea that our calculation of mean number of driver mutations across the cohort may be higher than the actual minimum number of driver events required to turn a cell malignant.
(b) Nevertheless, the correlation between the number of different types of mutations in tumors and patients' age across selected cancer cohorts (those with at least 100 samples) show striking differences between malignancies. (Odd-number rows present graphs with the correlation between the number of driver point mutations in tumors and the age of diagnosis across three cancer types, and pair-number rows present equivalent graphs with the correlation between the total number of point mutations (in log scale) in tumors and the age of diagnosis.) For certain tumor types (such as prostate adenocarcinomas and medulloblastomas) the mutation burden grows exponentially and the number of drivers grows linearly with the age of diagnosis of the tumor. For some malignancies in this group, such as medulloblastomas, the differences may be related with different courses of the disease in its pediatric and adult presentations. On the other hand, for some cancer types there is no discernible correlation between the total number of point mutations or the number of drivers and the patient's age (see, for example melanomas and pancreatic adenocarcinomas). This could mean that these tumors tend to be diagnosed mostly in later stages independently of patients' age. While these differences may reflect actual dissimilarities in the course of these diseases, these results must be taken with caution due to the small sample sizes of these cohorts.
(c) Correlations between the total number of driver mutations and the age of diagnosis of the tumor across the same cohorts represented in panel (b).

Extended Data Figure 6. Driver SCNAs across tumors with varying focal SCNAs burden
The graph represents the fraction of the genome of tumors with different number of driver SCNA events that is involved in different types of focal SCNA events. Each dot represents a tumor, colored following the same legend to represent tumor types as in previous figures. The three panels represent the relationship between the number of driver high-level SCNA gains (left), driver high-level SCNA losses (center), and driver low-level SCNA losses (right), with the fraction of the genome involved in focal events of the same nature in each case. The figure demonstrates the lack of a trend towards an increase of the number of driver events in tumors with higher fractions of the genome involved in focal events.

Extended Data Figure 7. Landscape of anti-cancer actionability of the PCAWG cohort
(a) The landscape of all anti-cancer therapeutic opportunities uncovered by the panorama of driver mutations is represented as a row-stacked bar plot containing the fraction of tumors in each cancer type –and in the pan-cancer cohort– with therapeutic opportunities of different type (color-coded). These include opportunities arising from alterations that are known biomarkers of the response of a tumor to a drug (strong colors) and opportunities of

repurposing of these drugs to different tumor types or different types of alterations (faded colors). In this bar plot, each tumor is only counted once with the targeting opportunity that is closest to the clinical use of the drug.
(b) The heatmap represents the fraction of samples with therapeutic opportunities of each category (rows) in each tumor type. Here, all opportunities to target each tumor are accounted for.
(c) Furthermore, we identified the tumors with a mutation burden above a threshold reported to accurately predict the effectiveness of immune checkpoint therapies to treat a variety of tumors. We found that 7% of tumors in the cohort including 65% of melanomas and 57% of lung squamous cell carcinomas tumors might in principle benefit from these therapies. The fraction of tumors of each cohort susceptible of responding to either immune checkpoint inhibitors or a combination of these drugs with classic anti-cancer drugs considered in the landscape described above is represented as a rowstacked bar plot.

Extended Data Figure 8. The landscape of arm-level SCNAs across tumors
The graphs show the distribution of the number of arm-level SCNA events observed in each cohort.
(a) Distribution of the number of arm high-level SCNA loss events in each cohort.
(b) Distribution of the number of arm high-level SCNA gain events in each cohort.
(c) Distribution of the fraction of the number of arm low-level SCNA loss events in each cohort.
(d) Distribution of the number of arm low-level CNA gain events in each cohort.
The boxplots representing the distribution of events in tumors of each cohort follow the same color legend as previous figures.

Extended Data Figure 9. Distribution of biallelic inactivation of tumor suppressor genes and co-occurrence of driver mutations
(a) Number of genes with biallelic inactivation per samples in each cohort. Only cohorts with at least 10 samples affected by one or more biallelic inactivation are included.
(b) Heatmaps representing the co-occurrence and mutual exclusivity of selected GEs across tumors of four cohorts. The colors of each type of driver mutation follow the code used in previous figures, with the addition of biallelic inactivation events, which are represented in gray.

Supplementary Table 1. The Compendium of Mutational Driver Elements

Supplementary Table 2. Annotated driver point mutations

Supplementary Table 3. The panorama of driver mutations in PCAWG

METHODS

Genomics data

From several PCAWG working groups we obtained the somatic point mutations (single and multiple nucleotide variants and short indels), copy number alterations, structural variants, cancer cell fraction, and expression data across 2,583 tumors in the cohort. (Eight samples incorrectly classified as malignant were discarded from downstream analysis.) Genomic elements that potentially harbor driver point mutations were defined as agreed within the Consortium, using the coordinates of Gencode v19. The list of these genomic elements included the coding DNA sequence (CDS) of protein-coding genes, their 5'UTR and 3'UTR, the coordinates of their intronic acceptor and donor splice sites, their promoter regions, known enhancers, and the exonic regions of non-coding RNAs, and the promoter regions of the latter.

Compendium of Mutational Driver GEs

We retrieved the list of mutational discovery genomic elements identified as drivers of PCAWG tumors from the PCAWG Drivers and Functional Interpretation Group. This list of genomic elements covers coding genes, promoters, enhancers, 3' and 5' UTRs, and ncRNAs. Each element in this list was detected because its mutational patterns exhibited signals of positive selection across one of several cohorts within PCAWG, or across the entire cohort (pan-cancer analysis). They were therefore annotated as mutational drivers in each of these cohorts, and we retrieved the excess of somatic mutations for each of them in the respective cohorts, estimated by the NBR algorithm[15]. These elements constitute the discovery compendium. Five coding genes and one promoter region of a coding gene identified in the lymphoma cohorts with a high proportion of mutations introduced by AID (>40%) were removed from the Compendium.

In an effort to complement the list of discovery drivers and to add prior knowledge about the drivers of tumorigenesis to the compendium, we employed a manually collated and curated list of previously known cancer elements –driving tumorigenesis through mutations, SCNAs and/or balanced genomic rearrangements– across tumor types. This list of drivers comprises: a) elements with validated tumorigenic effect, obtained from the Cancer Gene Census (CGC)[2] and literature reports (with special attention paid to recently reported instances of driver non-coding genomic elements), and b) genes whose mutational patterns show signals of positive selection across previously analyzed cohorts of tumor exomes or genomes[17,18,32]. To add CNA driver elements to the Discovery Compendium, we relied on the identification of amplified or deleted chromosomic regions (peaks) with significant recurrence across TCGA cohorts (see below). To add genomic rearrangement driver events to the Compendium, we manually gathered known events from the literature and searched for their occurrence in the PCAWG cohort. Only rearrangements that appear in tumors of the cohort were included in the panorama (see below).

Functional annotation of point mutations

The Variant Effect Predictor[31] v86 was used with GRCh37 coordinates to determine the genes affected by point mutations and their effect. The Combined Annotation Dependent Depletion (CADD) v. 1.0 score of all single and multiple nucleotide variants –as well as the maximum possible CADD score of short indels– was obtained at cadd.gs.washington.edu[33].

Identification of driver point mutations

We developed a method, onCohortDrive to identify driver mutations affecting driver GEs in individual tumors. Supplementary Note 1 contains the full description of the method and its benchmark. The algorithm comprises a rank-based approach and a rule-based approach. The rank-based approach is applied to point mutations in GEs with an observed excess of point mutations above the number expected across the cohort. Briefly, consider that N is the total number of point mutations affecting a GE, and *n* is the point mutations in excess above the background rate across the tumors in a cohort, with *k* known driver point mutations. The *k* known driver point mutations are first nominated as drivers and withdrawn from the rank-based process. Then, the number of point mutations available to the rank-based approach is N-k, and the excess after identifying known tumorigenic point mutations is n-k. These N-k point mutations are then ranked, based on a combination of i) functional impact scores, ii) whether they belong to mutational clusters, iii) their unlikeliness and iv) element specific features (e.g. the impact on TFBS for promoter and enhancers, the impact on secondary RNA structure for microRNAs, etc). OnCohortDrive then nominates as driver-by-rank the n-k top ranking mutations. Finally, the number of mutations in excess in the pan-cancer cohort after the excess in all cohorts has been processed (named residual excess, or *r*) are processed as explained above across all tumors in the pan-cancer cohort, excluding those in cohorts already processed.

Point mutations observed in prior knowledge GEs or in GEs with no mutational excess are not eligible for the rank-based approach and subsequently undergo a rule-based approach. The rules used to nominate driver-by-rule point mutations are based on the same features used for ranking. Point mutations that are not known drivers, drivers-by-rank, or drivers-by-rule are deemed passengers.

To make the unbiased assessment of the contribution of non-coding driver point mutations to tumorigenesis, we applied onCohortDrive solely to point mutations observed in discovery GEs across the PCAWG cohort, skipping the identification of known point mutations in GEs.

To evaluate the performance of onCohortDrive we carried out three benchmarks (described in full in Supplementary Note 1; Extended Data Fig. 1g). First, we assessed the accuracy of the approach (rank-and-cut and rule-based) to correctly classified as drivers and passengers 1,032 known cancer mutations and 12 known benign mutations in tumors in the PCAWG cohort from a total of 22,854 SNVs in coding genes of

the Compendium. We found that while only 16% of the 22,854 SNVs are classified as drivers, 95% of known drivers (982 out of 1,032) are correctly classified as drivers. All 12 benign mutations are correctly classified as passengers. Second, to verify the stringency of the rules designed within onCohortDrive, we benchmarked the capacity of the rule-based approach to correctly classify the same set of driver and passenger point mutations. As expected, a lower proportion of SNVs are classified as drivers by rules, 2,412 (11%), still maintaining a high proportion of the known driver mutations correctly classified, 805 (78%), and all 12 benign variants classified correctly as passengers. Third, we evaluated the performance of the approach to correctly label as passengers all benign SNVs (ClinVar[34]; n=314) and a list of common coding polymorphisms obtained from (ExAC[35]; n=1,413) in genes of the Compendium. OncohortDrive correctly classifies 97% and 99% of those SNVs, respectively as passengers.

Bias towards high CCF of point mutations in driver GEs

We computed the mean cancer cell fraction (CCF) of all point mutations affecting GEs identified across cancer types in PCAWG. To determine whether the mean CCF of a GE exhibited a significant bias towards high or low clonality, we randomly sampled the same number of mutations observed in each element from all the mutations in the same cancer type. By iterating this process 10,000 times, we computed an expected distribution of mean CCFs of each element and assessed whether a significant deviation from the observed mean existed.

Identification of driver SCNAs

The Compendium of SCNA GEs was populated from the analysis of chromosomal regions that exhibit signals of positive selection in their patterns of focal amplifications or deletions across cohorts of tumors analyzed by TCGA. Significant SCNA peaks were detected by applying the GISTIC algorithm[32] to both individual cohorts and to the TCGA pan-cancer cohort. Peaks detected in this manner were then merged into metapeaks and known driver genes within each metapeak were identified, when known. The list of metapeaks was further filtered by checking whether the expression of the driver contained within the metapeak exhibited a significant shift in coherence with the type of peak (amplification or deletion) in the same cohort in which it had been annotated as a driver. The metapeaks passing this filter integrated the Compendium of driver SCNAs. Finally, the copy number of the genomic segments overlapping with the loci of the metapeaks in the Compendium was assessed in the PCAWG tumors. (See Supplementary Note 2.)

Identification of driver genomic rearrangements events

The driver genomic rearrangements we considered included genic fusions involving an oncogene, truncation of tumor suppressors, and cis-activating rearrangements (e.g., promoter-rearrangement and enhancer hijacking). These events were obtained from literature reports, curated databases (Cancer Gene Census)[2], and a set of high-confidence novel genomic rearrangements that were identified in the PCAWG cohorts

(provided by PCAWG Structural Variants analysis working group). Using this information, each tumor within the PCAWG cohorts was assessed for the presence of driver rearrangements. In the case of gene fusions, the gene coordinates plus a 50-kb flanking window on either side of each member of the pair of fused genes were scanned for the presence of rearrangements. Furthermore, the rearrangements were annotated when they produced a sense in-frame fusion.

This resulted in 331 fusion events in 319 samples. For 214 of these events (in 204 samples) we could not find the expression evidence (i.e., expression of fusion transcripts) due to the lack of expression data (RNA-seq) available for those samples. However, for the 117 events (in 115 samples) with expression data available, we identified 41 events (in 40 samples) that have fusion transcript match, based on the results provided by PCAWG3 (syn10003873). For the remaining 76 fusion events (in 76 samples), the lack of fusion transcript match can be explained by promoter/enhancer hijacking events that resulted in an over-expression of the target oncogene. The majority of the fusion events that fall under this category are related to the fusion with IGH/IGL locus. On the other hand, we have included fusion events in seven samples based on the evidence from fusion transcripts (PCAWG3 results), but were not detected based on the aforementioned SV analysis.

