Pan-cancer analysis of whole genomes

https://doi.org/10.1038/s41586-020-1969-6

Received: 29 July 2018

Accepted: 11 December 2019

Published online: 5 February 2020

Open access

The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium

Cancer is driven by genetic change, and the advent of massively parallel sequencing has enabled systematic documentation of this variation at the whole-genome scale[1–3]. Here we report the integrative analysis of 2,658 whole-cancer genomes and their matching normal tissues across 38 tumour types from the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium of the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA). We describe the generation of the PCAWG resource, facilitated by international data sharing using compute clouds. On average, cancer genomes contained 4–5 driver mutations when combining coding and non-coding genomic elements; however, in around 5% of cases no drivers were identified, suggesting that cancer driver discovery is not yet complete. Chromothripsis, in which many clustered structural variants arise in a single catastrophic event, is frequently an early event in tumour evolution; in acral melanoma, for example, these events precede most somatic point mutations and affect several cancer-associated genes simultaneously. Cancers with abnormal telomere maintenance often originate from tissues with low replicative activity and show several mechanisms of preventing telomere attrition to critical levels. Common and rare germline variants affect patterns of somatic mutation, including point mutations, structural variants and somatic retrotransposition. A collection of papers from the PCAWG Consortium describes non-coding mutations that drive cancer beyond those in the *TERT* promoter[4]; identifies new signatures of mutational processes that cause base substitutions, small insertions and deletions and structural variation[5,6]; analyses timings and patterns of tumour evolution[7]; describes the diverse transcriptional consequences of somatic mutation on splicing, expression levels, fusion genes and promoter activity[8,9]; and evaluates a range of more-specialized features of cancer genomes[8,10–18].

Cancer is the second most-frequent cause of death worldwide, killing more than 8 million people every year; the incidence of cancer is expected to increase by more than 50% over the coming decades[19,20]. 'Cancer' is a catch-all term used to denote a set of diseases characterized by autonomous expansion and spread of a somatic clone. To achieve this behaviour, the cancer clone must co-opt multiple cellular pathways that enable it to disregard the normal constraints on cell growth, modify the local microenvironment to favour its own proliferation, invade through tissue barriers, spread to other organs and evade immune surveillance[21]. No single cellular program directs these behaviours. Rather, there is a large pool of potential pathogenic abnormalities from which individual cancers draw their own combinations: the commonalities of macroscopic features across tumours belie a vastly heterogeneous landscape of cellular abnormalities.

This heterogeneity arises from the stochastic nature of Darwinian evolution. There are three preconditions for Darwinian evolution: characteristics must vary within a population; this variation must be heritable from parent to offspring; and there must be competition for survival within the population. In the context of somatic cells, heritable variation arises from mutations acquired stochastically throughout life, notwithstanding additional contributions from germline and epigenetic variation. A subset of these mutations alter the cellular phenotype, and a small subset of those variants confer an advantage on clones during the competition to escape the tight physiological controls wired into somatic cells. Mutations that provide a selective advantage to the clone are termed driver mutations, as opposed to selectively neutral passenger mutations.

Initial studies using massively parallel sequencing demonstrated the feasibility of identifying every somatic point mutation, copy-number change and structural variant (SV) in a given cancer[1–3]. In 2008, recognizing the opportunity that this advance in technology provided, the global cancer genomics community established the ICGC with the goal of systematically documenting the somatic mutations that drive common tumour types[22].

The pan-cancer analysis of whole genomes

The expansion of whole-genome sequencing studies from individual ICGC and TCGA working groups presented the opportunity to undertake a meta-analysis of genomic features across tumour types. To achieve this, the PCAWG Consortium was established. A Technical Working Group implemented the informatics analyses by aggregating the raw sequencing data from different working groups that studied individual tumour types, aligning the sequences to the human genome and delivering a set of high-quality somatic mutation calls for downstream analysis (Extended Data Fig. 1). Given the recent meta-analysis

A list of members and their affiliations appears in the online version of the paper and lists of working groups appear in the Supplementary Information.

Box 1

Online resources for data access, visualization and analysis

The PCAWG landing page (http://docs.icgc.org/pcawg) provides links to several data resources for interactive online browsing, analysis and download of PCAWG data and results (Supplementary Table 4).

Direct download of PCAWG data

Aligned PCAWG read data in BAM format are also available at the European Genome Phenome Archive (EGA; https://www.ebi.ac.uk/ega/search/site/pcawg under accession number EGAS00001001692). In addition, all open-tier PCAWG genomics data, as well as reference datasets used for analysis, can be downloaded from the ICGC Data Portal at http://docs.icgc.org/pcawg/data/. Controlled-tier genomic data, including SNVs and indels that originated from TCGA projects (in VCF format) and aligned reads (in BAM format) can be downloaded using the Score (https://www.overture.bio/) software package, which has accelerated and secure file transfer, as well as BAM slicing facilities to selectively download defined regions of genomic alignments.

PCAWG computational pipelines

The core alignment, somatic variant-calling, quality-control and variant consensus-generation pipelines used by PCAWG have each been packaged into portable cross-platform images using the Dockstore system[84] and released under an Open Source licence that enables unrestricted use and redistribution. All PCAWG Dockstore images are available to the public at https://dockstore.org/organizations/PCAWG/collections/PCAWG.

ICGC Data Portal

The ICGC Data Portal[85] (https://dcc.icgc.org) serves as the main entry point for accessing PCAWG datasets with a single uniform web interface and a high-performance data-download client. This uniform interface provides users with easy access to the myriad of PCAWG sequencing data and variant calls that reside in many repositories and compute clouds worldwide. Streaming technology[86] provides users with high-level visualizations in real time of BAM and VCF files stored remotely on the Cancer Genome Collaboratory.

UCSC Xena

UCSC Xena[87] (https://pcawg.xenahubs.net) visualizes all PCAWG primary results, including copy-number, gene-expression, gene-fusion and promoter-usage alterations, simple somatic mutations, large somatic structural variations, mutational signatures and phenotypic data. These open-access data are available through a public Xena hub, and consensus simple somatic mutations can be loaded to the local computer of a user via a private Xena hub. Kaplan–Meier plots, histograms, box plots, scatter plots and transcript-specific views offer additional visualization options and statistical analyses.

The Expression Atlas

The Expression Atlas (https://www.ebi.ac.uk/gxa/home) contains RNA-sequencing and expression microarray data for querying gene expression across tissues, cell types, developmental stages and/or experimental conditions[88]. Two different views of the data are provided: summarized expression levels for each tumour type and gene expression at the level of individual samples, including reference-gene expression datasets for matching normal tissues.

PCAWG Scout

PCAWG Scout (http://pcawgscout.bsc.es/) provides a framework for -omics workflow and website templating to generate on-demand, in-depth analyses of the PCAWG data that are openly available to the whole research community. Views of protected data are available that still safeguard sensitive data. Through the PCAWG Scout web interface, users can access an array of reports and visualizations that leverage on-demand bioinformatic computing infrastructure to produce results in real time, allowing users to discover trends as well as form and test hypotheses.

Chromothripsis Explorer

Chromothripsis Explorer (http://compbio.med.harvard.edu/chromothripsis/) is a portal that allows structural variation in the PCAWG dataset to be explored on an individual patient basis through the use of circos plots. Patterns of chromothripsis can also be explored in aggregated formats.

of exome data from the TCGA Pan-Cancer Atlas[23–25], scientific working groups concentrated their efforts on analyses best-informed by whole-genome sequencing data.

We collected genome data from 2,834 donors (Extended Data Table 1), of which 176 were excluded after quality assurance. A further 75 had minor issues that could affect some of the analyses (grey-listed donors) and 2,583 had data of optimal quality (white-listed donors) (Supplementary Table 1). Across the 2,658 white- and grey-listed donors, whole-genome sequencing data were available from 2,605 primary tumours and 173 metastases or local recurrences. Mean read coverage was 39× for normal samples, whereas tumours had a bimodal coverage distribution with modes at 38× and 60× (Supplementary Fig. 1). RNA-sequencing data were available for 1,222 donors. The final cohort comprised 1,469 men (55%) and 1,189 women (45%), with a mean age of 56 years (range, 1–90 years) across 38 tumour types (Extended Data Table 1 and Supplementary Table 1).

To identify somatic mutations, we analysed all 6,835 samples using a uniform set of algorithms for alignment, variant calling and quality control (Extended Data Fig. 1, Supplementary Fig. 2 and Supplementary Methods 2). We used three established pipelines to call somatic single-nucleotide variations (SNVs), small insertions and deletions (indels), copy-number alterations (CNAs) and SVs. Somatic retrotransposition events, mitochondrial DNA mutations and telomere lengths were also called by bespoke algorithms. RNA-sequencing data were uniformly processed to call transcriptomic alterations. Germline variants identified by the three separate pipelines included single-nucleotide polymorphisms, indels, SVs and mobile-element insertions (Supplementary Table 2).

The requirement to uniformly realign and call variants on approximately 5,800 whole genomes presented considerable computational challenges, and raised ethical issues owing to the use of data from different jurisdictions (Extended Data Table 2). We used cloud computing[26,27] to distribute alignment and variant calling across 13 data centres on 3 continents (Supplementary Table 3). Core pipelines were packaged into Docker containers[28] as reproducible, stand-alone packages, which we have made available for download. Data repositories for raw and derived datasets, together with portals for data visualization and exploration, have also been created (Box 1 and Supplementary Table 4).

Benchmarking of genetic variant calls

To benchmark mutation calling, we ran the 3 core pipelines, together with 10 additional pipelines, on 63 representative tumour–normal genome pairs (Supplementary Note 1). For 50 of these cases, we performed validation by hybridization of tumour and matched normal DNA to a custom bait set with deep sequencing[29]. The 3 core somatic variant-calling pipelines had individual estimates of sensitivity of 80–90% to detect a true somatic SNV called by any of the 13 pipelines; more

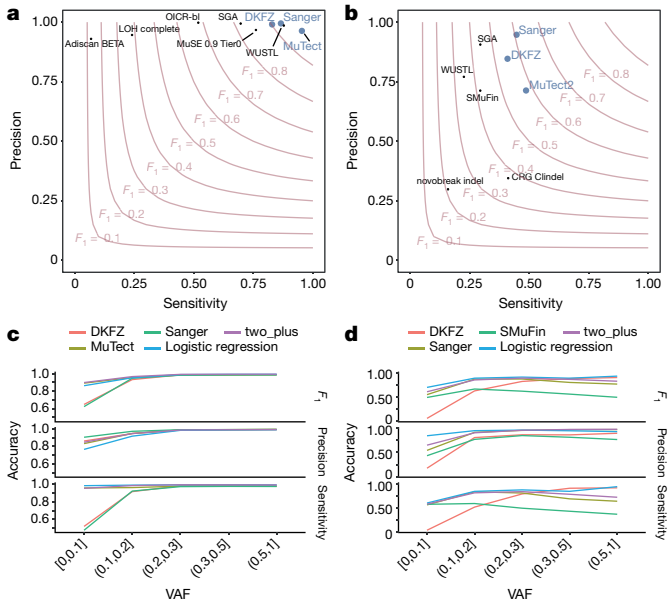


Fig. 1 | Validation of variant-calling pipelines in PCAWG. a, Scatter plot of estimated sensitivity and precision for somatic SNVs across individual algorithms assessed in the validation exercise across $n = 63$ PCAWG samples. Core algorithms included in the final PCAWG call set are shown in blue. **b**, Sensitivity and precision estimates across individual algorithms for somatic indels. **c**, Accuracy (precision, sensitivity and F_1 score, defined as 2 × sensitivity × precision/(sensitivity + precision)) of somatic SNV calls across variant allele fractions (VAFs) for the core algorithms. The accuracy of two methods of combining variant calls (two-plus, which was used in the final dataset, and logistic regression) is also shown. **d**, Accuracy of indel calls across variant allele fractions.

than 95% of SNV calls made by each of the core pipelines were genuine somatic variants (Fig. 1a). For indels—a more-challenging class of variants to identify with short-read sequencing—the 3 core algorithms had individual sensitivity estimates in the range of 40–50%, with precision of 70–95% (Fig. 1b). For individual SV algorithms, we estimated precision to be in the range 80–95% for samples in the 63-sample pilot dataset.

Next, we defined a strategy to merge results from the three pipelines into one final call-set to be used for downstream scientific analyses (Methods and Supplementary Note 2). Sensitivity and precision of consensus somatic variant calls were 95% (90% confidence interval, 88–98%) and 95% (90% confidence interval, 71–99%), respectively, for SNVs (Extended Data Fig. 2). For somatic indels, sensitivity and precision were 60% (34–72%) and 91% (73–96%), respectively (Extended Data Fig. 2). Regarding somatic SVs, we estimate the sensitivity of merged calls to be 90% for true calls generated by any one pipeline; precision was estimated as 97.5%. The improvement in calling accuracy from combining different pipelines was most noticeable in variants with low variant allele fractions, which probably originate from tumour subclones (Fig. 1c, d). Germline variant calls, phased using a haplotype-reference panel, displayed a precision of more than 99% and a sensitivity of 92–98% (Supplementary Note 2).

Analysis of PCAWG data

The uniformly generated, high-quality set of variant calls across more than 2,500 donors provided the springboard for a series of scientific working groups to explore the biology of cancer. A comprehensive suite of companion papers that describe the analyses and discoveries across these thematic areas is copublished with this paper[4–18] (Extended Data Table 3).

Pan-cancer burden of somatic mutations

Across the 2,583 white-listed PCAWG donors, we called 43,778,859 somatic SNVs, 410,123 somatic multinucleotide variants, 2,418,247 somatic indels, 288,416 somatic SVs, 19,166 somatic retrotransposition events and 8,185 de novo mitochondrial DNA mutations (Supplementary Table 1). There was considerable heterogeneity in the burden of somatic mutations across patients and tumour types, with a broad correlation in mutation burden among different classes of somatic variation (Extended Data Fig. 3). Analysed at a per-patient level, this correlation held, even when considering tumours with similar purity and ploidy (Supplementary Fig. 3). Why such correlation should apply on a pan-cancer basis is unclear. It is likely that age has some role, as we observe a correlation between most classes of somatic mutation and age at diagnosis (around 190 SNVs per year, $P = 0.02$; about 22 indels per year, $P = 5 \times 10^{-5}$; 1.5 SVs per year, $P < 2 \times 10^{-16}$; linear regression with likelihood ratio tests; Supplementary Fig. 4). Other factors are also likely to contribute to the correlations among classes of somatic mutation, as there is evidence that some DNA-repair defects can cause multiple types of somatic mutation[30], and a single carcinogen can cause a range of DNA lesions[31].

Panorama of driver mutations in cancer

We extracted the subset of somatic mutations in PCAWG tumours that have high confidence to be driver events on the basis of current knowledge. One challenge to pinpointing the specific driver mutations in an individual tumour is that not all point mutations in recurrently mutated cancer-associated genes are drivers[32]. For genomic elements significantly mutated in PCAWG data, we developed a 'rank-and-cut' approach to identify the probable drivers (Supplementary Methods 8.1). This approach works by ranking the observed mutations in a given genomic element based on recurrence, estimated functional consequence and expected pattern of drivers in that element. We then estimate the excess burden of somatic mutations in that genomic element above that expected for the background mutation rate, and cut the ranked mutations at this level. Mutations in each element with the highest driver ranking were then assigned as probable drivers; those below the threshold will probably have arisen through chance and were assigned as probable passengers. Improvements to features that are used to rank the mutations and the methods used to measure them will contribute to further development of the rank-and-cut approach.

We also needed to account for the fact that some bona fide cancer genomic elements were not rediscovered in PCAWG data because of low statistical power. We therefore added previously known cancer-associated genes to the discovery set, creating a 'compendium of mutational driver elements' (Supplementary Methods 8.2). Then, using stringent rules to nominate driver point mutations that affect these genomic elements on the basis of prior knowledge[33], we separated probable driver from passenger point mutations. To cover all classes of variant, we also created a compendium of known driver SVs, using analogous rules to identify which somatic CNAs and SVs are most likely to act as drivers in each tumour. For probable pathogenic germline variants, we identified all truncating germline point mutations and SVs that affect high-penetrance germline cancer-associated genes.

This analysis defined a set of mutations that we could confidently assert, based on current knowledge, drove tumorigenesis in the more than 2,500 tumours of PCAWG. We found that 91% of tumours had at least one identified driver mutation, with an average of 4.6 drivers per tumour identified, showing extensive variation across cancer types (Fig. 2a). For coding point mutations, the average was 2.6 drivers per tumour, similar to numbers estimated in known cancer-associated genes in tumours in the TCGA using analogous approaches[32].

To address the frequency of non-coding driver point mutations, we combined promoters and enhancers that are known targets of

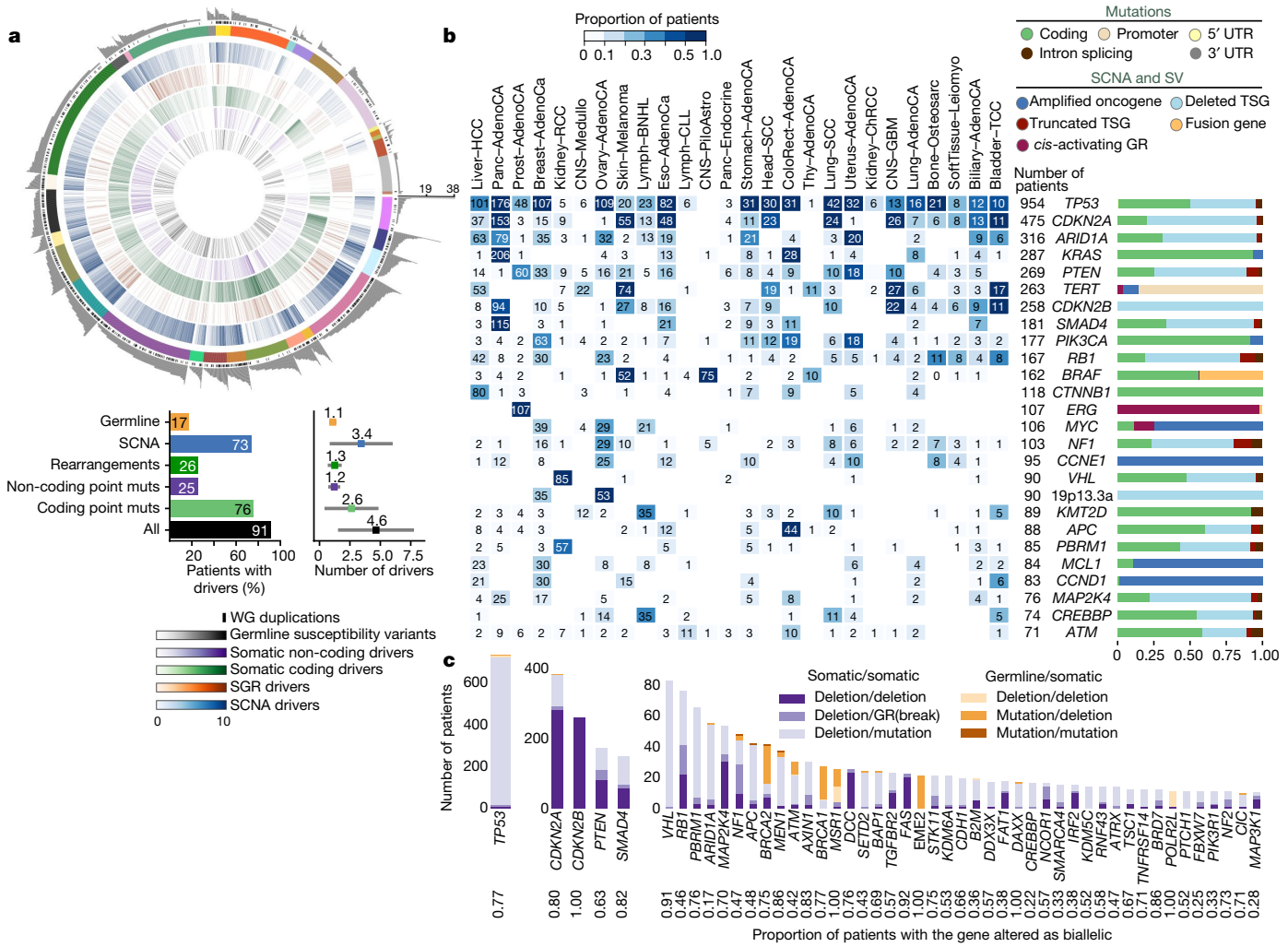


Fig. 2 | Panorama of driver mutations in PCAWG. a, Top, putative driver mutations in PCAWG, represented as a circos plot. Each sector represents a tumour in the cohort. From the periphery to the centre of the plot the concentric rings represent: (1) the total number of driver alterations; (2) the presence of whole-genome (WG) duplication; (3) the tumour type; (4) the number of driver CNAs; (5) the number of driver genomic rearrangements; (6) driver coding point mutations; (7) driver non-coding point mutations; and (8) pathogenic germline variants. Bottom, snapshots of the panorama of driver mutations. The horizontal bar plot (left) represents the proportion of patients with different types of drivers. The dot plot (right) represents the mean number of each type of driver mutation across tumours with at least one event (the square dot) and the standard deviation (grey whiskers), based on $n = 2{,}583$ patients. **b**, Genomic elements targeted by different types of mutations in the cohort altered in more than 65 tumours. Both germline and somatic variants are included. Left, the heat map shows the recurrence of alterations across cancer types. The colour indicates the proportion of mutated tumours and the number indicates the absolute count of mutated tumours. Right, the proportion of each type of alteration that affects each genomic element. **c**, Tumour-suppressor genes with biallelic inactivation in 10 or more patients. The values included under the gene labels represent the proportions of patients who have biallelic mutations in the gene out of all patients with a somatic mutation in that gene. GR, genomic rearrangement; SCNA, somatic copy-number alteration; SGR, somatic genome rearrangement; TSG, tumour suppressor gene; UTR, untranslated region.

non-coding drivers[34–37] with those newly discovered in PCAWG data; this is reported in a companion paper[4]. Using this approach, only 13% (785 out of 5,913) of driver point mutations were non-coding in PCAWG. Nonetheless, 25% of PCAWG tumours bear at least one putative non-coding driver point mutation, and one third (237 out of 785) affected the *TERT* promoter (9% of PCAWG tumours). Overall, non-coding driver point mutations are less frequent than coding driver mutations. With the exception of the *TERT* promoter, individual enhancers and promoters are only infrequent targets of driver mutations[4].

Across tumour types, SVs and point mutations have different relative contributions to tumorigenesis. Driver SVs are more prevalent in breast adenocarcinomas (6.4 ± 3.7 SVs (mean ± s.d.) compared with 2.2 ± 1.3 point mutations; $P < 1 \times 10^{-16}$, Mann–Whitney *U*-test) and ovary adenocarcinomas (5.8 ± 2.6 SVs compared with 1.9 ± 1.0 point mutations; $P < 1 \times 10^{-16}$), whereas driver point mutations have a larger contribution in colorectal adenocarcinomas (2.4 ± 1.4 SVs compared with 7.4 ± 7.0 point mutations; $P = 4 \times 10^{-10}$) and mature B cell lymphomas (2.2 ± 1.3 SVs compared with 6 ± 3.8 point mutations; $P < 1 \times 10^{-16}$), as previously shown[38]. Across tumour types, there are differences in which classes of mutation affect a given genomic element (Fig. 2b).

We confirmed that many driver mutations that affect tumour-suppressor genes are two-hit inactivation events (Fig. 2c). For example, of the 954 tumours in the cohort with driver mutations in *TP53*, 736 (77%) had both alleles mutated, 96% of which (707 out of 736) combined a somatic point mutation that affected one allele with somatic deletion of the other allele. Overall, 17% of patients had rare germline protein-truncating variants (PTVs) in cancer-predisposition genes[39], DNA-damage response genes[40] and somatic driver genes. Biallelic inactivation due to somatic alteration on top of a germline PTV was observed in 4.5% of patients overall, with 81% of

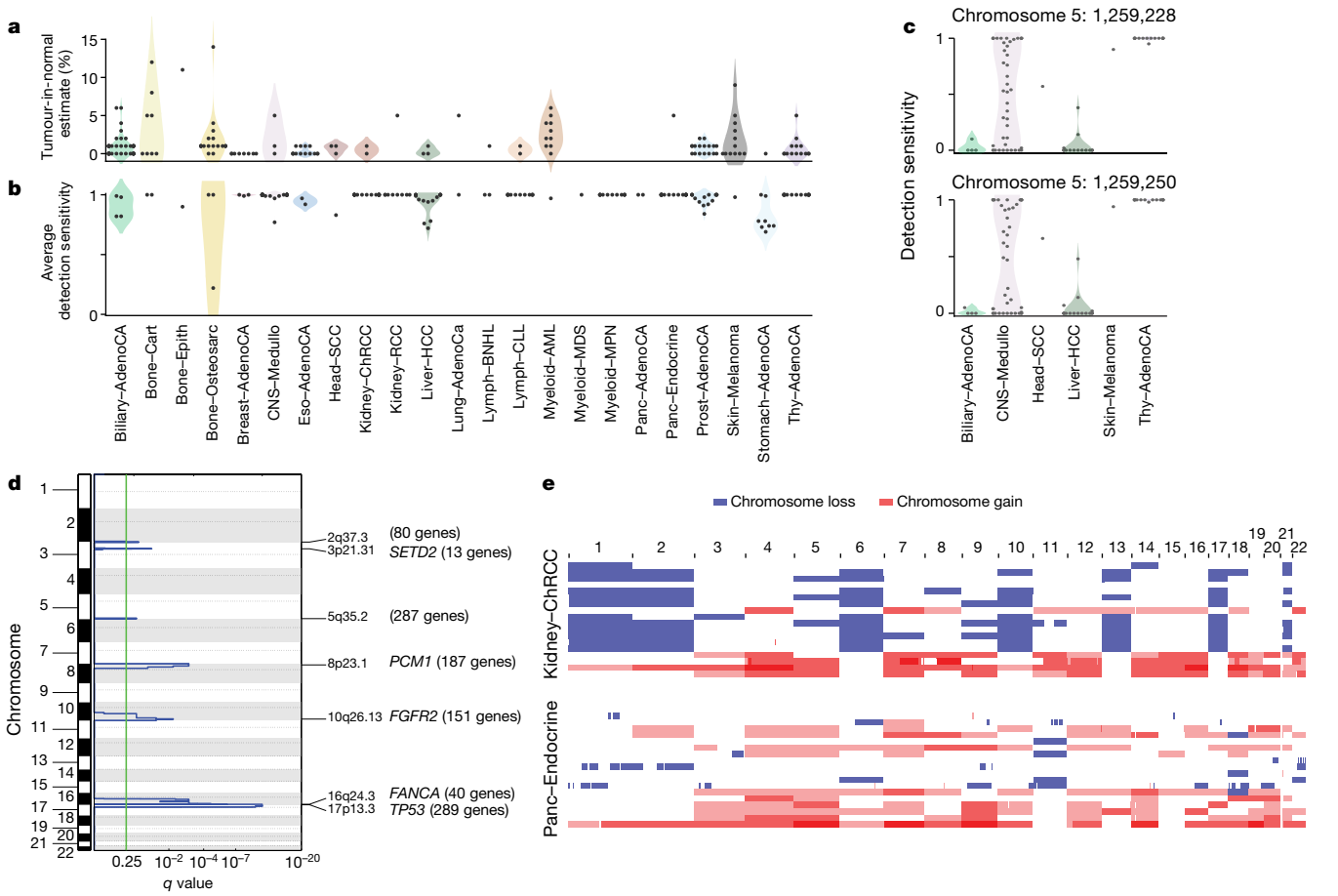


Fig. 3 | Analysis of patients with no detected driver mutations. a, Individual estimates of the percentage of tumour-in-normal contamination across patients with no driver mutations in PCAWG ($n = 181$). No data were available for myelodysplastic syndromes and acute myeloid leukaemia. Points represent estimates for individual patients, and the coloured areas are estimated density distributions (violin plots). Abbreviations of the tumour types are defined in Extended Data Table 1. **b**, Average detection sensitivity by tumour type for tumours without known drivers ($n = 181$). Each dot represents a given sample and is the average sensitivity of detecting clonal substitutions across the genome, taking into account purity and ploidy. Coloured areas are estimated density distributions, shown for cohorts with at least five cases. **c**, Detection sensitivity for *TERT* promoter hotspots in tumour types in which *TERT* is frequently mutated. Coloured areas are estimated density distributions. **d**, Significant copy-number losses identified by two-sided hypothesis testing using GISTIC2.0, corrected for multiple-hypothesis testing. Numbers in parentheses indicate the number of genes in significant regions when analysing medulloblastomas without known drivers ($n = 42$). Significant regions with known cancer-associated genes are labelled with the representative cancer-associated gene. **e**, Aneuploidy in chromophobe renal cell carcinomas and pancreatic neuroendocrine tumours without known drivers. Patients are ordered on the *y* axis by tumour type and then by presence of whole-genome duplication (bottom) or not (top).

these affecting known cancer-predisposition genes (such as *BRCA1*, *BRCA2* and *ATM*).