In the case of tumor suppressor genes, the breakpoints affecting the exons were considered as drivers. In addition, we analyzed rearrangements affecting the cis-regulatory elements of the coding genes in the compendium. This included rearrangements in the promoter regions (promoter-SV) and those causing a juxtaposition of enhancers close to a gene (enhancer-hijacking). In the latter case, we focused on genes that CESAM analysis[11] has shown to become over-expressed through the enhancer-hijacking process, which takes into account the breakpoints, CNAs, target gene (mRNA) expression, and chromatin interaction data from topologically associating domains (TADs). For those genes, we performed CESAM analysis to identify PCAWG samples with genes that showed over-expression. Furthermore, we annotated the presence of any tissue-specific enhancer or super enhancer close to the breakpoints.

Identification of biallelic driver events

To identify tumor suppressor two-hit events[25], we defined biallelic inactivation as a gene locus $G^{A/B}$, where alleles A and B are genetically altered, leading to a genetic $G^{mut/mut}$ state.

The biallelic inactivation assessment includes three genetic inactivation event types consisting of somatic or germline deletions ("Loss"), somatic or germline SVs ("Break") and somatic or germline SNVs ("Mutation"). Given a heterozygous $G^{A/B}$ locus, we required a loss of the A allele of the gene, leading to a hemizygous $G^{-/B}$ state, and genetic inactivation of the remaining B allele, specifically requiring the second event to overlap the loss on the A allele, leading to biallelic inactivation.

We considered four classes of biallelic inactivations: i) Loss/Mutation, nonsynonymous driver mutations of the B allele; ii) Loss/Loss, two deletion events that overlap an exon and the copy-number derived allele count is 0 both for A and B allele; iii) Loss/Break, SVs where one or both breakpoints are situated in an exon of the B allele; and iv) Mutation/Mutation, a nonsynonymous germline SNV and a nonsynonymous driver somatic SNV of the same gene. We infer the germline mutation to occur on the A allele and the somatic mutation on the B allele, with the assumption that two independent driver mutation events are highly unlikely to occur on the same allele.

All biallelic inactivation events involving at least one Loss event which had not been detected in the process of identification of driver SCNAs –either because the SCNA GE was under the statistical power of detection in the corresponding cohort, failed the expression filter, or because the Loss event involved an arm-level deletion (see Supplementary Note 2)– were included in the panorama as driver SCNA loss events.

The panorama of driver mutations

All driver mutations identified across all tumors in the cohort (point mutations, CNAs, and genomic rearrangements, as well as biallelic events, described in previous sections) were integrated at the tumor level to obtain the whole-genome panorama of driver events. This panorama is presented at the level of single samples in Figure 3a and summarized as tumor-type bar plot in Figure 3b. Throughout the paper we have used the panorama to answer open questions in cancer genomics, identify GEs affected by different types of driver events, and pairs or groups of driver mutations that either co-occur or are mutually exclusive across tumors. The panorama of driver mutations in the PCAWG tumors can be explored via prepared Gitools[36] interactive heatmaps (http//www.gitools.org/pcawg) and browsed in IntOGen[16], at http://www.intogen.org/pcawg.

Gene expression changes caused by different types of driver alteration

To test the hypothesis that events affecting the same coding gene cause the same alteration to the expression of the gene –as in the case of promoter driver mutations, CNAs, and SVs– we compared the expression of several genes (Zscore normalized across cohorts) in various sets of tumors with different genotypes via a Wilcoxon-Mann-Whitney test.

Therapeutic landscape of PCAWG tumors

To match drugs employed in the clinic, both approved and under trials, as well as small molecules in pre-clinical research to the driver alterations in tumors of the cohort, we employed the Cancer Biomarkers Database and the Cancer Actionability Database[22]. We also considered the possibilities of repurposing drugs in any of these three groups to counteract a different alteration or to use in another tumor type. To compute the contribution of different types of mutations to the therapeutic landscape, we only accounted for matches

that i) involve a drug approved for its use in the clinic or in clinical trials; ii) elicit sensitivity to the drug; iii) occur in the tumor type of the clinical guidelines specified for the biomarker.

REFERENCES

1. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458,** 719–24 (2009).
2. Futreal, A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4,** 177–183 (2004).
3. Zack, T. I. *et al.* Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45,** 1134–40 (2013).
4. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–8 (2013).
5. Gonzalez-Perez, A. & Lopez-Bigas, N. Functional impact bias reveals cancer drivers. *Nucleic Acids Res.* **40,** e169 (2012).
6. Gonzalez-Perez, A. *et al.* Computational approaches to identify functional genetic variants in cancer genomes. *Nat. Methods* **10,** 723–729 (2013).
7. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502,** 333–9 (2013).
8. Horn, S. *et al.* TERT promoter mutations in familial and sporadic melanoma. *Science* **339,** 959–61 (2013).
9. Huang, F. W. *et al.* Highly recurrent TERT promoter mutations in human melanoma. *Science* **339,** 957–9 (2013).
10. Puente, X. S. *et al.* Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* **526,** 519–24 (2015).
11. Weischenfeldt, J. *et al.* Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat. Genet.* **49,** 65–74 (2016).
12. Davis, C. *et al.* The Somatic Genomic Landscape of Chromophobe Renal Cell Carcinoma. *Cancer Cell* **26,** 319–330 (2014).
13. Northcott, P. A. *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nat. Publ. Gr.* **547,** (2017).
14. Morganella, S. *et al.* The topography of mutational processes in breast cancer genomes. *Nat. Commun.* **7,** 11383 (2016).
15. Martincorena, I. *et al.* Universal Patterns Of Selection In Cancer And Somatic Tissues. *bioRxiv* (2017).
16. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505,** 495–501 (2014).
17. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* **10,** 1081–2 (2013).
18. Rubio-Perez, C. *et al.* In Silico Prescription of Anticancer Drugs to Cohorts of 28 Tumor Types Reveals Targeting Opportunities. *Cancer Cell* **27,** 382–396 (2015).
19. Gerstung, M. *et al.* The evolutionary history of 2,658 cancers. (2017). doi:10.1101/161562

20. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science (80-.).* **355,** eaaf8399 (2017).

21. Ciriello, G. *et al.* Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45,** 1127–33 (2013).

22. Tamborero, D. *et al.* Cancer Genome Interpreter Annotates The Biological And Clinical Relevance Of Tumor Alterations. *bioRxiv* (2017).

23. Yamaguchi, S. *et al.* Eribulin Mesylate Targets Human Telomerase Reverse Transcriptase in Ovarian Cancer Cells. *PLoS One* **9,** e112438 (2014).

24. Colli, L. M. *et al.* Burden of Nonsynonymous Mutations among TCGA Cancers and Candidate Immune Checkpoint Inhibitor Responses. *Cancer Res.* **76,** 3767–3772 (2016).

25. Knudson, A. G. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. U. S. A.* **68,** 820–3 (1971).

26. Lindeboom, R. G. H., Supek, F. & Lehner, B. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* **48,** 1112–1118 (2016).

27. Jiao, Y. *et al.* DAXX/ATRX, MEN1, and mTOR Pathway Genes Are Frequently Altered in Pancreatic Neuroendocrine Tumors. *Science (80-.).* **331,** 1199–1203 (2011).

28. Feng, C. *et al.* PI3K? Inhibitor TGX221 Selectively Inhibits Renal Cell Carcinoma Cells with Both VHL and SETD2 mutations and Links Multiple Pathways. *Sci. Rep.* **5,** 9465 (2015).

29. Castro-Giner, F., Ratcliffe, P. & Tomlinson, I. The mini-driver model of polygenic cancer evolution. *Nat. Rev. Cancer* **15,** 680–685 (2015).

30. DeGregori, J. Connecting cancer to its causes requires incorporation of effects on tissue microenvironments. *Cancer Res.* (2017).

31. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. doi:10.1186/s13059-016-0974-4

32. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12,** R41 (2011).

33. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46,** 310–5 (2014).

34. Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42,** D980-5 (2014).

35. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nat. Publ. Gr.* **536,** (2016).

36. Perez-Llamas, C. & Lopez-Bigas, N. Gitools: analysis and visualisation of genomic data using interactive heat-maps. *PLoS One* **6,** e19541 (2011).

Acknowledgments

NL-B acknowledges funding by the European Research Council (Consolidator Grant 682398). IRB Barcelona is a recipient of a Severo Ochoa Centre of Excellence Award from the Spanish Ministry of Economy and Competitiveness (MINECO; Government of Spain) and is supported by CERCA (Generalitat de Catalunya). R.S. is supported by an EMBO Long-Term Fellowship (ALTF 568-2014) co-funded by the European Commission (EMBOCOFUND2012, GA-2012-600394) support from Marie Curie Actions. A.G.-P. is supported by a Ramón y Cajal contract (RYC-2013-14554). O.P. is supported by the BIST PhD Fellowship. S.M.W. received funding through an SNSF Early Postdoc Mobility Fellowship (P2ELP3_155365) and an EMBO Long-Term Fellowship (ALTF 755-2014). The results published here are in part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/. We thank, in particular, members of the Technical Working Group of PCAWG for their work in generating the somatic mutation calls used in this work.