PCAWG tumours with no apparent drivers

Although more than 90% of PCAWG cases had identified drivers, we found none in 181 tumours (Extended Data Fig. 4a). Reasons for missing drivers have not yet been systematically evaluated in a pan-cancer cohort, and could arise from either technical or biological causes.

Technical explanations could include poor-quality samples, inadequate sequencing or failures in the bioinformatic algorithms used. We assessed the quality of the samples and found that 4 of the 181 cases with no known drivers had more than 5% tumour DNA contamination in their matched normal sample (Fig. 3a). Using an algorithm designed to correct for this contamination[41], we identified previously missed mutations in genes relevant to the respective cancer types. Similarly, if the fraction of tumour cells in the cancer sample is low through stromal contamination, the detection of driver mutations can be impaired. Most tumours with no known drivers had an average power to detect mutations close to 100%; however, a few had power in the 70–90% range (Fig. 3b and Extended Data Fig. 4b). Even in adequately sequenced genomes, lack of read depth at specific driver loci can impair mutation detection. For example, only around 50% of PCAWG tumours had sufficient coverage to call a mutation (≥90% power) at the two *TERT* promoter hotspots, probably because the high GC content of this region causes biased coverage (Fig. 3c). In fact, 6 hepatocellular carcinomas and 2 biliary cholangiocarcinomas among the 181 cases with no known drivers actually did contain *TERT* mutations, which were discovered after deep targeted sequencing[42].

Finally, technical reasons for missing driver mutations include failures in the bioinformatic algorithms. This affected 35 myeloproliferative neoplasms in PCAWG, in which the $JAK2^{V617F}$ driver mutation should have been called. Our somatic variant-calling algorithms rely on 'panels of normals', typically from blood samples, to remove recurrent sequencing artefacts. As 2–5% of healthy individuals carry occult haematopoietic clones[43], recurrent driver mutations in these clones can enter panels of normals.

With regard to biological causes, tumours may be driven by mutations in cancer-associated genes that are not yet described for that tumour type. Using driver discovery algorithms on tumours with no known drivers, no individual genes reached significance for point mutations. However, we identified a recurrent CNA that spanned *SETD2* in

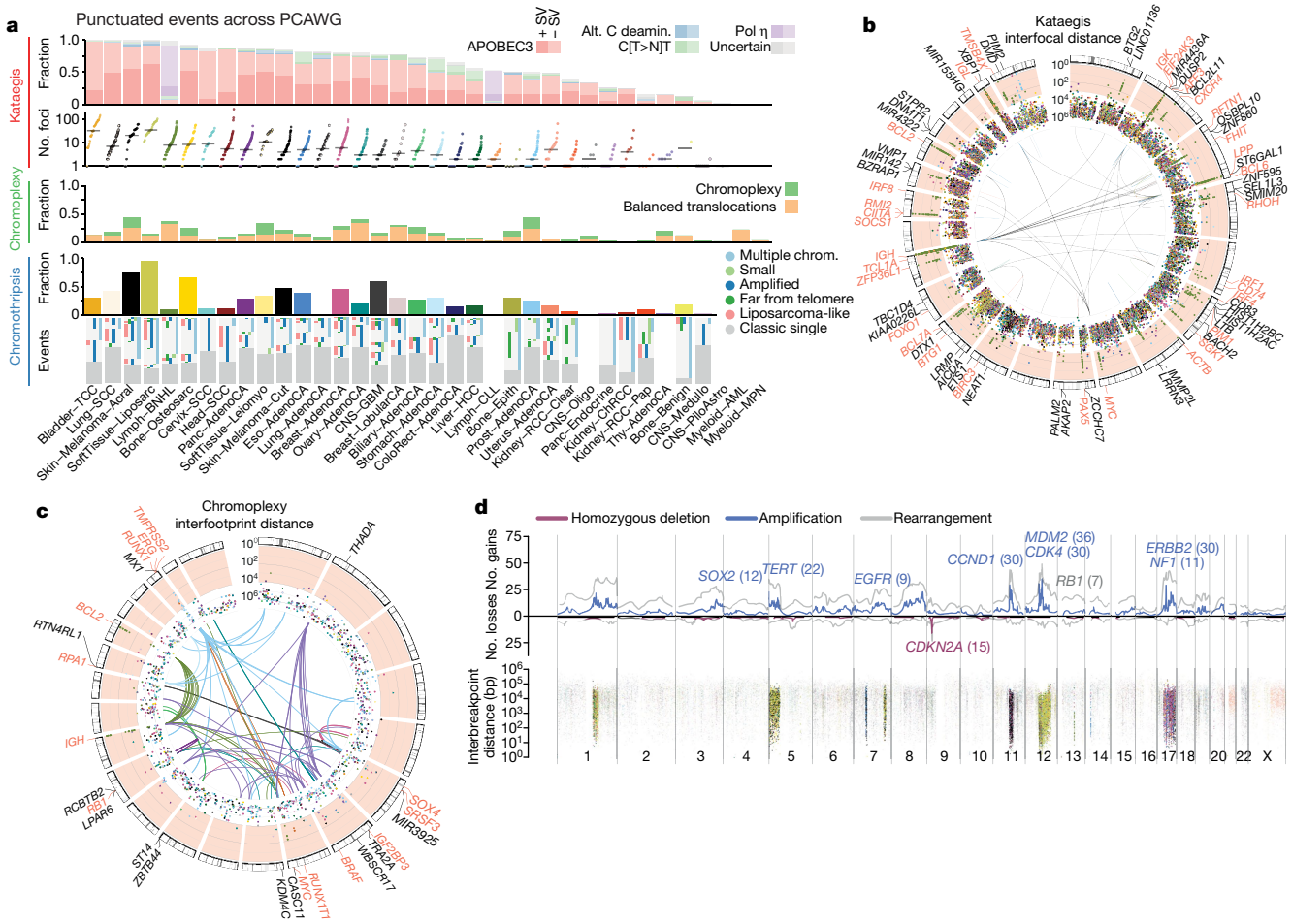


Fig. 4 | Patterns of clustered mutational processes in PCAWG. a, Kataegis. Top, prevalence of different types of kataegis and their association with SVs (≤1 kb from the focus). Bottom, the distribution of the number of foci of kataegis per sample. Chromoplexy. Prevalence of chromoplexy across cancer types, subdivided into balanced translocations and more complex events. Chromothripsis. Top, frequency of chromothripsis across cancer types. Bottom, for each cancer type a column is shown, in which each row is a chromothripsis region represented by five coloured rectangles relating to its categorization. **b**, Circos rainfall plot showing the distances between consecutive kataegis events across PCAWG compared with their genomic position. Lymphoid tumours (khaki, B cell non-Hodgkin's lymphoma; orange, chronic lymphocytic leukaemia) have hypermutation hot spots (≥3 foci with distance ≤1 kb; pale red zone), many of which are near known cancer-associated genes (red annotations) and have associated SVs (≤10 kb from the focus; shown as arcs in the centre). **c**, Circos rainfall plot as in **b** that shows the distance versus the position of consecutive chromoplexy and reciprocal translocation footprints across PCAWG. Lymphoid, prostate and thyroid cancers exhibit recurrent events (≥2 footprints with distance ≤10 kb; pale red zone) that are likely to be driver SVs and are annotated with nearby genes and associated SVs, which are shown as bold and thin arcs for chromoplexy and reciprocal translocations, respectively (colours as in **a**). **d**, Effect of chromothripsis along the genome and involvement of PCAWG driver genes. Top, number of chromothripsis-induced gains or losses (grey) and amplifications (blue) or deletions (red). Within the identified chromothripsis regions, selected recurrently rearranged (light grey), amplified (blue) and homozygously deleted (magenta) driver genes are indicated. Bottom, interbreakpoint distance between all subsequent breakpoints within chromothripsis regions across cancer types, coloured by cancer type. Regions with an average interbreakpoint distance <10 kb are highlighted. C[T>N]T, kataegis with a pattern of thymine mutations in a Cp **T**pT context.

medulloblastomas that lacked known drivers (Fig. 3d), indicating that restricting hypothesis testing to missing-driver cases can improve power if undiscovered genes are enriched in such tumours. Inactivation of *SETD2* in medulloblastoma significantly decreased gene expression ($P = 0.002$) (Extended Data Fig. 4c). Notably, *SETD2* mutations occurred exclusively in medulloblastoma group-4 tumours ($P < 1 \times 10^{-4}$). Group-4 medulloblastomas are known for frequent mutations in other chromatin-modifying genes[44], and our results suggest that *SETD2* loss of function is an additional driver that affects chromatin regulators in this subgroup.

Two tumour types had a surprisingly high fraction of patients without identified driver mutations: chromophobe renal cell carcinoma (44%; 19 out of 43) and pancreatic neuroendocrine cancers (22%; 18 out of 81) (Extended Data Fig. 4a). A notable feature of the missing-driver cases in both tumour types was a remarkably consistent profile of chromosomal aneuploidy—patterns that have previously been reported[45,46] (Fig. 3e). The absence of other identified driver mutations in these patients raises the possibility that certain combinations of whole-chromosome gains and losses may be sufficient to initiate a cancer in the absence of more-targeted driver events such as point mutations or fusion genes of focal CNAs.

Even after accounting for technical issues and novel drivers, 5.3% of PCAWG tumours still had no identifiable driver events. In a research setting, in which we are interested in drawing conclusions about populations of patients, the consequences of technical issues that affect occasional samples will be mitigated by sample size. In a clinical setting, in which we are interested in the driver mutations in a specific patient, these issues become substantially more important. Careful and critical appraisal of the whole pipeline—including sample acquisition, genome sequencing, mapping, variant calling and driver annotation, as done

here—should be required for laboratories that offer clinical sequencing of cancer genomes.

Patterns of clustered mutations and SVs

Some somatic mutational processes generate multiple mutations in a single catastrophic event, typically clustered in genomic space, leading to substantial reconfiguration of the genome. Three such processes have previously been described: (1) chromoplexy, in which repair of co-occurring double-stranded DNA breaks—typically on different chromosomes—results in shuffled chains of rearrangements[47,48] (Extended Data Fig. 5a); (2) kataegis, a focal hypermutation process that leads to locally clustered nucleotide substitutions, biased towards a single DNA strand[49–51] (Extended Data Fig. 5b); and (3) chromothripsis, in which tens to hundreds of DNA breaks occur simultaneously, clustered on one or a few chromosomes, with near-random stitching together of the resulting fragments[52–55] (Extended Data Fig. 5c). We characterized the PCAWG genomes for these three processes (Fig. 4).

Chromoplexy events and reciprocal translocations were identified in 467 (17.8%) samples (Fig. 4a, c). Chromoplexy was prominent in prostate adenocarcinoma and lymphoid malignancies, as previously described[47,48], and—unexpectedly—thyroid adenocarcinoma. Different genomic loci were recurrently rearranged by chromoplexy across the three tumour types, mediated by positive selection for particular fusion genes or enhancer-hijacking events. Of 13 fusion genes or enhancer hijacking events in 48 thyroid adenocarcinomas, at least 4 (31%) were caused by chromoplexy, with a further 4 (31%) part of complexes that contained chromoplexy footprints (Extended Data Fig. 5a). These events generated fusion genes that involved *RET* (two cases) and *NTRK3* (one case)[56], and the juxtaposition of the oncogene *IGF2BP3* with regulatory elements from highly expressed genes (five cases).