Figure 1

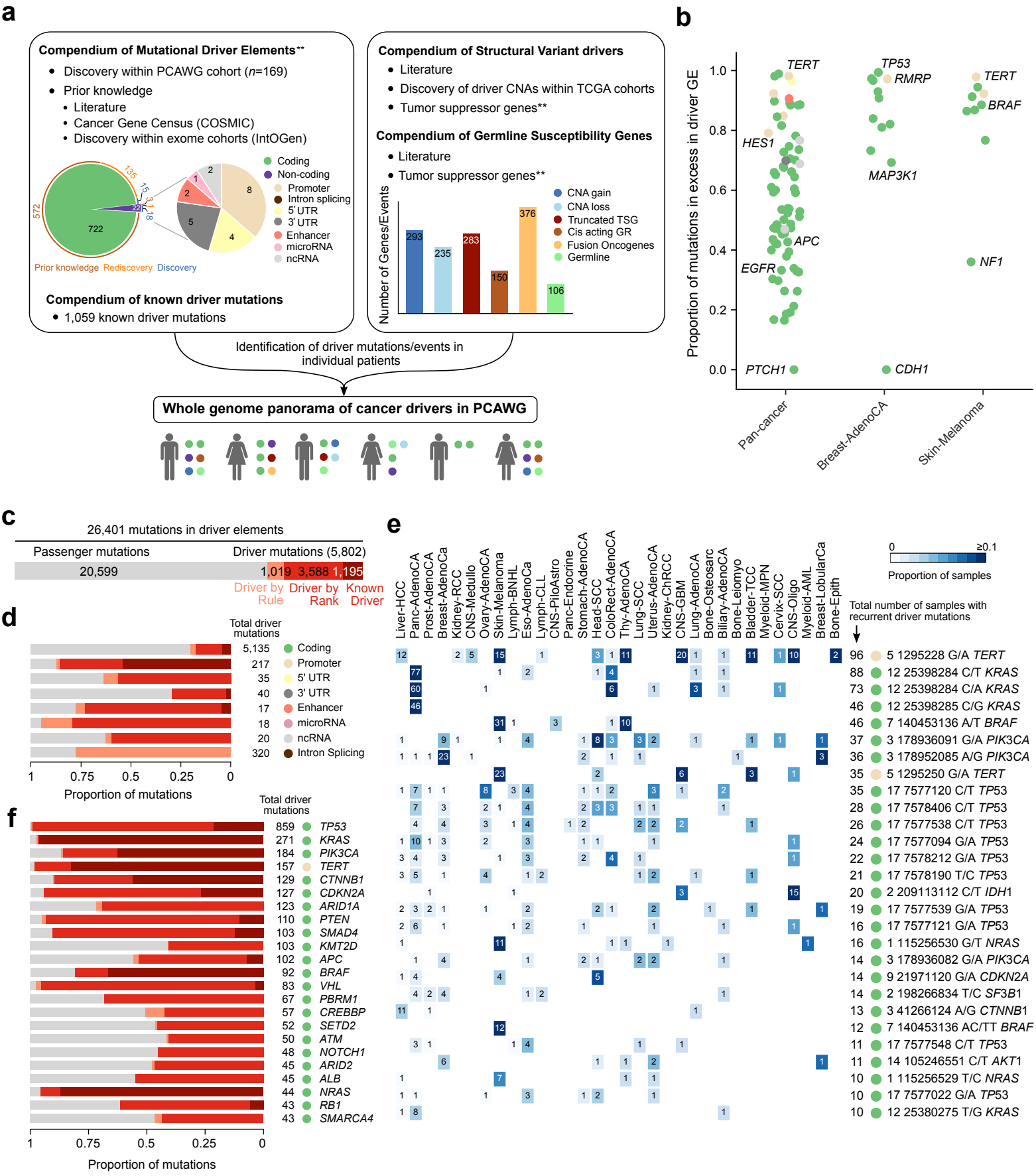

a
Compendium of Mutational Driver Elements**
Discovery within PCAWG cohort (n=169)
Prior knowledge
Literature
Cancer Gene Census (COSMIC)
Discovery within exome cohorts (IntOGen)
Coding
Non-coding
Promoter
Intron splicing
5' UTR
3' UTR
Enhancer
microRNA
ncRNA
Prior knowledge
Rediscovery
Discovery
Compendium of known driver mutations
1,059 known driver mutations
Compendium of Structural Variant drivers
Literature
Discovery of driver CNAs within TCGA cohorts
Tumor suppressor genes**
Compendium of Germline Susceptibility Genes
Literature
Tumor suppressor genes**
Number of Genes/Events
CNA gain
CNA loss
Truncated TSG
Cis acting GR
Fusion Oncogenes
Germline
Identification of driver mutations/events in individual patients
Whole genome panorama of cancer drivers in PCAWG
b
Proportion of mutations in excess in driver GE
TERT
TP53
RMRP
TERT
BRAF
HES1
MAP3K1
APC
EGFR
NF1
PTCH1
CDH1
Pan-cancer
Breast-AdenoCA
Skin-Melanoma
c
26,401 mutations in driver elements
Passenger mutations
Driver mutations (5,802)
20,599
1,019
3,588
1,195
Driver by Rule
Driver by Rank
Known Driver
d
Total driver mutations
5,135 Coding
217 Promoter
35 5' UTR
40 3' UTR
17 Enhancer
18 microRNA
20 ncRNA
320 Intron Splicing
Proportion of mutations
e
Liver-HCC
Panc-AdenoCA
Prost-AdenoCA
Breast-AdenoCa
Kidney-RCC
CNS-Medullo
Ovary-AdenoCA
Skin-Melanoma
Lymph-BNHL
Eso-AdenoCa
Lymph-CLL
CNS-PiloAstro
Panc-Endocrine
Stomach-AdenoCA
Head-SCC
ColoRect-AdenoCA
Thy-AdenoCA
Lung-SCC
Uterus-AdenoCA
Kidney-ChRCC
CNS-GBM
Lung-AdenoCA
Bone-Osteosarc
Biliary-AdenoCA
Bone-Leiomyo
Bladder-TCC
Myeloid-MPN
Cervix-SCC
CNS-Oligo
Myeloid-AML
Breast-LobularCa
Bone-Epith
Proportion of samples
Total number of samples with recurrent driver mutations
96 5 1295228 G/A TERT
88 12 25398284 C/T KRAS
73 12 25398284 C/A KRAS
46 12 25398285 C/G KRAS
46 7 140453136 A/T BRAF
37 3 178936091 G/A PIK3CA
36 3 178952085 A/G PIK3CA
35 5 1295250 G/A TERT
35 17 7577120 C/T TP53
28 17 7578406 C/T TP53
26 17 7577538 C/T TP53
24 17 7577094 G/A TP53
22 17 7578212 C/T TP53
21 17 7578190 T/C TP53
20 2 209113112 C/T IDH1
19 17 7577539 G/A TP53
16 17 7577121 G/A TP53
16 1 115256530 G/T NRAS
14 3 178936082 G/A PIK3CA
14 9 21971120 G/A CDKN2A
14 2 198266834 T/C SF3B1
13 3 41266124 A/G CTNNB1
12 7 140453136 AC/TT BRAF
11 17 7577548 C/T TP53
11 14 105246551 C/T AKT1
10 1 115256529 T/C NRAS
10 17 7577022 G/A TP53
10 12 25380275 T/G KRAS
f
Total driver mutations
859 TP53
271 KRAS
184 PIK3CA
157 TERT
129 CTNNB1
127 CDKN2A
123 ARID1A
110 PTEN
103 SMAD4
103 KMT2D
102 APC
92 BRAF
83 VHL
67 PBRM1
57 CREBBP
52 SETD2
50 ATM
48 NOTCH1
45 ARID2
45 ALB
44 NRAS
43 RB1
43 SMARCA4
Proportion of mutations

Figure 2

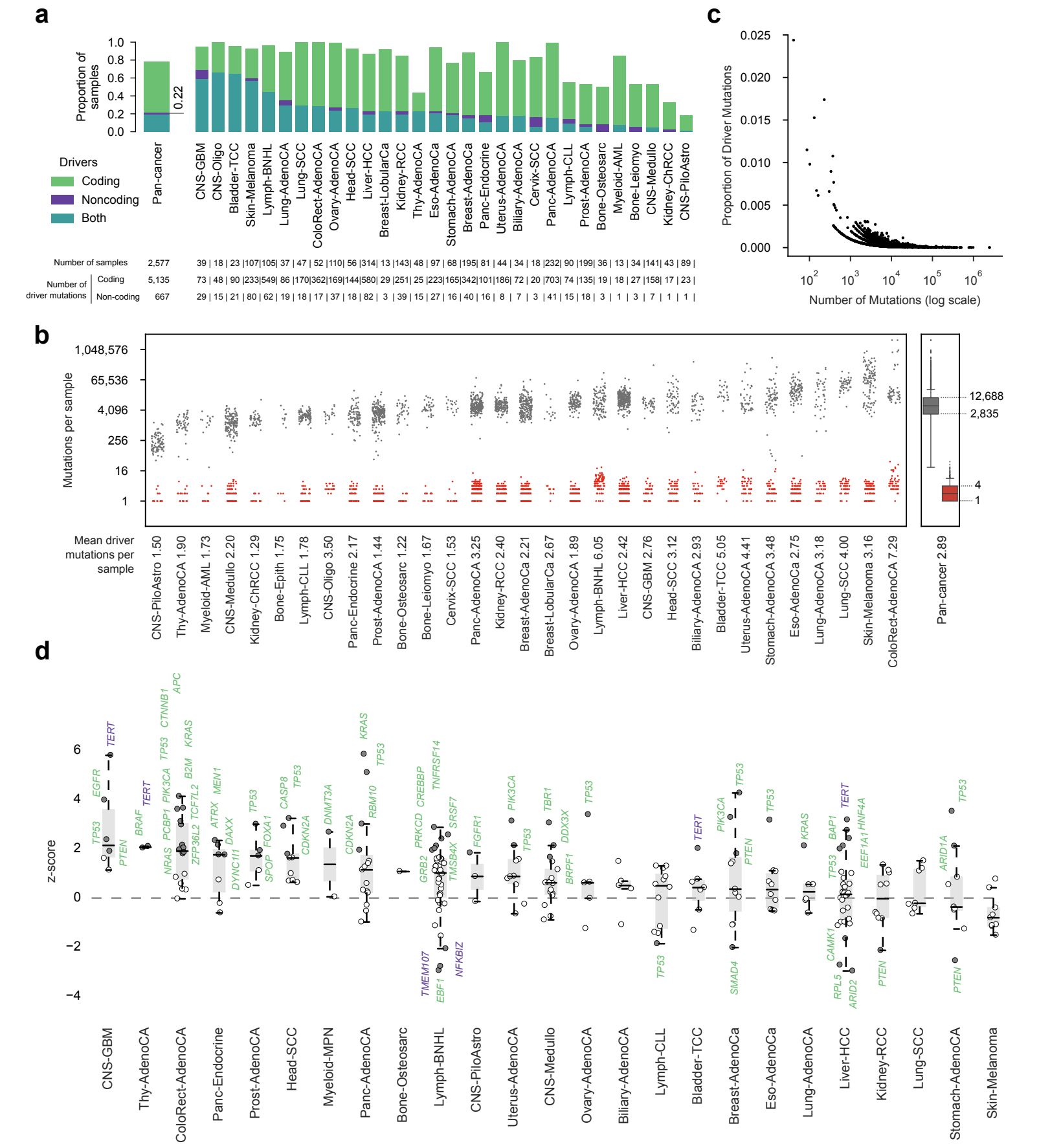

a
Proportion of samples
1.0
0.8
0.6
0.4
0.2
0.0
0.22
Drivers
Coding
Noncoding
Both
Pan-cancer
CNS-GBM
CNS-Oligo
Bladder-TCC
Skin-Melanoma
Lymph-BNHL
Lung-AdenoCA
Lung-SCC
ColoRect-AdenoCA
Ovary-AdenoCA
Head-SCC
Liver-HCC
Breast-LobularCa
Kidney-RCC
Thy-AdenoCA
Eso-AdenoCa
Stomach-AdenoCA
Breast-AdenoCa
Panc-Endocrine
Uterus-AdenoCA
Biliary-AdenoCA
Cervix-SCC
Panc-AdenoCA
Lymph-CLL
Prost-AdenoCA
Bone-Osteosarc
Myeloid-AML
Bone-Leiomyo
CNS-Medullo
Kidney-ChRCC
CNS-PiloAstro
Number of samples 2,577 39 | 18 | 23 |107|105| 37 | 47 | 52 |110| 56 |314| 13 |143| 48 | 97 | 68 |195| 81 | 44 | 34 | 18 |232| 90 |199| 36 | 13 | 34 |141| 43 | 89 |
Number of driver mutations
Coding 5,135 73 | 48 | 90 |233|549| 86 |170|362|169|144|580| 29 |251| 25 |223|165|342|101|186| 72 | 20 |703| 74 |135| 19 | 18 | 27 |158| 17 | 23 |
Non-coding 667 29 | 15 | 21 | 80 | 62 | 19 | 18 | 17 | 37 | 18 | 82 | 3 | 39 | 15 | 27 | 16 | 40 | 16 | 8 | 7 | 3 | 41 | 15 | 18 | 3 | 1 | 3 | 7 | 1 | 1 |
c
Proportion of Driver Mutations
0.025
0.020
0.015
0.010
0.005
0.000
10^2
10^3
10^4
10^5
10^6
Number of Mutations (log scale)
b
Mutations per sample
1,048,576
65,536
4,096
256
16
1
12,688
2,835
4
1
Mean driver mutations per sample
CNS-PiloAstro 1.50
Thy-AdenoCA 1.90
Myeloid-AML 1.73
CNS-Medullo 2.20
Kidney-ChRCC 1.29
Bone-Epith 1.75
Lymph-CLL 1.78
CNS-Oligo 3.50
Panc-Endocrine 2.17
Prost-AdenoCA 1.44
Bone-Osteosarc 1.22
Bone-Leiomyo 1.67
Cervix-SCC 1.53
Panc-AdenoCA 3.25
Kidney-RCC 2.40
Breast-AdenoCa 2.21
Breast-LobularCa 2.67
Ovary-AdenoCA 1.89
Lymph-BNHL 6.05
Liver-HCC 2.42
CNS-GBM 2.76
Head-SCC 3.12
Biliary-AdenoCA 2.93
Bladder-TCC 5.05
Uterus-AdenoCA 4.41
Stomach-AdenoCA 3.48
Eso-AdenoCa 2.75
Lung-AdenoCA 3.18
Lung-SCC 4.00
Skin-Melanoma 3.16
ColoRect-AdenoCA 7.29
Pan-cancer 2.89
d
z-score
6
4
2
0
−2
−4
TP53 EGFR PTEN TERT
BRAF TERT
NRAS PCBP1 PIK3CA TP53 CTNNB1 APC
ZFP36L2 TCF7L2 B2M KRAS
ATRX MEN1 DAXX
DYNC1I1 TP53
SPOP FOXA1 CASP8 TP53
CDKN2A
DNMT3A
CDKN2A RBM10 KRAS TP53
GRB2 PRKCD CREBBP
TMEM107 EBF1 NFKBIZ TMSB4X SRSF7 TNFRSF14
FGFR1
PIK3CA TP53
TBR1 BRPF1 DDX3X TP53
TP53
TERT
SMAD4 PIK3CA PTEN TP53
TP53
KRAS
CAMK1 RPL5 ARID2 TP53 BAP1 TERT EEF1A1 HNF4A
PTEN
ARID1A PTEN TP53
CNS-GBM
Thy-AdenoCA
ColoRect-AdenoCA
Panc-Endocrine
Prost-AdenoCA
Head-SCC
Myeloid-MPN
Panc-AdenoCA
Bone-Osteosarc
Lymph-BNHL
CNS-PiloAstro
Uterus-AdenoCA
CNS-Medullo
Ovary-AdenoCA
Biliary-AdenoCA
Lymph-CLL
Bladder-TCC
Breast-AdenoCa
Eso-AdenoCa
Lung-AdenoCA
Liver-HCC
Kidney-RCC
Lung-SCC
Stomach-AdenoCA
Skin-Melanoma