Kataegis events were found in 60.5% of all cancers, with particularly high abundance in lung squamous cell carcinoma, bladder cancer, acral melanoma and sarcomas (Fig. 4a, b). Typically, kataegis comprises C > N mutations in a TpC context, which are probably caused by APOBEC activity[49–51], although a T > N conversion in a **T**pT or Cp**T** process (the affected T is highlighted in bold) attributed to error-prone polymerases has recently been described[57]. The APOBEC signature accounted for 81.7% of kataegis events and correlated positively with *APOBEC3B* expression levels, somatic SV burden and age at diagnosis (Supplementary Fig. 5). Furthermore, 5.7% of kataegis events involved the T > N error-prone polymerase signature and 2.3% of events, most notably in sarcomas, showed cytidine deamination in an alternative Gp**C** or Cp**C** context.

Kataegis events were frequently associated with somatic SV breakpoints (Fig. 4a and Supplementary Fig. 6a), as previously described[50,51]. Deletions and complex rearrangements were most-strongly associated with kataegis, whereas tandem duplications and other simple SV classes were only infrequently associated (Supplementary Fig. 6b). Kataegis inducing predominantly T > N mutations in Cp**T**pT context was enriched near deletions, specifically those in the 10–25-kilobase (kb) range (Supplementary Fig. 6c).

Samples with extreme kataegis burden (more than 30 foci) comprise four types of focal hypermutation (Extended Data Fig. 6): (1) off-target somatic hypermutation and foci of T > N at Cp**T**pT, found in B cell non-Hodgkin lymphoma and oesophageal adenocarcinomas, respectively; (2) APOBEC kataegis associated with complex rearrangements, notably found in sarcoma and melanoma; (3) rearrangement-independent APOBEC kataegis on the lagging strand and in early-replicating regions, mainly found in bladder and head and neck cancer; and (4) a mix of the last two types. Kataegis only occasionally led to driver mutations (Supplementary Table 5).

We identified chromothripsis in 587 samples (22.3%), most frequently among sarcoma, glioblastoma, lung squamous cell carcinoma, melanoma and breast adenocarcinoma[18]. Chromothripsis increased with whole-genome duplications in most cancer types (Extended Data Fig. 7a), as previously shown in medulloblastoma[58]. The most recurrently associated driver was *TP53*[52] (pan-cancer odds ratio = 3.22; pan-cancer $P = 8.3 \times 10^{-35}$; $q < 0.05$ in breast lobular (odds ratio = 13), colorectal (odds ratio = 25), prostate (odds ratio = 2.6) and hepatocellular (odds ratio = 3.9) cancers; Fisher–Boschloo tests). In two cancer types (osteosarcoma and B cell lymphoma), women had a higher incidence of chromothripsis than men (Extended Data Fig. 7b). In prostate cancer, we observed a higher incidence of chromothripsis in patients with late-onset than early-onset disease[59] (Extended Data Fig. 7c).

Chromothripsis regions coincided with 3.6% of all identified drivers in PCAWG and around 7% of copy-number drivers (Fig. 4d). These proportions are considerably enriched compared to expectation if selection were not acting on these events (Extended Data Fig. 7d). The majority of coinciding driver events were amplifications (58%), followed by homozygous deletions (34%) and SVs within genes or promoter regions (8%). We frequently observed a ≥2-fold increase or decrease in expression of amplified or deleted drivers, respectively, when these loci were part of a chromothripsis event, compared with samples without chromothripsis (Extended Data Fig. 7e).

Chromothripsis manifested in diverse patterns and frequencies across tumour types, which we categorized on the basis of five characteristics (Fig. 4a). In liposarcoma, for example, chromothripsis events often involved multiple chromosomes, with universal *MDM2* amplification[60] and co-amplification of *TERT* in 4 of 19 cases (Fig. 4d). By contrast, in glioblastoma the events tended to affect a smaller region on a single chromosome that was distant from the telomere, resulting in focal amplification of *EGFR* and *MDM2* and loss of *CDKN2A*. Acral melanomas frequently exhibited *CCND1* amplification, and lung squamous cell carcinomas *SOX2* amplifications. In both cases, these drivers were more-frequently altered by chromothripsis compared with other drivers in the same cancer type and to other cancer types for the same driver (Fig. 4d and Extended Data Fig. 7f). Finally, in chromophobe renal cell carcinoma, chromothripsis nearly always affected chromosome 5 (Supplementary Fig. 7): these samples had breakpoints immediately adjacent to *TERT*, increasing *TERT* expression by 80-fold on average compared with samples without rearrangements ($P = 0.0004$; Mann–Whitney *U*-test).

Timing clustered mutations in evolution

An unanswered question for clustered mutational processes is whether they occur early or late in cancer evolution. To address this, we used molecular clocks to define broad epochs in the life history of each tumour[49,61]. One transition point is between clonal and subclonal mutations: clonal mutations occurred before, and subclonal mutations after, the emergence of the most-recent common ancestor. In regions with copy-number gains, molecular time can be further divided according to whether mutations preceded the copy-number gain (and were themselves duplicated) or occurred after the gain (and therefore present on only one chromosomal copy)[7].

Chromothripsis tended to have greater relative odds of being clonal than subclonal, suggesting that it occurs early in cancer evolution, especially in liposarcomas, prostate adenocarcinoma and squamous cell lung cancer (Fig. 5a). As previously reported, chromothripsis was especially common in melanomas[62]. We identified 89 separate chromothripsis events that affected 66 melanomas (61%); 47 out of 89 events affected genes known to be recurrently altered in melanoma[63] (Supplementary Table 6). Involvement of a region on chromosome 11 that includes the cell-cycle regulator *CCND1* occurred in 21 cases (10 out of 86 cutaneous, and 11 out of 21 acral or mucosal melanomas), typically combining chromothripsis with amplification (19 out of 21 cases) (Extended Data Fig. 8). Co-involvement of other cancer-associated genes in the same chromothripsis event was also frequent, including

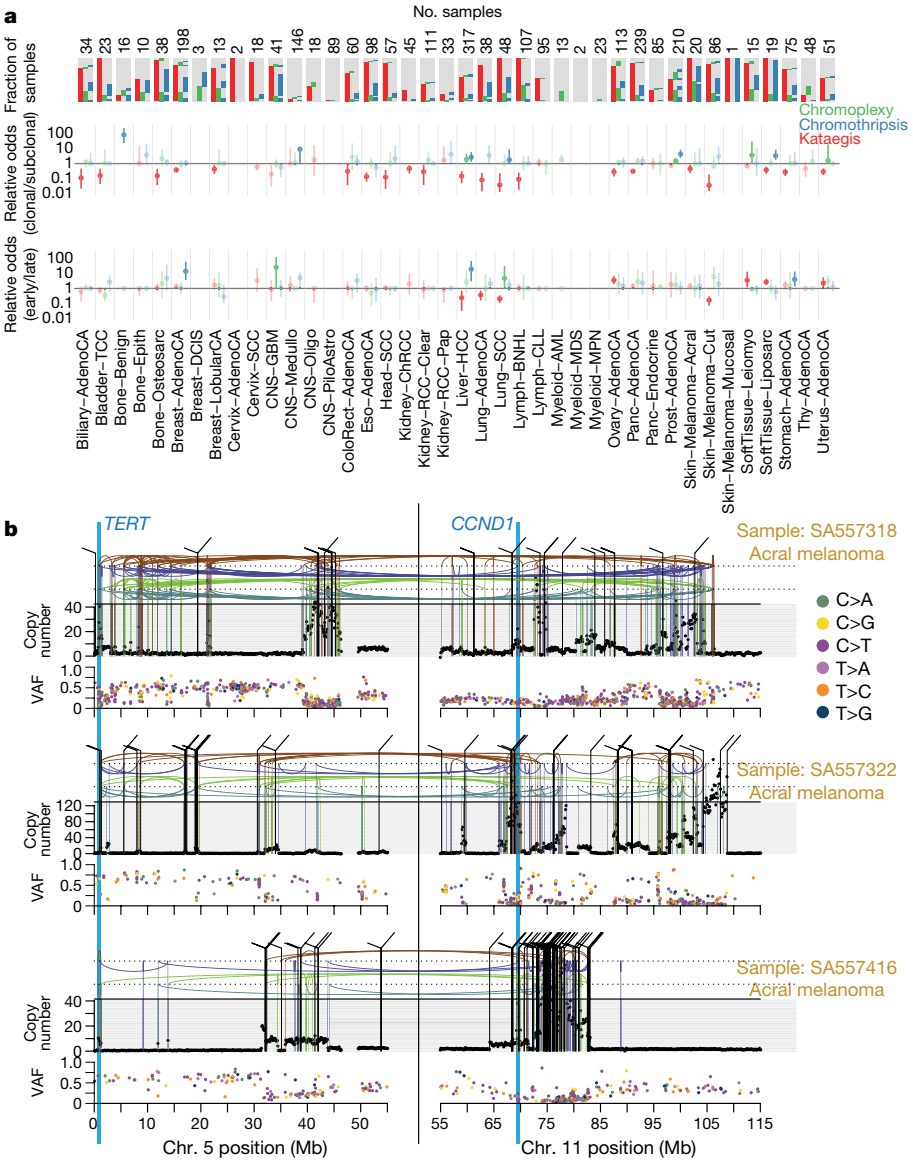


Fig. 5 | Timing of clustered events in PCAWG. a, Extent and timing of chromothripsis, kataegis and chromoplexy across PCAWG. Top, stacked bar charts illustrate co-occurrence of chromothripsis, kataegis and chromoplexy in the samples. Middle, relative odds of clustered events being clonal or subclonal are shown with bootstrapped 95% confidence intervals. Point estimates are highlighted when they do not overlap odds of 1:1. Bottom, relative odds of the events being early or late clonal are shown as above. Sample sizes (number of patients) are shown across the top. **b**, Three representative patients with acral melanoma and chromothripsis-induced amplification that simultaneously affects *TERT* and *CCND1*. The black points (top) represent sequence coverage from individual genomic bins, with SVs shown as coloured arcs (translocation in black, deletion in purple, duplication in brown, tail-to-tail inversion in cyan and head-to-head inversion in green). Bottom, the variant allele fractions of somatic point mutations.

TERT (five cases), *CDKN2A* (three cases), *TP53* (two cases) and *MYC* (two cases) (Fig. 5b). In these co-amplifications, a chromothripsis event involving multiple chromosomes initiated the process, creating a derivative chromosome in which hundreds of fragments were stitched together in a near-random order (Fig. 5b). This derivative then rearranged further, leading to massive co-amplification of the multiple target oncogenes together with regions located nearby on the derivative chromosome.

In these cases of amplified chromothripsis, we can use the inferred number of copies bearing each SNV to time the amplification process. SNVs present on the chromosome before amplification will themselves be amplified and are therefore reported in a high fraction of sequence reads (Fig. 5b and Extended Data Fig. 8). By contrast, late SNVs that occur after the amplification has concluded will be present on only one chromosome copy out of many, and thus have a low variant allele fraction. Regions of *CCND1* amplification had few—sometimes zero—mutations at high variant allele fraction in acral melanomas, in contrast to later *CCND1* amplifications in cutaneous melanomas, in which hundreds to thousands of mutations typically predated amplification (Fig. 5b and Extended Data Fig. 9a, b). Thus, both chromothripsis and the subsequent amplification generally occurred very early during the evolution of acral melanoma. By comparison, in lung squamous cell carcinomas, similar patterns of chromothripsis followed by *SOX2* amplification are characterized by many amplified SNVs, suggesting a later event in the evolution of these cancers (Extended Data Fig. 9c).

Notably, in cancer types in which the mutational load was sufficiently high, we could detect a larger-than-expected number of SNVs on an intermediate number of DNA copies, suggesting that they appeared during the amplification process (Supplementary Fig. 8).

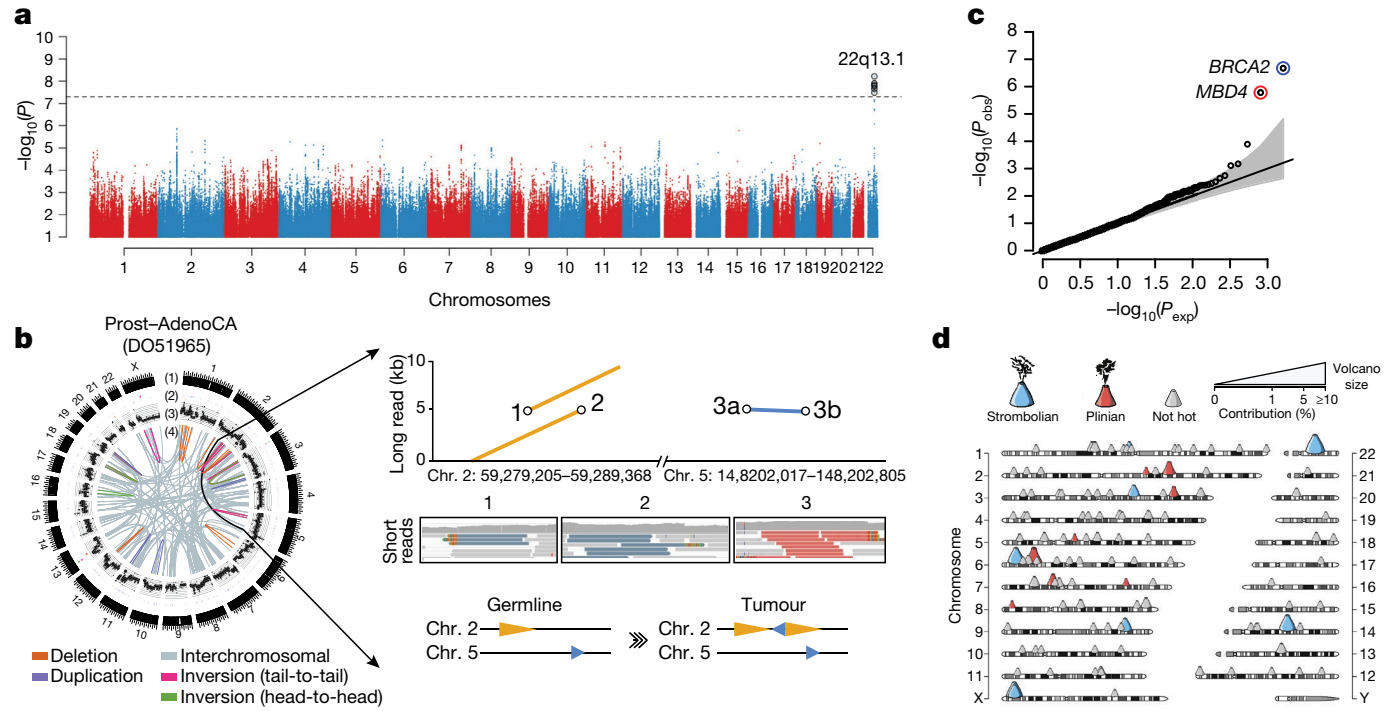


Fig. 6 | Germline determinants of the somatic mutation landscape. **a**, Association between common (MAF > 5%) germline variants and somatic APOBEC3B-like mutagenesis in individuals of European ancestry (n = 1,201). Two-sided hypothesis testing was performed with PLINK v.1.9. To mitigate multiple-hypothesis testing, the significance threshold was set to genome-wide significance ($P < 5 \times 10^{-8}$). **b**, Templated insertion SVs in a *BRCA1*-associated prostate cancer. Left, chromosome bands (1); SVs ≤ 10 megabases (Mb) (2); 1-kb read depth corrected to copy number 0–6 (3); inter- and intrachromosomal SVs > 10 Mb (4). Right, a complex somatic SV composed of a 2.2-kb tandem duplication on chromosome 2 together with a 232-base-pair (bp) inverted templated insertion SV that is derived from chromosome 5 and inserted inbetween the tandem duplication (bottom). Consensus sequence alignment of locally assembled Oxford Nanopore Technologies long sequencing reads to chromosomes 2 and 5 of the human reference genome (top). Breakpoints are circled and marked as 1 (beginning of tandem duplication), 2 (end of tandem duplication) or 3 (inverted templated insertion). For each breakpoint, the middle panel shows Illumina short reads at SV breakpoints. **c**, Association between rare germline PTVs (MAF < 0.5%) and somatic CpG mutagenesis (approximately with signature 1) in individuals of European ancestry (n = 1,201). Genes highlighted in blue or red were associated with lower or higher somatic mutation rates. Two-sided hypothesis testing was performed using linear-regression models with sex, age at diagnosis and cancer project as variables. To mitigate multiple-hypothesis testing, the significance threshold was set to exome-wide significance ($P < 2.5 \times 10^{-6}$). The black line represents the identity line that would be followed if the observed P values followed the null expectation; the shaded area shows the 95% confidence intervals. **d**, Catalogue of polymorphic germline L1 source elements that are active in cancer. The chromosomal map shows germline source L1 elements as volcano symbols. Each volcano is colour-coded according to the type of source L1 activity. The contribution of each source locus (expressed as a percentage) to the total number of transductions identified in PCAWG tumours is represented as a gradient of volcano size, with top contributing elements exhibiting larger sizes.

Germline effects on somatic mutations

We integrated the set of 88 million germline genetic variant calls with somatic mutations in PCAWG, to study germline determinants of somatic mutation rates and patterns. First, we performed a genome-wide association study of somatic mutational processes with common germline variants (minor allele frequency (MAF) > 5%) in individuals with inferred European ancestry. An independent genome-wide association study was performed in East Asian individuals from Asian cancer genome projects. We focused on two prevalent endogenous mutational processes: spontaneous deamination of 5-methylcytosine at CpG dinucleotides[5] (signature 1) and activity of the APOBEC3 family of cytidine deaminases[64] (signatures 2 and 13). No locus reached genome-wide significance ($P < 5 \times 10^{-8}$) for signature 1 (Extended Data Fig. 10a, b). However, a locus at 22q13.1 predicted an APOBEC3B-like mutagenesis at the pan-cancer level[65] (Fig. 6a). The strongest signal at 22q13.1 was driven by rs12628403, and the minor (non-reference) allele was protective against APOBEC3B-like mutagenesis ($\beta = -0.43$, $P = 5.6 \times 10^{-9}$, MAF = 8.2%, n = 1,201 donors) (Extended Data Fig. 10c). This variant tags a common, approximately 30-kb germline SV that deletes the *APOBEC3B* coding sequence and fuses the *APOBEC3B* 3′ untranslated region with the coding sequence of *APOBEC3A*. The deletion is known to increase breast cancer risk and APOBEC mutagenesis in breast cancer genomes[66,67]. Here, we found that rs12628403 reduces APOBEC3B-like mutagenesis specifically in cancer types with low levels of APOBEC mutagenesis ($\beta_{low} = -0.50$, $P_{low} = 1 \times 10^{-8}$; $\beta_{high} = 0.17$, $P_{high} = 0.2$), and increases APOBEC3A-like mutagenesis in cancer types with high levels of APOBEC mutagenesis ($\beta_{high} = 0.44$, $P_{high} = 8 \times 10^{-4}$; $\beta_{low} = -0.21$, $P_{low} = 0.02$). Moreover, we identified a second, novel locus at 22q13.1 that was associated with APOBEC3B-like mutagenesis across cancer types (rs2142833, $\beta = 0.23$, $P = 1.3 \times 10^{-8}$). We independently validated the association between both loci and APOBEC3B-like mutagenesis using East Asian individuals from Asian cancer genome projects ($\beta_{rs12628403} = 0.57$, $P_{rs12628403} = 4.2 \times 10^{-12}$; $\beta_{rs2142833} = 0.58$, $P_{rs2142833} = 8 \times 10^{-15}$) (Extended Data Fig. 10d). Notably, in a conditional analysis that accounted for rs12628403, we found that rs2142833 and rs12628403 are inherited independently in Europeans ($r^2 < 0.1$), and rs2142833 remained significantly associated with APOBEC3B-like mutagenesis in Europeans ($\beta_{EUR} = 0.17$, $P_{EUR} = 3 \times 10^{-5}$) and East Asians ($\beta_{ASN} = 0.25$, $P_{ASN} = 2 \times 10^{-3}$) (Extended Data Fig. 10e, f). Analysis of donor-matched expression data further suggests that rs2142833 is a *cis*-expression quantitative trait locus (eQTL) for *APOBEC3B* at the pan-cancer level ($\beta = 0.19$, $P = 2 \times 10^{-6}$) (Extended Data Fig. 10g, h), consistent with *cis*-eQTL studies in normal cells[68,69].

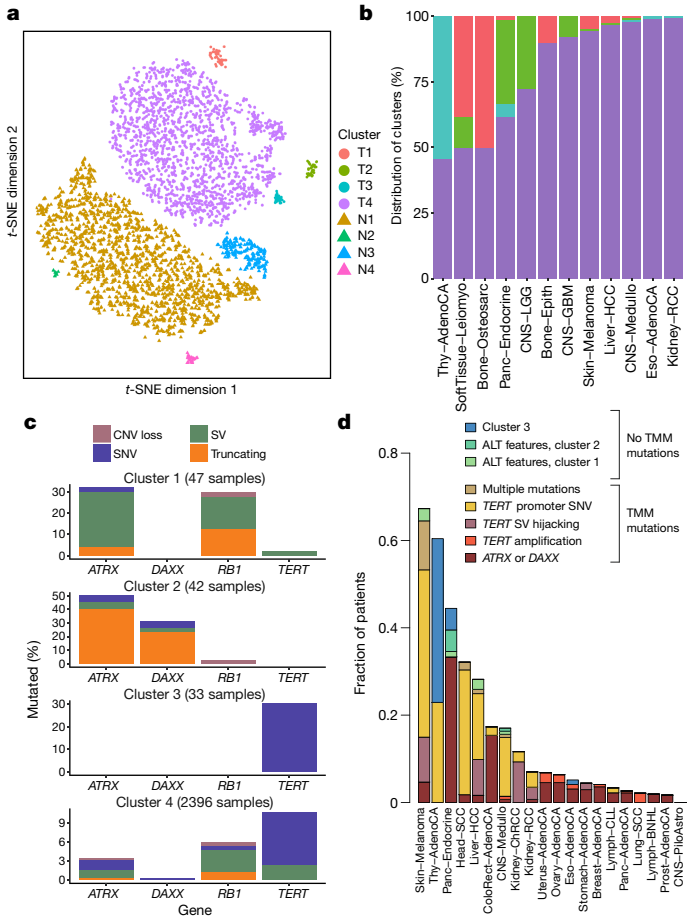


Fig. 7 | Telomere sequence patterns across PCAWG. a, Scatter plot of the clusters of telomere patterns identified across PCAWG using *t*-distributed stochastic neighbour embedding (*t*-SNE), based on $n = 2,518$ tumour samples and their matched normal samples. Axes have arbitrary dimensions such that samples with similar telomere profiles are clustered together and samples with dissimilar telomere profiles are far apart with high probability. **b**, Distribution of the four tumour-specific clusters of telomere patterns in selected tumour types from PCAWG. **c**, Distribution of relevant driver mutations associated with alternative lengthening of telomere and normal telomere maintenance across the four clusters. **d**, Distribution of telomere maintenance abnormalities across tumour types with more than 40 patients in PCAWG. Samples were classified as tumour clusters 1–3 if they fell into a relevant cluster without mutations in *TERT*, *ATRX* or *DAXX* and had no ALT phenotype. TMM, telomere maintenance mechanisms.

Second, we performed a rare-variant association study (MAF <0.5%) to investigate the relationship between germline PTVs and somatic DNA rearrangements in individuals with European ancestry (Extended Data Fig. 11a–c). Germline *BRCA2* and *BRCA1* PTVs were associated with an increased burden of small (less than 10 kb) somatic SV deletions ($P = 1 \times 10^{-8}$) and tandem duplications ($P = 6 \times 10^{-13}$), respectively, corroborating recent studies in breast and ovarian cancer[30,70]. In PCAWG data, this pattern also extends to other tumour types, including adenocarcinomas of the prostate and pancreas[6], typically in the setting of biallelic inactivation. In addition, tumours with high levels of small SV tandem duplications frequently exhibited a novel and distinct class of SVs termed 'cycles of templated insertions'[6]. These complex SV events consist of DNA templates that are copied from across the genome, joined into one contiguous sequence and inserted into a single derivative chromosome. We found a significant association between germline *BRCA1* PTVs and templated insertions at the pan-cancer level ($P = 4 \times 10^{-15}$) (Extended Data Fig. 11d, e). Whole-genome long-read sequencing data generated for a *BRCA1*-deficient PCAWG prostate tumour verified the small tandem-duplication and templated-insertion SV phenotypes (Fig. 6b). Almost all (20 out of 21) of *BRCA1*-associated tumours with a templated-insertion SV phenotype displayed combined germline and somatic hits in the gene. Together, these data suggest that biallelic inactivation of *BRCA1* is a driver of the templated-insertion SV phenotype.

Third, rare-variant association analysis revealed that patients with germline *MBD4* PTVs had increased rates of somatic C > T mutation rates at CpG dinucleotides ($P < 2.5 \times 10^{-6}$) (Fig. 6c and Extended Data Fig. 11f, g). Analysis of previously published whole-exome sequencing samples from the TCGA ($n = 8,134$) replicated the association between germline *MBD4* PTVs and increased somatic CpG mutagenesis at the pan-cancer level ($P = 7.1 \times 10^{-4}$) (Extended Data Fig. 11h). Moreover, gene-expression profiling revealed a significant but modest correlation between *MBD4* expression and somatic CpG mutation rates between and within PCAWG tumour types (Extended Data Fig. 11i–k). *MBD4* encodes a DNA-repair gene that removes thymidines from T:G mismatches within methylated CpG sites[71], a functionality that would be consistent with a CpG mutational signature in cancer.

Fourth, we assessed long interspersed nuclear elements (LINE-1; L1 hereafter) that mediate somatic retrotransposition events[72–74]. We identified 114 germline source L1 elements capable of active somatic retrotransposition, including 70 that represent insertions with respect to the human reference genome (Fig. 6d and Supplementary Table 7), and 53 that were tagged by single-nucleotide polymorphisms in strong linkage disequilibrium (Supplementary Table 7). Only 16 germline L1 elements accounted for 67% (2,440 out of 3,669) of all L1-mediated transductions[10] detected in the PCAWG dataset (Extended Data Fig. 12a). These 16 hot-L1 elements followed two broad patterns of somatic activity (8 of each), which we term Strombolian and Plinian in analogy to patterns of volcanic activity. Strombolian L1s are frequently active in cancer, but mediate only small-to-modest eruptions of somatic L1 activity in cancer samples (Extended Data Fig. 12b). By contrast, Plinian L1s are more rarely seen, but display aggressive somatic activity. Whereas Strombolian elements are typically relatively common (MAF > 2%) and sometimes even fixed in the human population, all Plinian elements were infrequent (MAF ≤ 2%) in PCAWG donors (Extended Data Fig. 12c; $P = 0.001$, Mann–Whitney *U*-test). This dichotomous pattern of activity and allele frequency may reflect differences in age and selective pressures, with Plinian elements potentially inserted into the human germline more recently. PCAWG donors bear on average between 50 and 60 L1 source elements and between 5 and 7 elements with hot activity (Extended Data Fig. 12d), but only 38% (1,075 out of 2,814) of PCAWG donors carried ≥1 Plinian element. Some L1 germline source loci caused somatic loss of tumour-suppressor genes (Extended Data Fig. 12e). Many are restricted to individual continental population ancestries (Extended Data Fig. 12f–j).