Figure 3

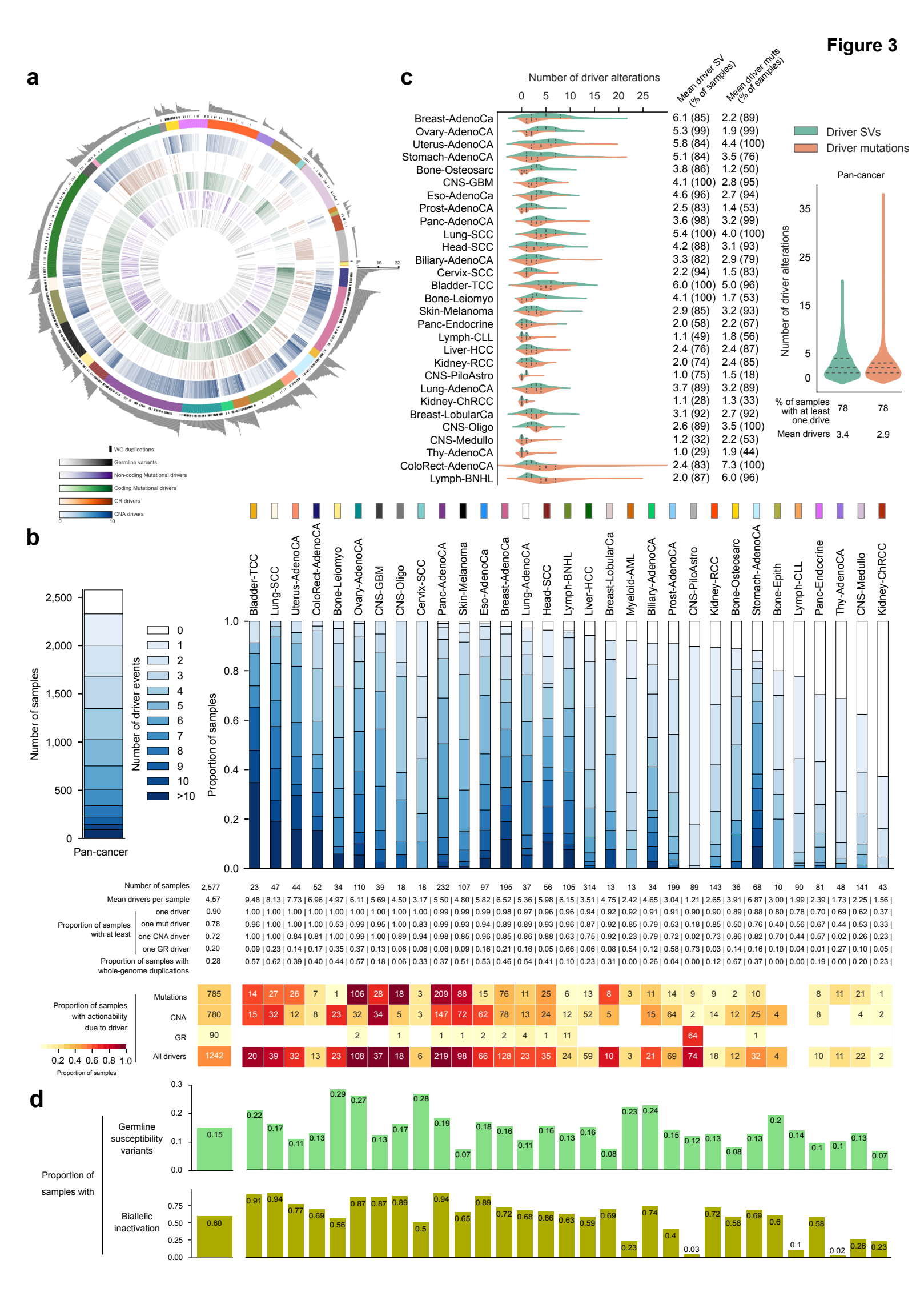

a
WG duplications
Germline variants
Non-coding Mutational drivers
Coding Mutational drivers
GR drivers
CNA drivers
c
Number of driver alterations
Mean driver SV (% of samples)
Mean driver muts (% of samples)
Breast-AdenoCa 6.1 (85) 2.2 (89)
Ovary-AdenoCA 5.3 (99) 1.9 (99)
Uterus-AdenoCA 5.8 (84) 4.4 (100)
Stomach-AdenoCA 5.1 (84) 3.5 (76)
Bone-Osteosarc 3.8 (86) 1.2 (50)
CNS-GBM 4.1 (100) 2.8 (95)
Eso-AdenoCa 4.6 (96) 2.7 (94)
Prost-AdenoCA 2.5 (83) 1.4 (53)
Panc-AdenoCA 3.6 (98) 3.2 (99)
Lung-SCC 5.4 (100) 4.0 (100)
Head-SCC 4.2 (88) 3.1 (93)
Biliary-AdenoCA 3.3 (82) 2.9 (79)
Cervix-SCC 2.2 (94) 1.5 (83)
Bladder-TCC 6.0 (100) 5.0 (96)
Bone-Leiomyo 4.1 (100) 1.7 (53)
Skin-Melanoma 2.9 (85) 3.2 (93)
Panc-Endocrine 2.0 (58) 2.2 (67)
Lymph-CLL 1.1 (49) 1.8 (56)
Liver-HCC 2.4 (76) 2.4 (87)
Kidney-RCC 2.0 (74) 2.4 (85)
CNS-PiloAstro 1.0 (75) 1.5 (18)
Lung-AdenoCA 3.7 (89) 3.2 (89)
Kidney-ChRCC 1.1 (28) 1.3 (33)
Breast-LobularCa 3.1 (92) 2.7 (92)
CNS-Oligo 2.6 (89) 3.5 (100)
CNS-Medullo 1.2 (32) 2.2 (53)
Thy-AdenoCA 1.0 (29) 1.9 (44)
ColoRect-AdenoCA 2.4 (83) 7.3 (100)
Lymph-BNHL 2.0 (87) 6.0 (96)
Driver SVs
Driver mutations
Pan-cancer
Number of driver alterations
% of samples with at least one drive 78 78
Mean drivers 3.4 2.9
b
Number of samples
Number of driver events
Pan-cancer
Proportion of samples
Number of samples 2,577 23 47 44 52 34 110 39 18 18 232 107 97 195 37 56 105 314 13 13 34 199 89 143 36 68 10 90 81 48 141 43
Mean drivers per sample 4.57 9.48 8.13 7.73 6.96 4.97 6.11 5.69 4.50 3.17 5.50 4.80 5.82 6.52 5.36 5.98 6.15 3.51 4.75 2.42 4.65 3.04 1.21 2.65 3.91 6.87 3.00 1.99 2.39 1.73 2.25 1.56
Proportion of samples with at least
one driver 0.90 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 0.99 0.99 0.99 0.98 0.97 0.96 0.96 0.94 0.92 0.92 0.91 0.91 0.90 0.90 0.89 0.88 0.80 0.78 0.70 0.69 0.62 0.37
one mut driver 0.78 0.96 1.00 1.00 1.00 0.53 0.99 0.95 1.00 0.83 0.99 0.93 0.94 0.89 0.89 0.93 0.96 0.87 0.92 0.85 0.79 0.53 0.18 0.85 0.50 0.76 0.40 0.56 0.67 0.44 0.53 0.33
one CNA driver 0.72 1.00 1.00 0.84 0.81 1.00 0.99 1.00 0.89 0.94 0.98 0.85 0.96 0.85 0.86 0.88 0.63 0.75 0.92 0.23 0.79 0.72 0.02 0.73 0.86 0.82 0.70 0.44 0.57 0.02 0.26 0.23
one GR driver 0.20 0.09 0.23 0.14 0.17 0.35 0.37 0.13 0.06 0.06 0.06 0.09 0.16 0.21 0.16 0.05 0.66 0.06 0.08 0.54 0.12 0.58 0.73 0.03 0.14 0.16 0.10 0.04 0.01 0.27 0.10 0.05
Proportion of samples with whole-genome duplications 0.28 0.57 0.62 0.39 0.40 0.44 0.57 0.18 0.06 0.33 0.37 0.51 0.53 0.46 0.54 0.41 0.10 0.23 0.31 0.00 0.26 0.04 0.00 0.12 0.67 0.37 0.00 0.00 0.19 0.00 0.20 0.23
Proportion of samples with actionability due to driver
Mutations 785 14 27 26 7 1 106 28 18 3 209 88 15 76 11 25 6 13 8 3 11 14 9 9 2 10 8 11 21 1
CNA 780 15 32 12 8 23 32 34 5 3 147 72 62 78 13 24 12 52 5 15 64 2 14 12 25 4 8 4 2
GR 90 2 1 1 1 1 2 2 4 1 11 64 1
All drivers 1242 20 39 32 13 23 108 37 18 6 219 98 66 128 23 35 24 59 10 3 21 69 74 18 12 32 4 10 11 22 2
Proportion of samples
d
Proportion of samples with
Germline susceptibility variants
Biallelic inactivation

Figure 4

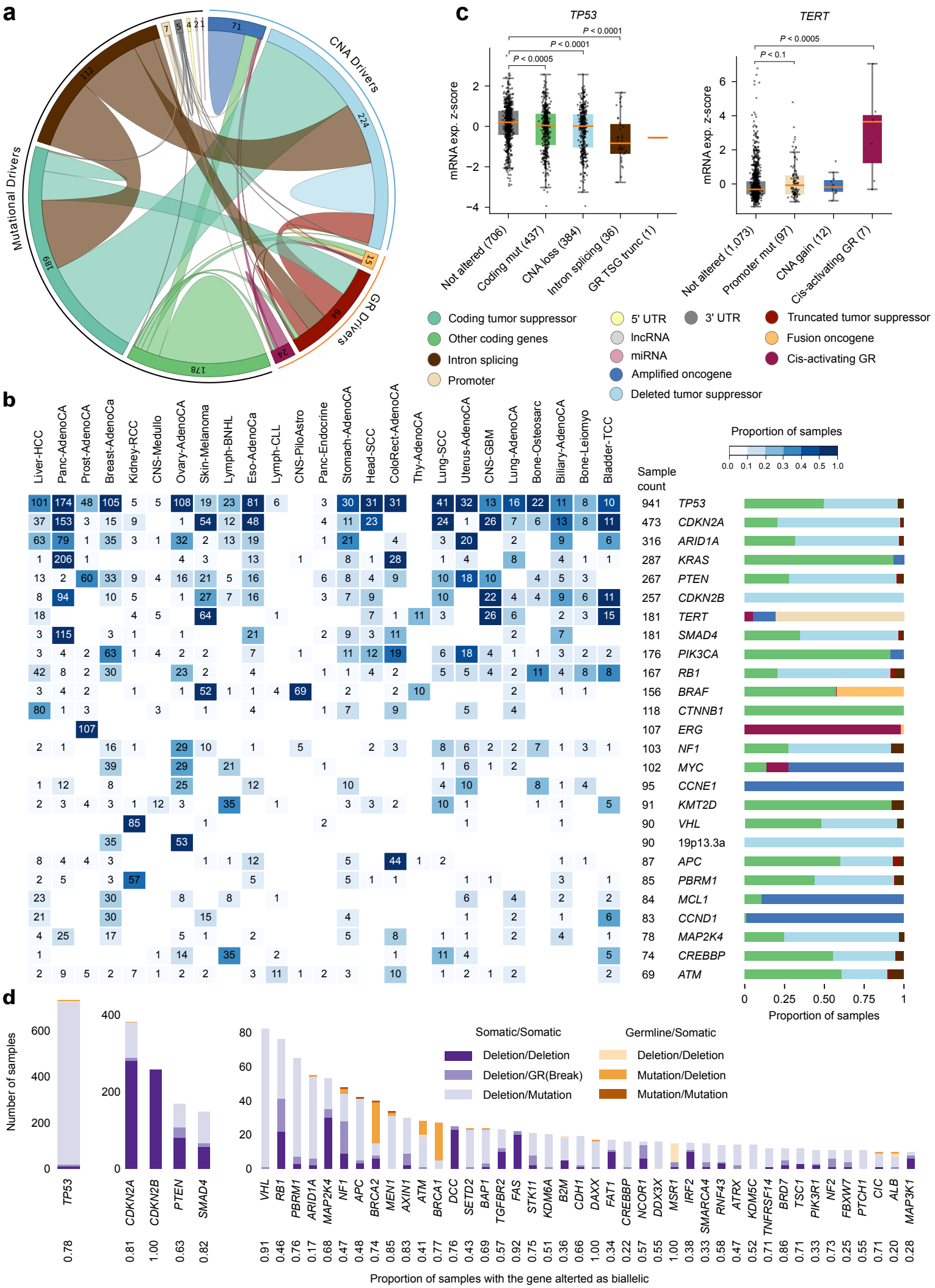
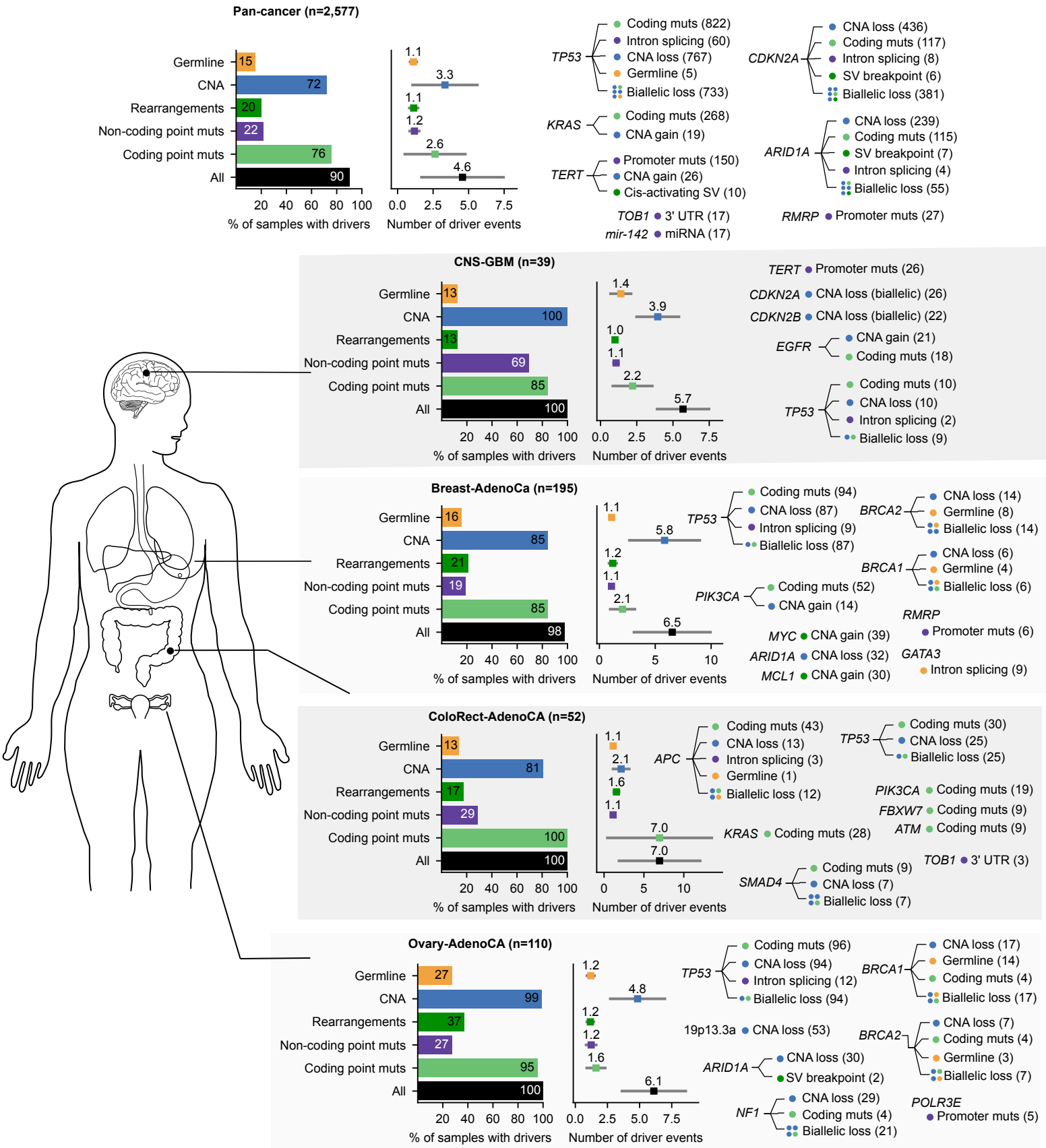
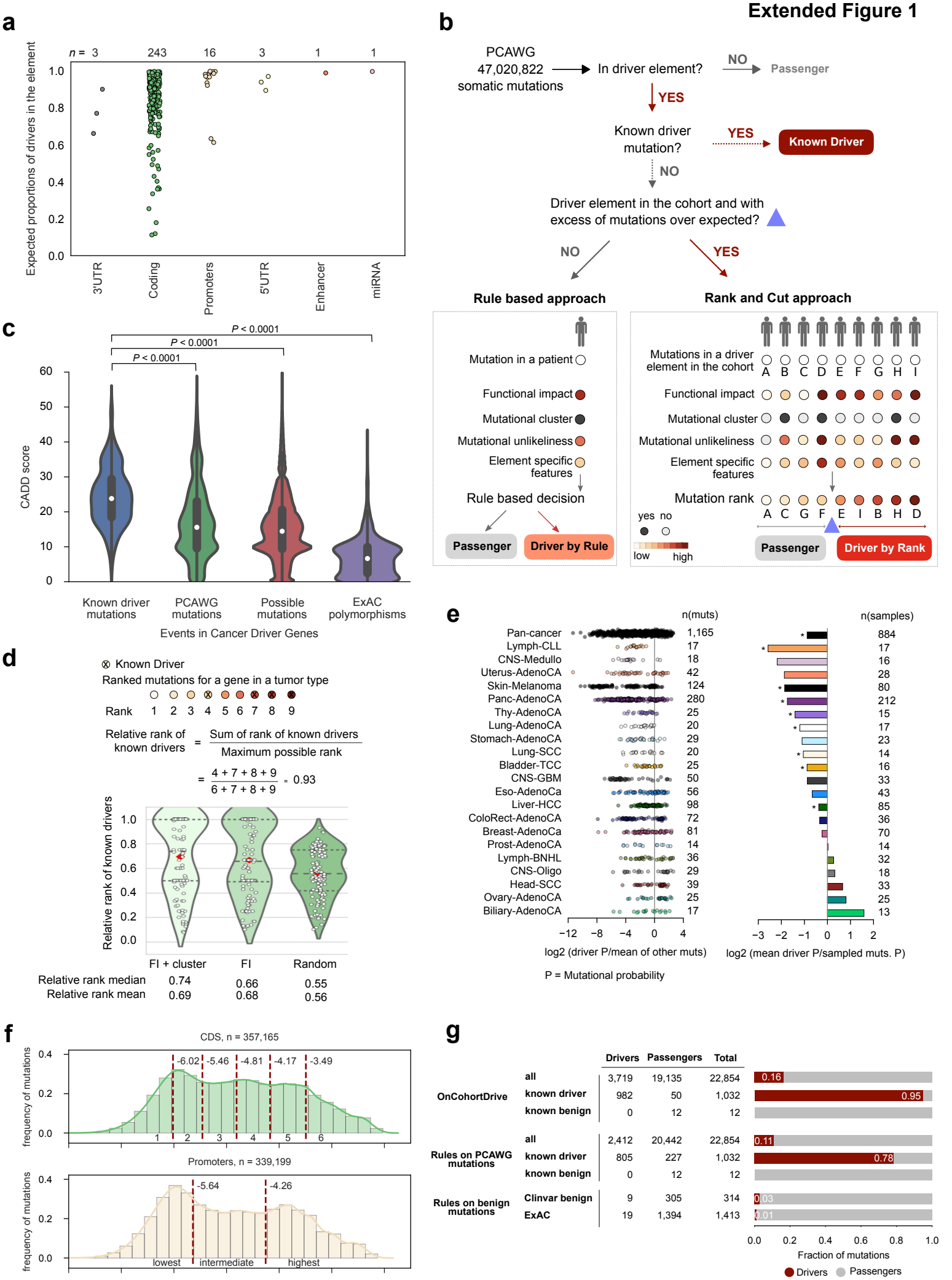

a
CNA Drivers
Mutational Drivers
GR Drivers
71
224
15
64
24
178
189
112
7
5
4
2
1
1
c
TP53
TERT
P < 0.0001
P < 0.0001
P < 0.0005
P < 0.0005
P < 0.1
mRNA exp. z-score
Not altered (706)
Coding mut (437)
CNA loss (384)
Intron splicing (36)
GR TSG trunc (1)
Not altered (1,073)
Promoter mut (97)
CNA gain (12)
Cis-activating GR (7)
Coding tumor suppressor
Other coding genes
Intron splicing
Promoter
5' UTR
lncRNA
miRNA
Amplified oncogene
Deleted tumor suppressor
3' UTR
Truncated tumor suppressor
Fusion oncogene
Cis-activating GR
b
Liver-HCC
Panc-AdenoCA
Prost-AdenoCA
Breast-AdenoCa
Kidney-RCC
CNS-Medullo
Ovary-AdenoCA
Skin-Melanoma
Lymph-BNHL
Eso-AdenoCa
Lymph-CLL
CNS-PiloAstro
Panc-Endocrine
Stomach-AdenoCA
Head-SCC
ColoRect-AdenoCA
Thy-AdenoCA
Lung-SCC
Uterus-AdenoCA
CNS-GBM
Lung-AdenoCA
Bone-Osteosarc
Biliary-AdenoCA
Bone-Leiomyo
Bladder-TCC
Sample count
Proportion of samples
0.0 0.1 0.2 0.3 0.4 0.5 1.0
941 TP53
473 CDKN2A
316 ARID1A
287 KRAS
267 PTEN
257 CDKN2B
181 TERT
181 SMAD4
176 PIK3CA
167 RB1
156 BRAF
118 CTNNB1
107 ERG
103 NF1
102 MYC
95 CCNE1
91 KMT2D
90 VHL
90 19p13.3a
87 APC
85 PBRM1
84 MCL1
83 CCND1
78 MAP2K4
74 CREBBP
69 ATM
0 0.25 0.50 0.75 1
Proportion of samples
d
Number of samples
Somatic/Somatic
Deletion/Deletion
Deletion/GR(Break)
Deletion/Mutation
Germline/Somatic
Deletion/Deletion
Mutation/Deletion
Mutation/Mutation
TP53 0.78
CDKN2A 0.81
CDKN2B 1.00
PTEN 0.63
SMAD4 0.82
VHL 0.91
RB1 0.46
PBRM1 0.76
ARID1A 0.17
MAP2K4 0.68
NF1 0.47
APC 0.48
BRCA2 0.74
MEN1 0.85
AXIN1 0.83
ATM 0.41
BRCA1 0.77
DCC 0.76
SETD2 0.43
BAP1 0.69
TGFBR2 0.57
FAS 0.92
STK11 0.75
KDM6A 0.51
B2M 0.36
CDH1 0.66
DAXX 1.00
FAT1 0.34
CREBBP 0.22
NCOR1 0.57
DDX3X 0.55
MSR1 1.00
IRF2 0.38
SMARCA4 0.33
RNF43 0.58
ATRX 0.47
KDM5C 0.52
TNFRSF14 0.71
BRD7 0.86
TSC1 0.71
PIK3R1 0.33
NF2 0.73
FBXW7 0.25
PTCH1 0.55
CIC 0.71
ALB 0.20
MAP3K1 0.28
Proportion of samples with the gene alterted as biallelic