Replicative immortality

One of the hallmarks of cancer is the ability of cancer to evade cellular senescence[21]. Normal somatic cells typically have finite cell division potential; telomere attrition is one mechanism to limit numbers of mitoses[75]. Cancers enlist multiple strategies to achieve replicative immortality. Overexpression of the telomerase gene, *TERT*, which maintains telomere lengths, is especially prevalent. This can be achieved through point mutations in the promoter that lead to de novo transcription factor binding[34,37]; hitching *TERT* to highly active regulatory elements elsewhere in the genome[46,76]; insertions of viral enhancers upstream of the gene[77,78]; and increased dosage through chromosomal amplification, as we have seen in melanoma (Fig. 5b). In addition, there is an 'alternative lengthening of telomeres' (ALT) pathway, in which telomeres are lengthened through homologous recombination, mediated by loss-of-function mutations in the *ATRX* and *DAXX* genes[79].

As reported in a companion paper[13], 16% of tumours in the PCAWG dataset exhibited somatic mutations in at least one of *ATRX*, *DAXX* and *TERT*. *TERT* alterations were detected in 270 samples, whereas 128 tumours had alterations in *ATRX* or *DAXX*, of which 71 were protein-truncating. In the companion paper, which focused on describing patterns of ALT and *TERT*-mediated telomere maintenance[13], 12 features of telomeric sequence were measured in the PCAWG cohort. These included counts of nine variants of the core hexameric sequence, the number of ectopic telomere-like insertions within the genome, the number of genomic breakpoints and telomere length as a ratio between tumour and normal. Here we used the 12 features as an overview of telomere integrity across all tumours in the PCAWG dataset.

On the basis of these 12 features, tumour samples formed 4 distinct subclusters (Fig. 7a and Extended Data Fig. 13a), suggesting that telomere-maintenance mechanisms are more diverse than the well-established *TERT* and ALT dichotomy. Clusters C1 (47 tumours) and C2 (42 tumours) were enriched for traits of the ALT pathway—having longer telomeres, more genomic breakpoints, more ectopic telomere insertions and variant telomere sequence motifs (Supplementary Fig. 9). C1 and C2 were distinguished from one another by the latter having a considerable increase in the number of TTCGGG and TGAGGG variant motifs among the telomeric hexamers. Thyroid adenocarcinomas were markedly enriched among C3 samples (26 out of 33 C3 samples; $P < 10^{-16}$); the C1 cluster (ALT subtype 1) was common among sarcomas; and both pancreatic endocrine neoplasms and low-grade gliomas had a high proportion of samples in the C2 cluster (ALT subtype 2) (Fig. 7b). Notably, some of the thyroid adenocarcinomas and pancreatic neuroendocrine tumours that cluster together (cluster C3) had matched normal samples that also cluster together (normal cluster N3) (Extended Data Fig. 13a) and which share common properties. For example, the GTAGGG repeat was overrepresented among samples in this group (Supplementary Fig. 10).

Somatic driver mutations were also unevenly distributed across the four clusters (Fig. 7c). C1 tumours were enriched for *RB1* mutations or SVs ($P = 3 \times 10^{-5}$), as well as frequent SVs that affected *ATRX* ($P = 6 \times 10^{-14}$), but not *DAXX*. *RB1* and *ATRX* mutations were largely mutually exclusive (Extended Data Fig. 13b). By contrast, C2 tumours were enriched for somatic point mutations in *ATRX* and *DAXX* ($P = 6 \times 10^{-5}$), but not *RB1*. The enrichment of *RB1* mutations in C1 remained significant when only leiomyosarcomas and osteosarcomas were considered, confirming that this enrichment is not merely a consequence of the different distribution of tumour types across clusters. C3 samples had frequent *TERT* promoter mutations (30%; $P = 2 \times 10^{-6}$).

There was a marked predominance of *RB1* mutations in C1. Nearly a third of the samples in C1 contained an *RB1* alteration, which were evenly distributed across truncating SNVs, SVs and shallow deletions (Extended Data Fig. 13c). Previous research has shown that *RB1* mutations are associated with long telomeres in the absence of *TERT* mutations and *ATRX* inactivation[80], and studies using mouse models have shown that knockout of Rb-family proteins causes elongated telomeres[81]. The association with the C1 cluster here suggests that *RB1* mutations can represent another route to activating the ALT pathway, which has subtly different properties of telomeric sequence compared with the inactivation of *DAXX*—these fall almost exclusively in cluster C2.

Tumour types with the highest rates of abnormal telomere maintenance mechanisms often originate in tissues that have low endogenous replicative activity (Fig. 7d). In support of this, we found an inverse correlation between previously estimated rates of stem cell division across tissues[82] and the frequency of telomere maintenance abnormalities ($P = 0.01$, Poisson regression) (Extended Data Fig. 13d). This suggests that restriction of telomere maintenance is an important tumour-suppression mechanism, particularly in tissues with low steady-state cellular proliferation, in which a clone must overcome this constraint to achieve replicative immortality.

Conclusions and future perspectives

The resource reported in this paper and its companion papers has yielded insights into the nature and timing of the many mutational processes that shape large- and small-scale somatic variation in the cancer genome; the patterns of selection that act on these variations; the widespread effect of somatic variants on transcription; the complementary roles of the coding and non-coding genome for both germline and somatic mutations; the ubiquity of intratumoral heterogeneity; and the distinctive evolutionary trajectory of each cancer type. Many of these insights can be obtained only from an integrated analysis of all classes of somatic mutation on a whole-genome scale, and would not be accessible with, for example, targeted exome sequencing.

The promise of precision medicine is to match patients to targeted therapies using genomics. A major barrier to its evidence-based implementation is the daunting heterogeneity of cancer chronicled in these papers, from tumour type to tumour type, from patient to patient, from clone to clone and from cell to cell. Building meaningful clinical predictors from genomic data can be achieved, but will require knowledge banks comprising tens of thousands of patients with comprehensive clinical characterization[83]. As these sample sizes will be too large for any single funding agency, pharmaceutical company or health system, international collaboration and data sharing will be required. The next phase of ICGC, ICGC-ARGO (https:// www.icgc-argo.org/), will bring the cancer genomics community together with healthcare providers, pharmaceutical companies, data science and clinical trials groups to build comprehensive knowledge banks of clinical outcome and treatment data from patients with a wide variety of cancers, matched with detailed molecular profiling.

Extending the story begun by TCGA, ICGC and other cancer genomics projects, the PCAWG has brought us closer to a comprehensive narrative of the causal biological changes that drive cancer phenotypes. We must now translate this knowledge into sustainable, meaningful clinical treatments.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-020-1969-6.

1. Pleasance, E. D. et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
2. Pleasance, E. D. et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
3. Ley, T. J. et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
4. Rheinbay, E. et al. Analyses of non-coding somatic drivers in 2,693 cancer whole genomes. *Nature* https://doi.org/10.1038/s41586-020-1965-x (2020).
5. Alexandrov, L. B. et al. The repertoire of mutational signatures in human cancer. *Nature* https://doi.org/10.1038/s41586-020-1943-3 (2020).
6. Li, Y. et al. Patterns of somatic structural variation in human cancer genomes. *Nature* https://doi.org/10.1038/s41586-019-1913-9 (2020).
7. Gerstung, M. et al. The evolutionary history of 2,658 cancers. *Nature* https://doi.org/10.1038/s41586-019-1907-7 (2020).
8. PCAWG Transcriptome Core Group et al. Genomic basis of RNA alterations in cancer. *Nature* https://doi.org/10.1038/s41586-020-1970-0 (2020).
9. Zhang, Y. et al. High-coverage whole-genome analysis of 1,220 cancers reveals hundreds of genes deregulated by rearrangement-mediated *cis*-regulatory alterations. *Nat. Commun*. https://doi.org/10.1038/s41467-019-13885-w (2020).
10. Rodriguez-Martin, B. et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat. Genet*. https://doi.org/10.1038/s41588-019-0562-0 (2020).
11. Zapatka, M. et al. The landscape of viral associations in human cancers. *Nat. Genet*. https://doi.org/10.1038/s41588-019-0558-9 (2020).
12. Jiao, W. et al. A deep learning system can accurately classify primary and metastatic cancers based on patterns of passenger mutations. *Nat. Commun*. https://doi.org/10.1038/s41467-019-13825-8 (2020).

13. Sieverling, L. et al. Genomic footprints of activated telomere maintenance mechanisms in cancer. *Nat. Commun*. https://doi.org/10.1038/s41467-019-13824-9 (2020).
14. Yuan, Y. et al. Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nat. Genet*. https://doi.org/10.1038/s41588-019-0557-x (2020).
15. Akdemir, K. C. et al. Chromatin folding domains disruptions by somatic genomic rearrangements in human cancers. *Nat. Genet*. https://doi.org/10.1038/s41588-019-0564-y (2020).
16. Reyna, M. A. et al. Pathway and network analysis of more than 2,500 whole cancer genomes. *Nat. Commun*. https://doi.org/10.1038/s41467-020-14351-8 (2020).
17. Bailey, M. H. et al. Retrospective evaluation of whole exome and genome mutation calls in 746 cancer samples. *Nat. Commun*. (2020).
18. Cortes-Ciriano, I. et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat. Genet*. https://doi.org/10.1038/s41588-019-0576-7 (2020).
19. Bray, F., Ren, J.-S., Masuyer, E. & Ferlay, J. Global estimates of cancer prevalence for 27 sites in the adult population in 2008. *Int. J. Cancer* **132**, 1133–1145 (2013).
20. Tarver, T. Cancer Facts & Figures 2012. American Cancer Society (ACS). *J. Consum. Health Internet* **16**, 366–367 (2012).
21. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
22. International Cancer Genome Consortium. International network of cancer genome projects. *Nature* **464**, 993–998 (2010).
23. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385 (2018).
24. Sanchez-Vega, F. et al. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell* **173**, 321–337 (2018).
25. Hoadley, K. A. et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* **173**, 291–304 (2018).
26. Stein, L. D., Knoppers, B. M., Campbell, P., Getz, G. & Korbel, J. O. Data analysis: create a cloud commons. *Nature* **523**, 149–151 (2015).
27. Phillips, M. et al. Genomics: data sharing needs international code of conduct. *Nature* https://doi.org/10.1038/d41586-020-00082-9 (2020).
28. Krochmalski, J. *Developing with Docker* (Packt Publishing, 2016).
29. Welch, J. S. et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278 (2012).
30. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
31. Meier, B. et al. *C. elegans* whole-genome sequencing reveals mutational signatures related to carcinogens and DNA repair deficiency. *Genome Res.* **24**, 1624–1636 (2014).
32. Martincorena, I. et al. Universal patterns of selection in cancer and somatic tissues. *Cell* **171**, 1029–1041 (2017).
33. Tamborero, D. et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med.* **10**, 25 (2018).
34. Huang, F. W. et al. Highly recurrent *TERT* promoter mutations in human melanoma. *Science* **339**, 957–959 (2013).
35. Rheinbay, E. et al. Recurrent and functional regulatory mutations in breast cancer. *Nature* **547**, 55–60 (2017).
36. Fredriksson, N. J., Ny, L., Nilsson, J. A. & Larsson, E. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nat. Genet.* **46**, 1258–1263 (2014).
37. Horn, S. et al. *TERT* promoter mutations in familial and sporadic melanoma. *Science* **339**, 959–961 (2013).
38. Ciriello, G. et al. Emerging landscape of oncogenic signatures across human cancers. *Nat. Genet.* **45**, 1127–1133 (2013).
39. Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* **505**, 302–308 (2014).
40. Pearl, L. H., Schierz, A. C., Ward, S. E., Al-Lazikani, B. & Pearl, F. M. G. Therapeutic opportunities within the DNA damage response. *Nat. Rev. Cancer* **15**, 166–180 (2015).
41. Taylor-Weiner, A. et al. DeTiN: overcoming tumor-in-normal contamination. *Nat. Methods* **15**, 531–534 (2018).
42. Fujimoto, A. et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* **48**, 500–509 (2016).
43. Shlush, L. I. Age-related clonal hematopoiesis. *Blood* **131**, 496–504 (2018).
44. Northcott, P. A. et al. The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).
45. Scarpa, A. et al. Whole-genome landscape of pancreatic neuroendocrine tumours. *Nature* **543**, 65–71 (2017).
46. Davis, C. F. et al. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* **26**, 319–330 (2014).
47. Berger, M. F. et al. The genomic complexity of primary human prostate cancer. *Nature* **470**, 214–220 (2011).
48. Baca, S. C. et al. Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677 (2013).
49. Nik-Zainal, S. et al. The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
50. Nik-Zainal, S. et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
51. Roberts, S. A. et al. Clustered mutations in yeast and in human cancers can arise from damaged long single-strand DNA regions. *Mol. Cell* **46**, 424–435 (2012).
52. Rausch, T. et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with *TP53* mutations. *Cell* **148**, 59–71 (2012).
53. Stephens, P. J. et al. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
54. Korbel, J. O. & Campbell, P. J. Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**, 1226–1236 (2013).
55. Zhang, C.-Z. et al. Chromothripsis from DNA damage in micronuclei. *Nature* **522**, 179–184 (2015).
56. The Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159**, 676–690 (2014).
57. Supek, F. & Lehner, B. Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. *Cell* **170**, 534–547 (2017).
58. Mardin, B. R. et al. A cell-based model system links chromothripsis with hyperploidy. *Mol. Syst. Biol.* **11**, 828 (2015).
59. Weischenfeldt, J. et al. Integrative genomic analyses reveal an androgen-driven somatic alteration landscape in early-onset prostate cancer. *Cancer Cell* **23**, 159–170 (2013).
60. Garsed, D. W. et al. The architecture and evolution of cancer neochromosomes. *Cancer Cell* **26**, 653–667 (2014).
61. Durinck, S. et al. Temporal dissection of tumorigenesis in primary cancers. *Cancer Discov.* **1**, 137–143 (2011).
62. Hayward, N. K. et al. Whole-genome landscapes of major melanoma subtypes. *Nature* **545**, 175–180 (2017).
63. The Cancer Genome Atlas Network. Genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696 (2015).
64. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
65. Chan, K. et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat. Genet.* **47**, 1067–1072 (2015).
66. Nik-Zainal, S. et al. Association of a germline copy number polymorphism of *APOBEC3A* and *APOBEC3B* with burden of putative APOBEC-dependent mutations in breast cancer. *Nat. Genet.* **46**, 487–491 (2014).
67. Middlebrooks, C. D. et al. Association of germline variants in the *APOBEC3* region with cancer risk and enrichment with APOBEC-signature mutations in tumors. *Nat. Genet.* **48**, 1330–1338 (2016).
68. Westra, H.-J. et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).
69. Stranger, B. E. et al. Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
70. Menghi, F. et al. The tandem duplicator phenotype as a distinct genomic configuration in cancer. *Proc. Natl Acad. Sci. USA* **113**, E2373–E2382 (2016).
71. Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J. & Bird, A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* **401**, 301–304 (1999).
72. Lee, E. et al. Landscape of somatic retrotransposition in human cancers. *Science* **337**, 967–971 (2012).
73. Tubio, J. M. C. et al. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**, 1251343–1251343 (2014).
74. Helman, E. et al. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res.* **24**, 1053–1063 (2014).
75. Shay, J. W. & Wright, W. E. Hayflick, his limit, and cellular ageing. *Nat. Rev. Mol. Cell Biol.* **1**, 72–76 (2000).
76. Peifer, M. et al. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526**, 700–704 (2015).
77. Totoki, Y. et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat. Genet.* **46**, 1267–1273 (2014).
78. Paterlini-Bréchot, P. et al. Hepatitis B virus-related insertional mutagenesis occurs frequently in human liver cancers and recurrently targets human telomerase gene. *Oncogene* **22**, 3911–3916 (2003).
79. Heaphy, C. M. et al. Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. *Am. J. Pathol.* **179**, 1608–1615 (2011).
80. Barthel, F. P. et al. Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat. Genet.* **49**, 349–357 (2017).
81. García-Cao, M., Gonzalo, S., Dean, D. & Blasco, M. A. A role for the Rb family of proteins in controlling telomere length. *Nat. Genet.* **32**, 415–419 (2002).
82. Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science* **347**, 78–81 (2015).
83. Gerstung, M. et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. *Nat. Genet.* **49**, 332–340 (2017).
84. O'Connor, B. D. et al. The Dockstore: enabling modular, community-focused sharing of Docker-based genomics tools and workflows. *F1000Res.* **6**, 52 (2017).
85. Zhang, J. et al. The International Cancer Genome Consortium Data Portal. *Nat. Biotechnol.* **37**, 367–369 (2019).
86. Miller, C. A., Qiao, Y., DiSera, T., D'Astous, B. & Marth, G. T. bam.iobio: a web-based, real-time, sequence alignment file inspector. *Nat. Methods* **11**, 1189–1189 (2014).
87. Goldman, M. et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. Preprint at https://www.biorxiv.org/content/10.1101/326470v6 (2019)
88. Papatheodorou, I. et al. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.* **46**, D246–D251 (2018).