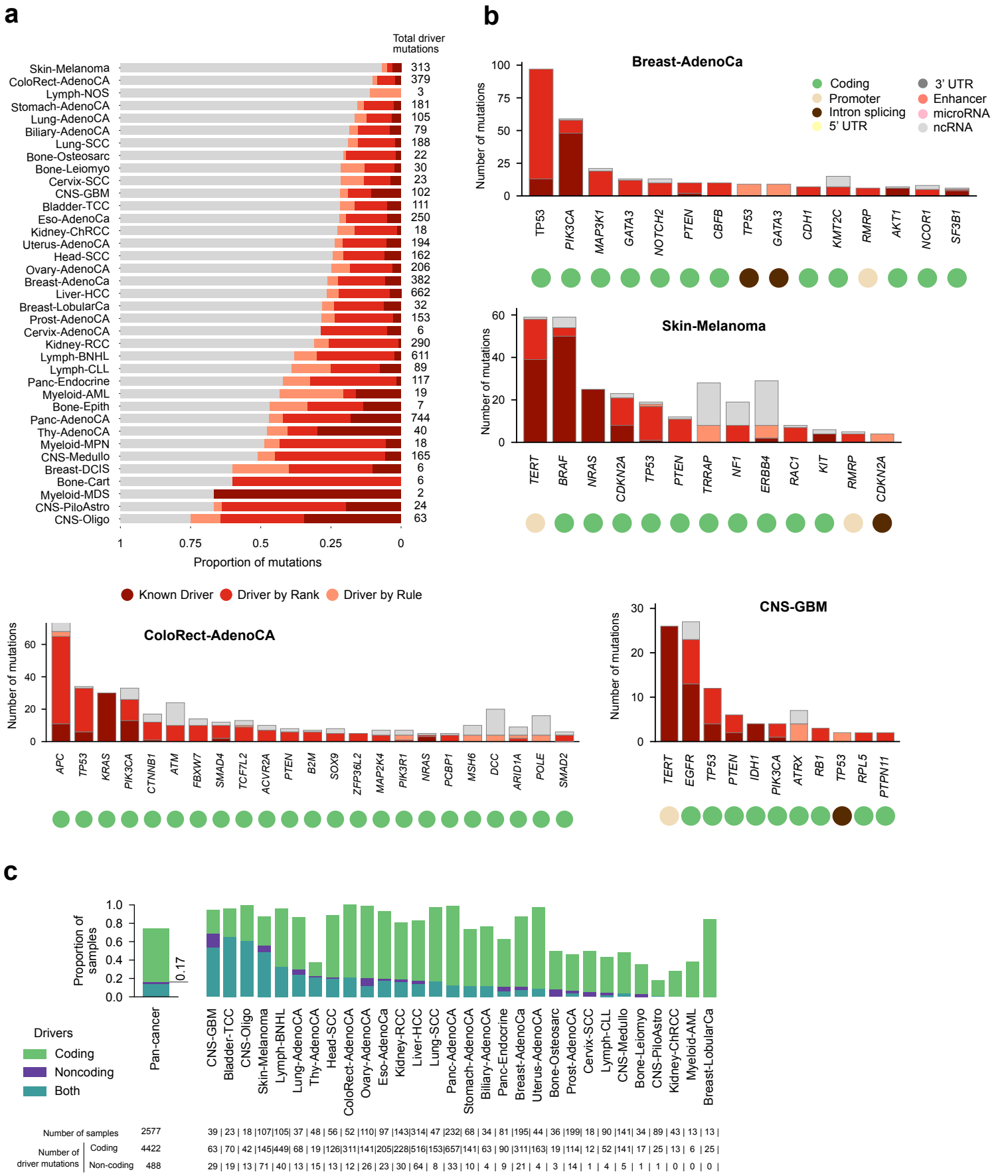
Figure 5


Pan-cancer (n=2,577)
Germline 15
CNA 72
Rearrangements 20
Non-coding point muts 22
Coding point muts 76
All 90
20 40 60 80 100
% of samples with drivers
1.1
3.3
1.1
1.2
2.6
4.6
0.0 2.5 5.0 7.5
Number of driver events
TP53 Coding muts (822)
Intron splicing (60)
CNA loss (767)
Germline (5)
Biallelic loss (733)
KRAS Coding muts (268)
CNA gain (19)
TERT Promoter muts (150)
CNA gain (26)
Cis-activating SV (10)
TOB1 3' UTR (17)
mir-142 miRNA (17)
CDKN2A CNA loss (436)
Coding muts (117)
Intron splicing (8)
SV breakpoint (6)
Biallelic loss (381)
ARID1A CNA loss (239)
Coding muts (115)
SV breakpoint (7)
Intron splicing (4)
Biallelic loss (55)
RMRP Promoter muts (27)
CNS-GBM (n=39)
Germline 13
CNA 100
Rearrangements 13
Non-coding point muts 69
Coding point muts 85
All 100
20 40 60 80 100
% of samples with drivers
1.4
3.9
1.0
1.1
2.2
5.7
0.0 2.5 5.0 7.5
Number of driver events
TERT Promoter muts (26)
CDKN2A CNA loss (biallelic) (26)
CDKN2B CNA loss (biallelic) (22)
EGFR CNA gain (21)
Coding muts (18)
TP53 Coding muts (10)
CNA loss (10)
Intron splicing (2)
Biallelic loss (9)
Breast-AdenoCa (n=195)
Germline 16
CNA 85
Rearrangements 21
Non-coding point muts 19
Coding point muts 85
All 98
20 40 60 80 100
% of samples with drivers
1.1
5.8
1.2
1.1
2.1
6.5
0 5 10
Number of driver events
TP53 Coding muts (94)
CNA loss (87)
Intron splicing (9)
Biallelic loss (87)
PIK3CA Coding muts (52)
CNA gain (14)
MYC CNA gain (39)
ARID1A CNA loss (32)
MCL1 CNA gain (30)
BRCA2 CNA loss (14)
Germline (8)
Biallelic loss (14)
BRCA1 CNA loss (6)
Germline (4)
Biallelic loss (6)
RMRP
Promoter muts (6)
GATA3
Intron splicing (9)
ColoRect-AdenoCA (n=52)
Germline 13
CNA 81
Rearrangements 17
Non-coding point muts 29
Coding point muts 100
All 100
20 40 60 80 100
% of samples with drivers
1.1
2.1
1.6
1.1
7.0
7.0
0 5 10
Number of driver events
APC Coding muts (43)
CNA loss (13)
Intron splicing (3)
Germline (1)
Biallelic loss (12)
KRAS Coding muts (28)
SMAD4 Coding muts (9)
CNA loss (7)
Biallelic loss (7)
TP53 Coding muts (30)
CNA loss (25)
Biallelic loss (25)
PIK3CA Coding muts (19)
FBXW7 Coding muts (9)
ATM Coding muts (9)
TOB1 3' UTR (3)
Ovary-AdenoCA (n=110)
Germline 27
CNA 99
Rearrangements 37
Non-coding point muts 27
Coding point muts 95
All 100
20 40 60 80 100
% of samples with drivers
1.2
4.8
1.2
1.2
1.6
6.1
0.0 2.5 5.0 7.5
Number of driver events
TP53 Coding muts (96)
CNA loss (94)
Intron splicing (12)
Biallelic loss (94)
19p13.3a CNA loss (53)
ARID1A CNA loss (30)
SV breakpoint (2)
NF1 CNA loss (29)
Coding muts (4)
Biallelic loss (21)
BRCA1 CNA loss (17)
Germline (14)
Coding muts (4)
Biallelic loss (17)
BRCA2 CNA loss (7)
Coding muts (4)
Germline (3)
Biallelic loss (7)
POLR3E
Promoter muts (5)

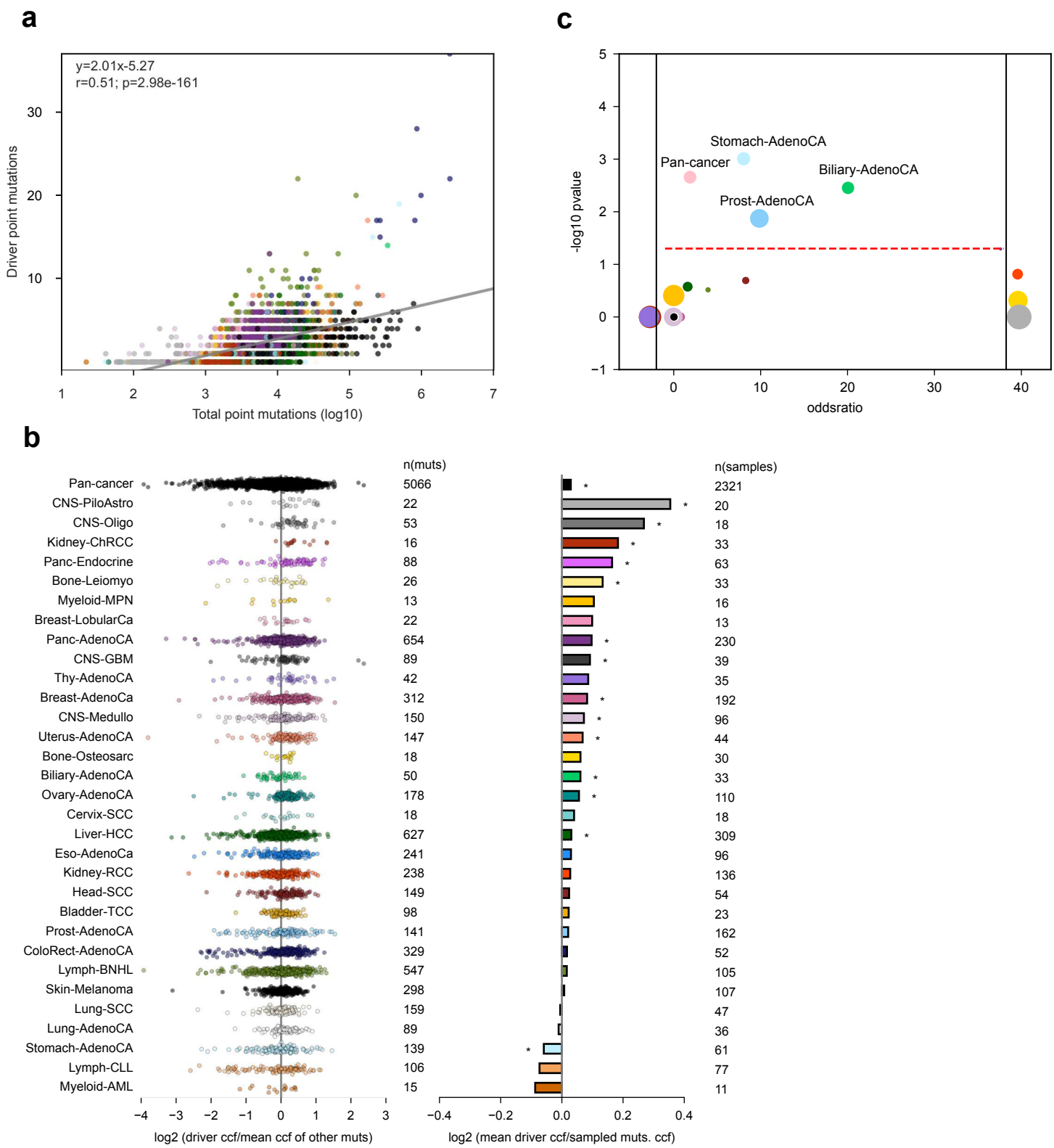
Extended Figure 1



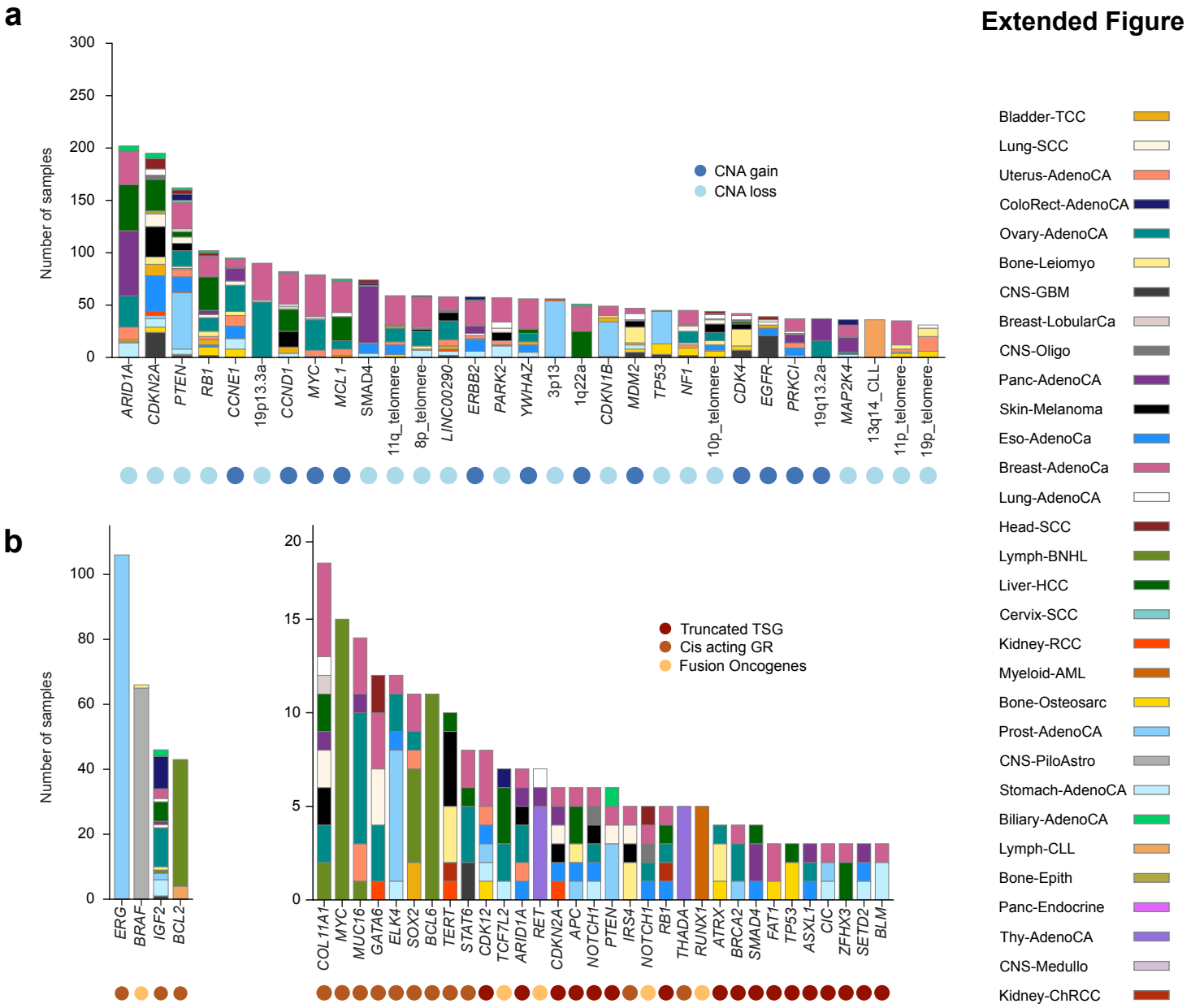
Extended Figure 2

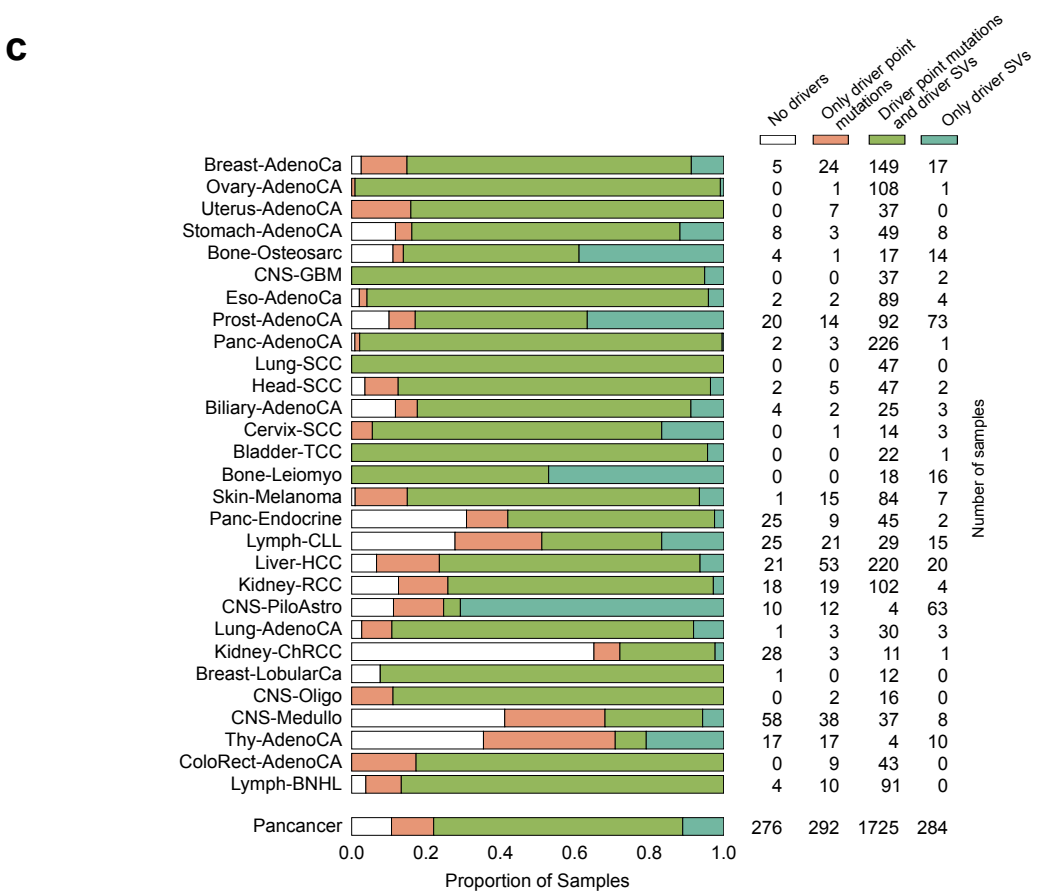


Extended Figure 3

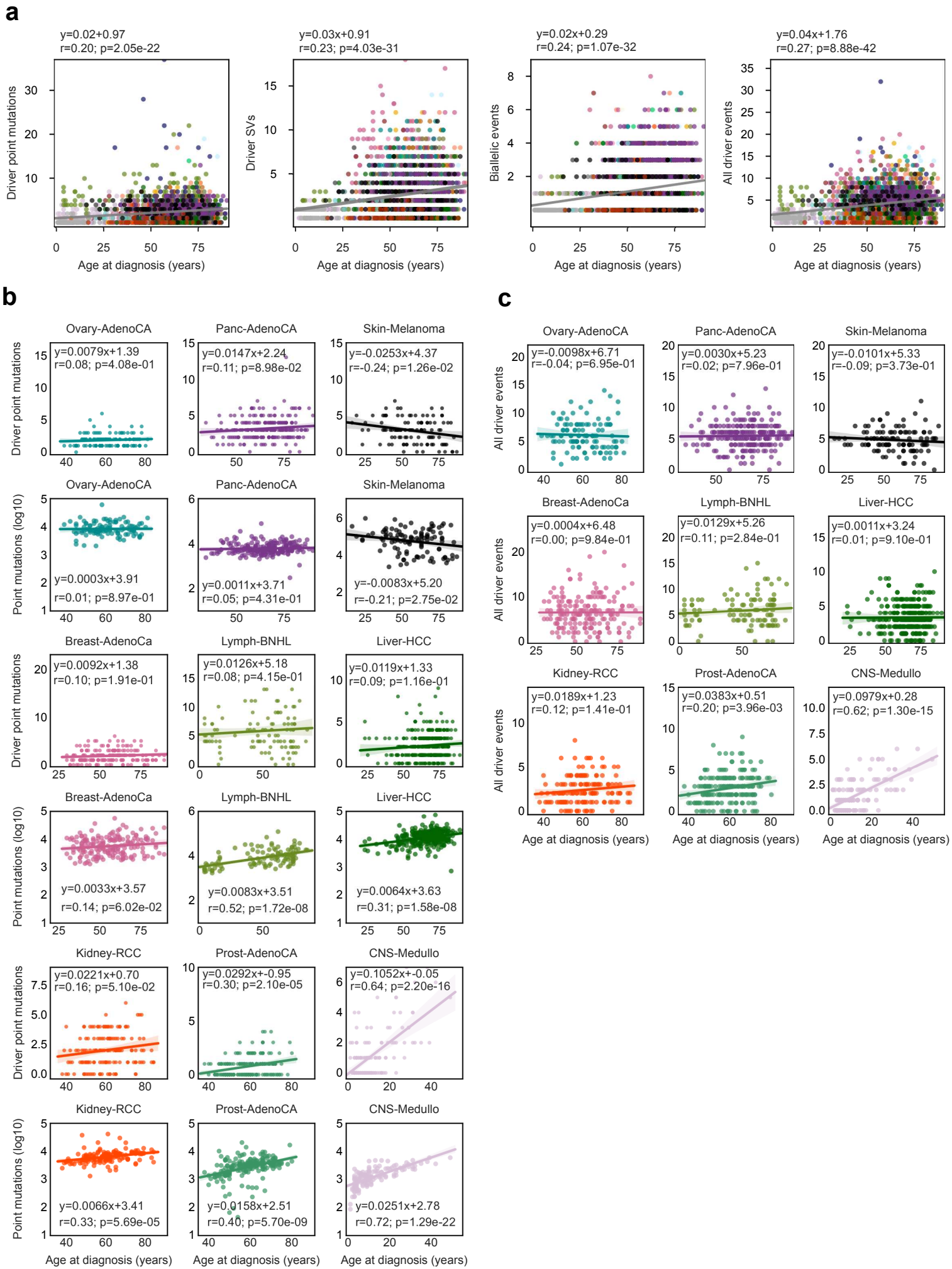


Extended Figure 4

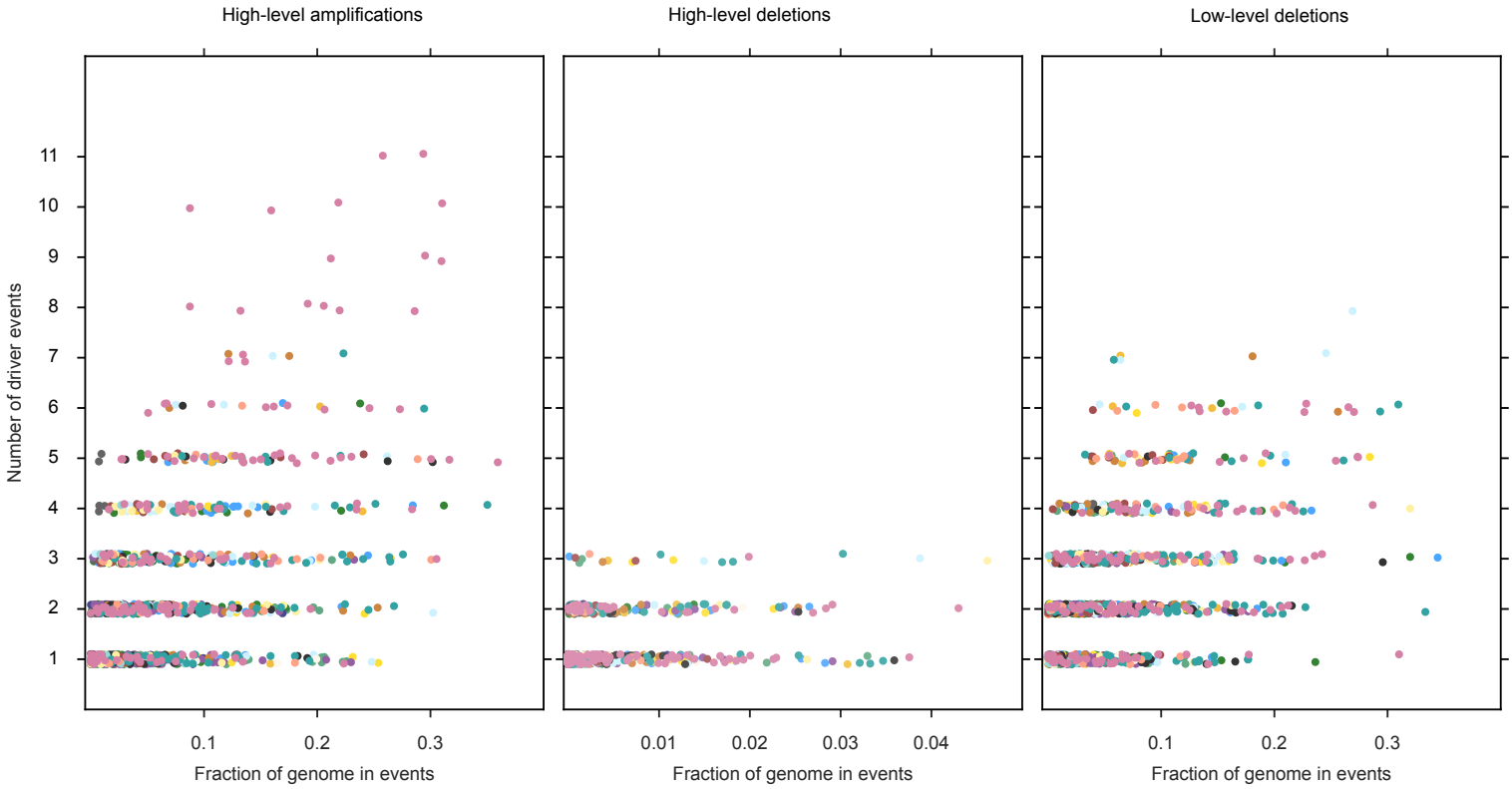




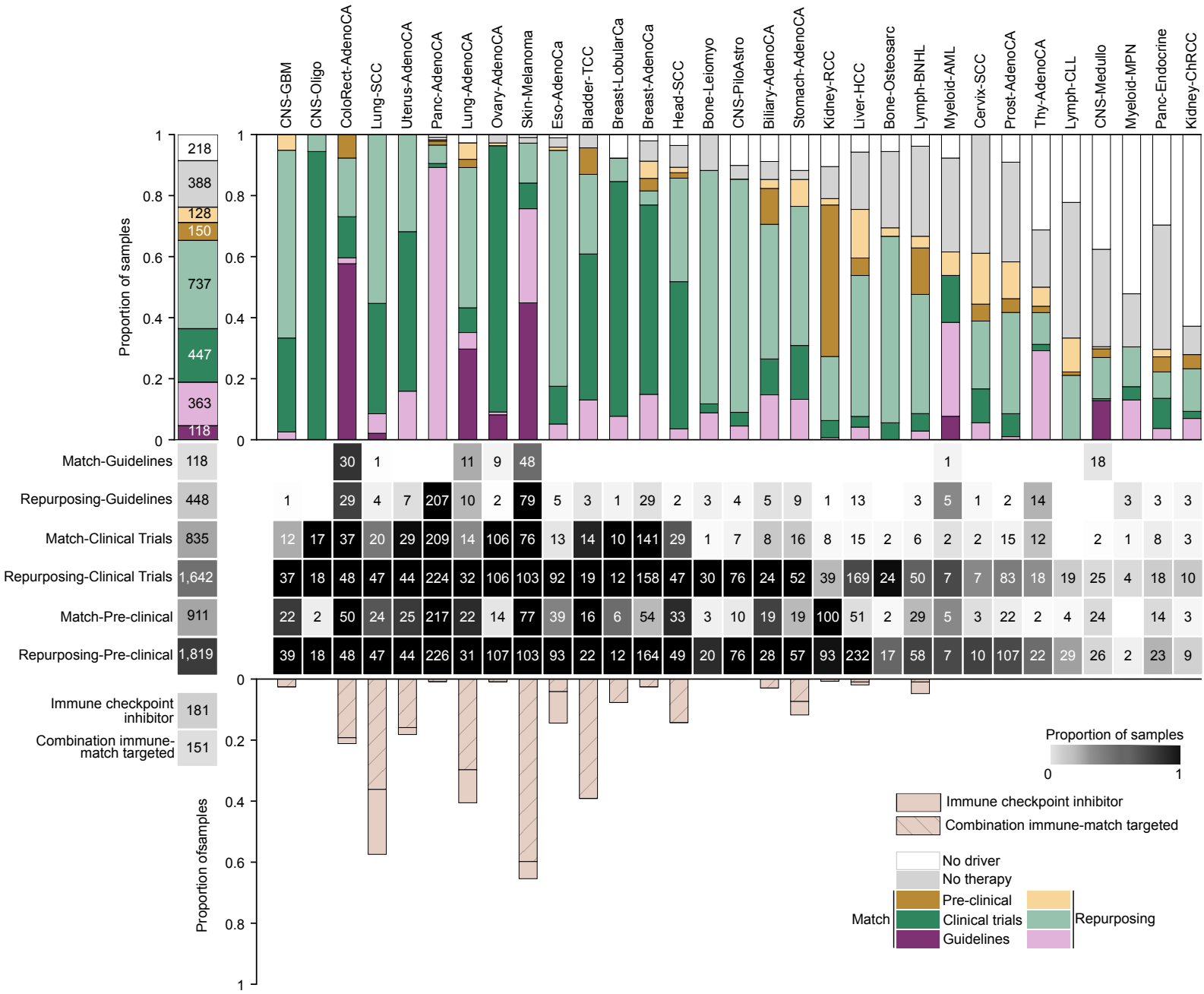
Extended Figure 5



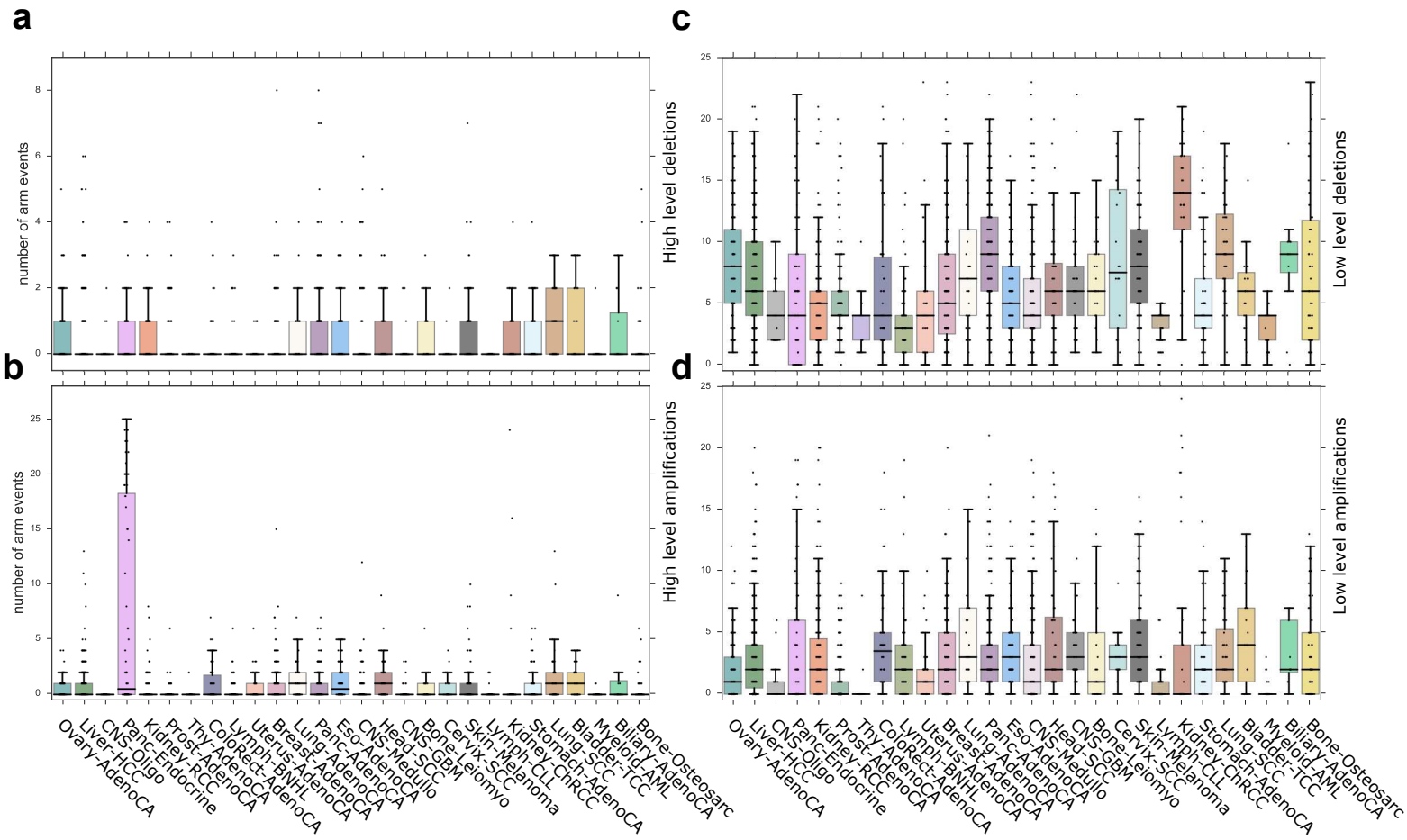
Extended Figure 6



Extended Figure 7

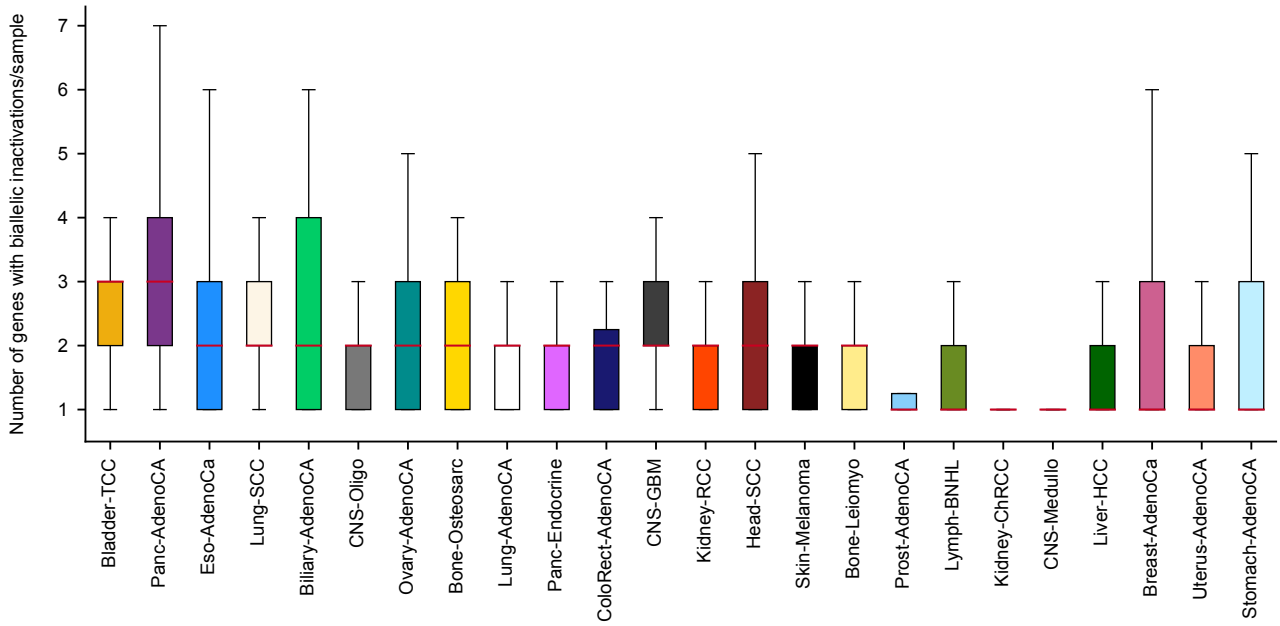


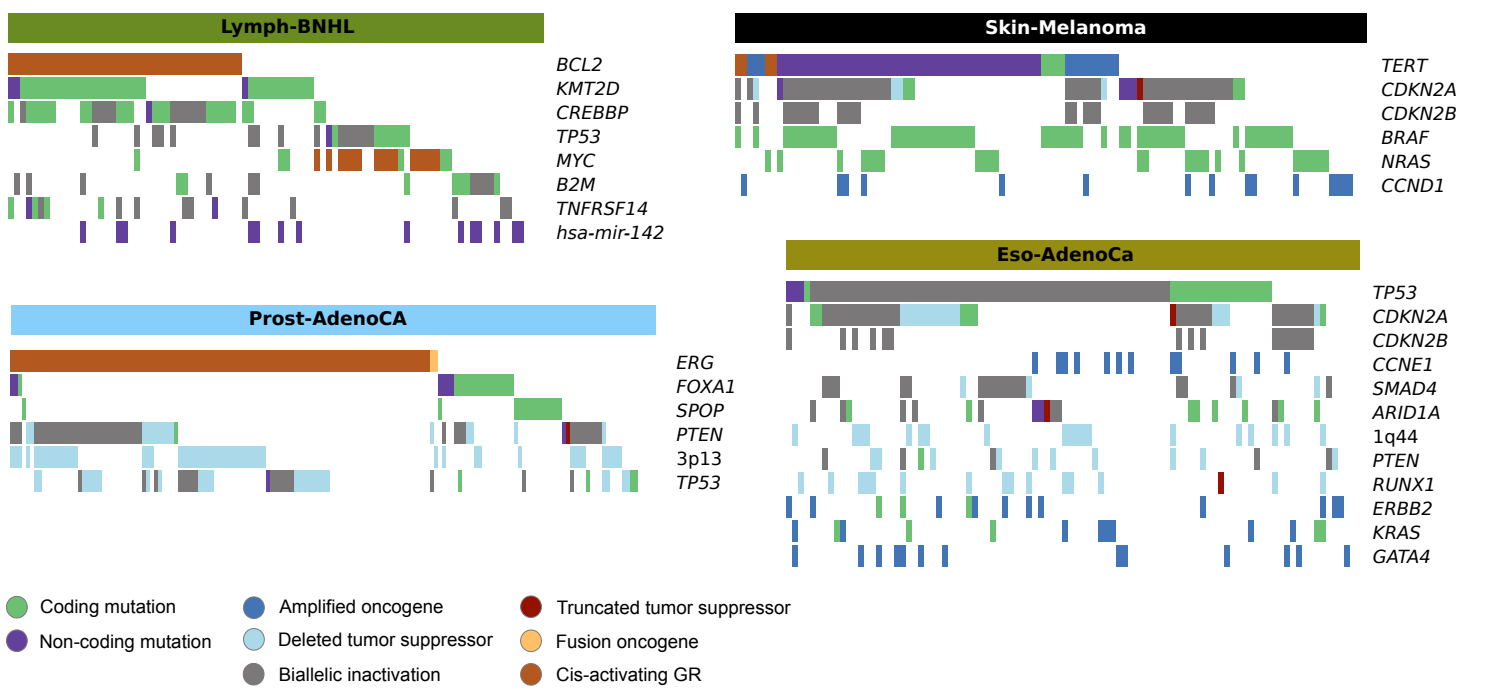
Extended Figure 8



Extended Figure 9

a



		Bladder-TCC	Panc-AdenoCA	Eso-AdenoCa	Lung-SCC	Biliary-AdenoCA	CNS-Oligo	Ovary-AdenoCA	Bone-Osteosarc	Lung-AdenoCA	Panc-Endocrine	ColoRect-AdenoCA	CNS-GBM	Kidney-RCC	Head-SCC	Skin-Melanoma	Bone-Leiomyo	Prost-AdenoCA	Lymph-BNHL	Kidney-ChRCC	CNS-Medullo	Liver-HCC	Breast-AdenoCa	Uterus-AdenoCA	Stomach-AdenoCA
Total samples in the cohort		23	232	97	47	34	18	110	36	37	81	52	39	143	56	107	34	199	105	43	141	314	195	44	68
Number of samples with biallelic incativation in	only 1 gene	21	217	86	44	25	16	96	21	25	47	36	34	103	37	70	19	80	66	10	36	185	140	34	47
	>1 gene	3	39	30	10	8	7	44	8	9	22	15	8	39	16	26	7	60	38	9	33	104	72	19	27
	Total	18	178	56	34	17	9	52	13	16	25	21	26	64	21	44	12	20	28	1	3	81	68	15	20

b