Article

The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium

Peter J. Campbell[1,2,745]*, Gad Getz[3,4,5,6,745]*, Jan O. Korbel[7,8,745]*, Joshua M. Stuart[9,745]*, Jennifer L. Jennings[10,11,745], Lincoln D. Stein[12,13,745]*, Marc D. Perry[14,15], Hardeep K. Nahal-Bose[15], B. F. Francis Ouellette[16,17], Constance H. Li[12,18], Esther Rheinbay[3,6,19], G. Petur Nielsen[19], Dennis C. Sgroi[19], Chin-Lee Wu[19], William C. Faquin[19], Vikram Deshpande[19], Paul C. Boutros[12,18,20,21], Alexander J. Lazar[22], Katherine A. Hoadley[23,24], David N. Louis[19], L. Jonathan Dursi[12,25], Christina K. Yung[15], Matthew H. Bailey[26,27], Gordon Saksena[3], Keiran M. Raine[1], Ivo Buchhalter[28,29,30], Kortine Kleinheinz[28,30], Matthias Schlesner[28,31], Junjun Zhang[15], Wenyi Wang[32], David A. Wheeler[33,34], Li Ding[26,27,35], Jared T. Simpson[12,36], Brian D. O'Connor[15,37], Sergei Yakneen[8], Kyle Ellrott[38], Naoki Miyoshi[39], Adam P. Butler[1], Romina Royo[40], Solomon I. Shorser[12], Miguel Vazquez[40,41], Tobias Rausch[8], Grace Tiao[3], Sebastian M. Waszak[8], Bernardo Rodriguez-Martin[42,43,44], Suyash Shringarpure[45], Dai-Ying Wu[46], German M. Demidov[47,48,49], Olivier Delaneau[50,51,52], Shuto Hayashi[39], Seiya Imoto[39], Nina Habermann[8], Ayellet V. Segre[3,53], Erik Garrison[1], Andy Cafferkey[7], Eva G. Alvarez[42,43,44], José María Heredia-Genestar[54], Francesc Muyas[47,48,49], Oliver Drechsel[47,49], Alicia L. Bruzos[42,43,44], Javier Temes[42,43], Jorge Zamora[1,42,43,44], Adrian Baez-Ortega[55], Hyung-Lae Kim[56], R. Jay Mashl[27,57], Kai Ye[58,59], Anthony DiBiase[60], Kuan-lin Huang[27,61], Ivica Letunic[62], Michael D. McLellan[26,27,35], Steven J. Newhouse[7], Tal Shmaya[46], Sushant Kumar[63,64], David C. Wedge[1,65,66], Mark H. Wright[45], Venkata D. Yellapantula[67,68], Mark Gerstein[63,64,69], Ekta Khurana[70,71,72,73], Tomas Marques-Bonet[74,75,76,77], Arcadi Navarro[74,75,76], Carlos D. Bustamante[45,78], Reiner Siebert[79,80], Hidewaki Nakagawa[81], Douglas F. Easton[82,83], Stephan Ossowski[47,48,49], Jose M. C. Tubio[42,43,44], Francisco M. De La Vega[45,46,78], Xavier Estivill[47,84], Denis Yuen[12], George L. Mihaiescu[15], Larsson Omberg[85], Vincent Ferretti[15,86], Radhakrishnan Sabarinathan[87,88,89], Oriol Pich[87,89], Abel Gonzalez-Perez[87,89], Amaro Taylor-Weiner[90], Matthew W. Fittall[91], Jonas Demeulemeester[91,92], Maxime Tarabichi[1,91], Nicola D. Roberts[1], Peter Van Loo[91,92], Isidro Cortés-Ciriano[93,94,95], Lara Urban[7,8], Peter Park[94,95], Bin Zhu[96], Esa Pitkänen[8], Yilong Li[1], Natalie Saini[97], Leszek J. Klimczak[98], Joachim Weischenfeldt[8,99,100], Nikos Sidiropoulos[100], Ludmil B. Alexandrov[1,101], Raquel Rabionet[47,49,102], Georgia Escaramis[47,103,104], Mattia Bosio[40,47,49], Aliaksei Z. Holik[47], Hana Susak[47,49], Aparna Prasad[49], Serap Erkek[8], Claudia Calabrese[7,8], Benjamin Raeder[8], Eoghan Harrington[105], Simon Mayes[106], Daniel Turner[106], Sissel Juul[105], Steven A. Roberts[107], Lei Song[96], Roelof Koster[108], Lisa Mirabello[96], Xing Hua[96], Tomas J. Tanskanen[109], Marta Tojo[44], Jieming Chen[64,110], Lauri A. Aaltonen[111], Gunnar Rätsch[112,113,114,115,116,117], Roland F. Schwarz[7,118,119,120], Atul J. Butte[121], Alvis Brazma[7], Stephen J. Chanock[96], Nilanjan Chatterjee[122,123], Oliver Stegle[7,8,124], Olivier Harismendy[125], G. Steven Bova[126], Dmitry A. Gordenin[97], David Haan[9], Lina Sieverling[127,128], Lars Feuerbach[127], Don Chalmers[129], Yann Joly[130], Bartha Knoppers[130], Fruzsina Molnár-Gábor[131], Mark Phillips[130], Adrian Thorogood[130], David Townend[130], Mary Goldman[132], Nuno A. Fonseca[7,133], Qian Xiang[15], Brian Craft[132], Elena Piñeiro-Yáñez[134], Alfonso Muñoz[7], Robert Petryszak[7], Anja Füllgrabe[7], Fatima Al-Shahrour[134], Maria Keays[7], David Haussler[132,135], John Weinstein[136,137], Wolfgang Huber[8], Alfonso Valencia[40,76], Irene Papatheodorou[7], Jingchun Zhu[132], Yu Fan[32], David Torrents[40,76], Matthias Bieg[138,139], Ken Chen[140], Zechen Chong[141], Kristian Cibulskis[3], Roland Eils[28,30,142,143], Robert S. Fulton[26,27,35], Josep L. Gelpi[40,144], Santiago Gonzalez[7,8], Ivo G. Gut[49,74], Faraz Hach[145,146], Michael Heinold[28,30], Taobo Hu[147], Vincent Huang[12], Barbara Hutter[139,148,149], Natalie Jäger[28], Jongsun Jung[150], Yogesh Kumar[147], Christopher Lalansingh[12], Ignaty Leshchiner[3], Dimitri Livitz[3], Eric Z. Ma[147], Yosef E. Maruvka[3,19,151], Ana Milovanovic[40], Morten Muhlig Nielsen[152], Nagarajan Paramasivam[28,139], Jakob Skou Pedersen[152,153], Montserrat Puiggròs[40], S. Cenk Sahinalp[146,154,155], Iman Sarrafi[146,155], Chip Stewart[3], Miranda D. Stobbe[49,74], Jeremiah A. Wala[3,6,156], Jiayin Wang[27,58,157], Michael Wendl[27,158,159], Johannes Werner[28,160], Zhenggang Wu[147], Hong Xue[147], Takafumi N. Yamaguchi[12], Venkata Yellapantula[67,68], Brandi N. Davis-Dusenbery[161], Robert L. Grossman[162], Youngwook Kim[163,164], Michael C. Heinold[28,30], Jonathan Hinton[1], David R. Jones[1], Andrew Menzies[1], Lucy Stebbings[1], Julian M. Hess[3,151], Mara Rosenberg[3,19], Andrew J. Dunford[3], Manaswi Gupta[3], Marcin Imielinski[165,166], Matthew Meyerson[3,6,156], Rameen Beroukhim[3,6,167], Jüri Reimand[12,18], Priyanka Dhingra[71,73], Francesco Favero[168], Stefan Dentro[1,65,91], Jeff Wintersinger[169,170,171], Vasilisa Rudneva[8], Ji Wan Park[172], Eun Pyo Hong[172], Seong Gu Heo[172], André Kahles[112,113,114,115,116], Kjong-Van Lehmann[112,114,115,173,174], Cameron M. Soulette[37], Yuichi Shiraishi[39], Fenglin Liu[175,176], Yao He[175], Deniz Demircioğlu[177,178], Natalie R. Davidson[112,114,115,117,173], Liliana Greger[7], Siliang Li[179,180], Dongbing Liu[179,180], Stefan G. Stark[115,173,181,182], Fan Zhang[175], Samirkumar B. Amin[183,184,185], Peter Bailey[186], Aurélien Chateigner[15], Milana Frenkel-Morgenstern[187], Yong Hou[179,180], Matthew R. Huska[118], Helena Kilpinen[188], Fabien C. Lamaze[12], Chang Li[179,180], Xiaobo Li[179,180], Xinyue Li[179], Xingmin Liu[179,180], Maximillian G. Marin[37], Julia Markowski[118], Tannistha Nandi[189], Akinyemi I. Ojesina[190,191,192], Qiang Pan-Hammarström[179,193], Peter J. Park[94,95], Chandra Sekhar Pedamallu[3,6,167], Hong Su[179,180], Patrick Tan[189,194,195,196], Bin Tean Teh[194,195,196,197,198], Jian Wang[179], Heng Xiong[179,180], Chen Ye[179,180], Christina Yung[15], Xiuqing Zhang[179], Liangtao Zheng[175], Shida Zhu[179,180], Philip Awadalla[12,13], Chad J. Creighton[199], Kui Wu[179,180], Huanming Yang[179], Jonathan Göke[177,200], Zemin Zhang[175,201], Angela N. Brooks[3,37,156], Matthew W Fittall[91], Iñigo Martincorena[1], Carlota Rubio-Perez[87,89,202], Malene Juul[152], Steven Schumacher[3,203], Ofer Shapira[3,156], David Tamborero[87,89], Loris Mularoni[87,89], Henrik Hornshøj[152], Jordi Deu-Pons[89,204], Ferran Muiños[87,89], Johanna Bertl[152,205], Qianyun Guo[153], Abel Gonzalez-Perez[87,89,206], Qian Xiang[207], Wojciech Bazant[7], Elisabet Barrera[7], Sultan T. Al-Sedairy[208], Axel Aretz[209], Cindy Bell[210], Miguel Betancourt[211], Christiane Buchholz[212], Fabien Calvo[213], Christine Chomienne[214], Michael Dunn[215], Stuart Edmonds[216], Eric Green[217], Shailja Gupta[218], Carolyn M. Hutter[217], Karine Jegalian[219], Nic Jones[220], Youyong Lu[221,222,223], Hitoshi Nakagama[224], Gerd Nettekoven[225], Laura Planko[225], David Scott[220], Tatsuhiro Shibata[226,227], Kiyo Shimizu[228], Michael R. Stratton[1], Takashi Yugawa[228], Giampaolo Tortora[229,230], K. VijayRaghavan[218], Jean C. Zenklusen[231], David Townend[232], Bartha M. Knoppers[130], Brice Aminou[15], Javier Bartolome[40], Keith A. Boroevich[81,233], Rich Boyce[7], Alex Buchanan[38], Niall J. Byrne[15], Zhaohong Chen[234], Sunghoon Cho[235], Wan Choi[236], Peter Clapham[1], Michelle T. Dow[234], Lewis Jonathan Dursi[12,25], Juergen Eils[142,143], Claudiu Farcas[234], Nodirjon Fayzullaev[15], Paul Flicek[7], Allison P. Heath[237], Oliver Hofmann[238], Jongwhi H. Hong[239], Thomas J. Hudson[240,241], Daniel Hübschmann[30,120,142,242,243], Sinisa Ivkovic[244], Seung-Hyup Jeon[236], Wei Jiao[12], Rolf Kabbe[28], Andre Kahles[112,113,114,115,174], Jules N. A. Kerssemakers[28], Hyunghwan Kim[236], Jihoon Kim[245], Michael Koscher[246], Antonios Koures[234], Milena Kovacevic[244], Chris Lawerenz[143], Jia Liu[247], Sanja Mijalkovic[244], Ana Mijalkovic Mijalkovic-Lazic[244], Satoru Miyano[39], Mia Nastic[244], Jonathan Nicholson[1], David Ocana[7], Kazuhiro Ohi[39], Lucila Ohno-Machado[234], Todd D. Pihl[248], Manuel Prinz[28], Petar Radovic[244], Charles Short[7], Heidi J. Sofia[217], Jonathan Spring[162], Adam J. Struck[38], Nebojsa Tijanic[244], David Vicente[40], Zhining Wang[231], Ashley Williams[234], Youngchoon Woo[236], Adam J. Wright[12], Liming Yang[231], Mark P. Hamilton[249], Todd A. Johnson[233], Abdullah Kahraman[250,251,252], Manolis Kellis[3,253], Paz Polak[3,4,6], Richard Sallari[3], Nasa Sinnott-Armstrong[3,45], Christian von Mering[252,254], Sergi Beltran[49,74], Daniela S. Gerhard[255], Marta Gut[49,74], Jean-Rémi Trotta[74], Justin P. Whalley[74], Beifang Niu[256], Shadrielle M. G. Espiritu[12], Shengjie Gao[179], Yi Huang[157,257], Christopher M. Lalansingh[12], Jon W. Teague[1], Michael C. Wendl[27,158,159], Federico Abascal[1], Gary D. Bader[13], Pratiti Bandopadhayay[3,258,259], Jonathan Barenboim[12], Søren Brunak[260,261], Joana Carlevaro-Fita[262,263,264], Dimple Chakravarty[265,266], Calvin Wing Yiu Chan[28,128], Jung Kyoon Choi[267], Klev Diamanti[268], J. Lynn Fink[40,269], Joan Frigola[204], Carlo Gambacorti-Passerini[270], Dale W. Garsed[271], Nicholas J. Haradhvala[3,19], Arif O. Harmanci[64,272], Mohamed Helmy[170], Carl Herrmann[28,30,273], Asger Hobolth[153,205], Ermin Hodzic[155], Chen Hong[127,128], Keren Isaev[12,18], Jose M. G. Izarzugaza[260], Rory Johnson[263,274], Randi Istrup Juul[152], Jaegil Kim[3], Jong K. Kim[275], Jan Komorowski[268,276], Andrés Lanzós[263,264,274], Erik Larsson[112], Donghoon Lee[64], Shantao Li[64], Xiaotong Li[64], Ziao Lin[3,277], Eric Minwei Liu[71,73,278], Lucas Lochovsky[63,64,185], Shaoke Lou[63,64], Tobias Madsen[152], Kathleen Marchal[279,280], Alexander Martinez-Fundichely[71,72,73], Patrick D. McGillivray[63], William Meyerson[64,281], Marta Paczkowska[12], Keunchil Park[282,283], Kiejung Park[284], Tirso Pons[285], Sergio Pulido-Tamayo[279,280], Iker Reyes-Salazar[87], Matthew A. Reyna[286], Mark A. Rubin[274,287,288,289,290], Leonidas Salichos[63,64], Chris Sander[112,156,291,292], Steven E. Schumacher[3,203], Mark Shackleton[271], Ciyue Shen[292,293], Raunak Shrestha[146], Shimin Shuai[12,13], Tatsuhiko Tsunoda[233,294,295,296], Husen M. Umer[268,297], Liis Uusküla-Reimand[298,299], Lieven P. C. Verbeke[280,300], Claes Wadelius[301], Lina Wadi[12], Jonathan Warrell[63,64], Guanming Wu[302], Jun Yu[303], Jing Zhang[64], Xuanping Zhang[157,304], Yan Zhang[64,305,306], Zhongming Zhao[307], Lihua Zou[308], Michael S. Lawrence[3,19,233], Benjamin J. Raphael[286], Peter J. Bailey[186], David Craft[3,309], Mary J. Goldman[132], Hiroyuki Aburatani[310], Hans Binder[311,312], Huy Q. Dinh[313], Simon C. Heath[49,74], Steve Hoffmann[311,312,314,315], Charles David Imbusch[127], Helene Kretzmer[312,315], Peter W. Laird[316], Jose I. Martin-Subero[76,317], Genta Nagae[310,318], Hui Shen[319], Qi Wang[246], Dieter Weichenhan[320], Wanding Zhou[319], Benjamin P. Berman[313,321,322], Benedikt Brors[127,149,323], Christoph Plass[320], Kadir C. Akdemir[140], David D. L. Bowtell[271], Kathleen H. Burns[324,325], John Busanovich[3,326], Kin Chan[327], Ana Dueso-Barroso[40], Paul A. Edwards[328,329], Dariush Etemadmoghadam[271], James E. Haber[330], David T. W. Jones[331,332], Young Seok Ju[1,267], Marat D. Kazanov[333,334,335], Youngil Koh[336,337], Kiran Kumar[3], Eunjung Alice Lee[338], Jake June-Koo Lee[94,95], Andy G. Lynch[328,329,339], Geoff Macintyre[328], Florian Markowetz[328,329], Fabio C. P. Navarro[63], John V. Pearson[340,341], Karsten Rippe[120], Ralph Scully[342], Izar Villasante[40], Nicola Waddell[340,341], Lixing Yang[343], Xiaotong Yao[165,344], Sung-Soo Yoon[337], Cheng-Zhong Zhang[3,6,156], Erik N. Bergstrom[345], Arnoud Boot[195,346], Kyle Covington[34], Akihiro Fujimoto[81], Mi Ni Huang[195,346], S. M. Ashiqul Islam[101], John R. McPherson[195,346], Sandro Morganella[1], Ville Mustonen[347,348,349], Alvin Wei Tian Ng[350], Stephenie D. Prokopec[12], Ignacio Vázquez-García[1,67,351,352], Yang Wu[195,346], Fouad Yousif[12], Willie Yu[353], Steven G. Rozen[195,196,346], Vasilisa A. Rudneva[8], Suyash S. Shringarpure[45], Daniel J. Turner[106], Tian Xia[354], Gurnit Atwal[12,13,171], David K. Chang[186,355], Susanna L. Cooke[186], Bishoy M. Faltas[117], Syed Haider[12], Vera B. Kaiser[356], Rosa Karlić[357], Mamoru Kato[358], Kirsten Kübler[3,6,19], Adam Margolin[38], Sancha Martin[1,359], Serena Nik-Zainal[1,360,361,362], Christine P'ng[12], Colin A. Semple[356], Jaclyn Smith[38], Ren X. Sun[12], Kevin Thai[15], Derek W. Wright[363,364], Ke Yuan[328,359,365], Andrew V. Biankin[186,355,366,367], Levi Garraway[156], Sean M. Grimmond[368], David J. Adams[1], Pavana Anur[369], Shaolong Cao[32], Elizabeth L. Christie[271], Marek Cmero[370,371,372], Yupeng Cun[373], Kevin J. Dawson[1], Stefan C. Dentro[1,65,91], Amit G. Deshwar[374], Nilgun Donmez[146,155], Ruben M. Drews[328], Moritz Gerstung[7,8], Gavin Ha[3], Kerstin Haase[91], Lara Jerman[8,375], Yuan Ji[376,377], Clemency Jolly[91], Juhee Lee[378], Henry Lee-Six[1], Salem Malikic[146,155], Thomas J. Mitchell[1,329,379], Quaid D. Morris[171,380], Layla Oesper[381], Martin Peifer[373], Myron Peto[382], Daniel Rosebrock[3], Yulia Rubanova[36,171], Adriana Salcedo[12], Subhajit Sengupta[383], Ruian Shi[380], Seung Jun Shin[182], Oliver Spiro[3], Shankar Vembu[380,384], Jeffrey A. Wintersinger[169,170,171], Tsun-Po Yang[373], Kaixian Yu[385], Hongtu Zhu[386,387], Paul T. Spellman[388], John N. Weinstein[136,137], Yiwen Chen[32], Masashi Fujita[81], Leng Han[304], Takanori Hasegawa[39], Mitsuhiro Komura[39], Jun Li[32], Shinichi Mizuno[389], Eigo Shimizu[39], Yumeng Wang[32,390], Yanxun Xu[391], Rui Yamaguchi[39], Fan Yang[380], Yang Yang[304], Christopher J. Yoon[267], Yuan Yuan[32], Han Liang[32], Malik Alawi[392,393], Ivan Borozan[12], Daniel S. Brewer[394,395], Colin S. Cooper[395,396,397], Nikita Desai[15], Adam Grundhoff[392,398], Murat Iskar[399], Xiaoping Su[400], Marc Zapatka[399], Peter Lichter[148,399], Kathryn Alsop[271], Timothy J. C. Bruxner[269], Angelika N. Christ[269], Stephen M. Cordner[401], Prue A. Cowin[402], Ronny Drapkin[403], Sian Fereday[402], Joshy George[185], Anne Hamilton[402], Oliver Holmes[340,341], Jillian A. Hung[404,405], Karin S. Kassahn[269,406], Stephen H. Kazakoff[340,341], Catherine J. Kennedy[407,408], Conrad R. Leonard[340,341], Linda Mileshkin[271], David K. Miller[269,355,409], Gisela Mir Arnau[402], Chris Mitchell[402], Felicity Newell[340,341], Katia Nones[340,341], Ann-Marie Patch[340,341], Michael C. Quinn[340,341], Darrin F. Taylor[269], Heather Thorne[402], Nadia Traficante[402],

Ravikiran Vedururu[402], Nick M. Waddell[341], Paul M. Waring[410], Scott Wood[340,341], Qinying Xu[340,341], Anna deFazio[411,412,413], Matthew J. Anderson[269], Davide Antonello[414], Andrew P. Barbour[415,416], Claudio Bassi[414], Samantha Bersani[417], Ivana Cataldo[417,418], Lorraine A. Chantrill[355,419], Yoke-Eng Chiew[411], Angela Chou[355,420], Sara Cingarlini[229], Nicole Cloonan[421], Vincenzo Corbo[418,422], Maria Vittoria Davi[423], Fraser R. Duthie[186,424], Anthony J. Gill[355,420], Janet S. Graham[186,425], Ivon Harliwong[269], Nigel B. Jamieson[186,367,426], Amber L. Johns[355,409], James G. Kench[355,420,427], Luca Landoni[414], Rita T. Lawlor[418], Andrea Mafficini[418], Neil D. Merrett[414,428], Marco Miotto[414], Elizabeth A. Musgrove[186], Adnan M. Nagrial[355], Karin A. Oien[410,429], Marina Pajic[355], Mark Pinese[430], Alan J. Robertson[269], Ilse Rooman[355], Borislav C. Rusev[418], Jaswinder S. Samra[414,420], Maria Scardoni[417], Christopher J. Scarlett[355,431], Aldo Scarpa[418], Elisabetta Sereni[414], Katarzyna O. Sikora[418], Michele Simbolo[422], Morgan L. Taschuk[15], Christopher W. Toon[355], Caterina Vicentini[418], Jianmin Wu[355], Nikolajs Zeps[432,433], Andreas Behren[434], Hazel Burke[435], Jonathan Cebon[434], Rebecca A. Dagg[436], Ricardo De Paoli-Iseppi[437], Ken Dutton-Regester[340], Matthew A. Field[438], Anna Fitzgerald[439], Peter Hersey[435], Valerie Jakrot[435], Peter A. Johansson[340], Hojabr Kakavand[437], Richard F. Kefford[440], Loretta M. S. Lau[441], Georgina V. Long[442], Hilda A. Pickett[441], Antonia L. Pritchard[340], Gulietta M. Pupo[443], Robyn P. M. Saw[442], Sarah-Jane Schramm[444], Catherine A. Shang[439], Ping Shang[442], Andrew J. Spillane[442], Jonathan R. Stretch[442], Varsha Tembe[411,444], John F. Thompson[442], Ricardo E. Vilain[445], James S. Wilmott[442], Jean Y. Yang[446], Nicholas K. Hayward[340,435], Graham J. Mann[411,447], Richard A. Scolyer[412,442,445,448], John Bartlett[449,450], Prashant Bavi[451], Dianne E. Chadwick[452], Michelle Chan-Seng-Yue[451], Sean Cleary[451,453], Ashton A. Connor[453,454], Karolina Czajka[241], Robert E. Denroche[451], Neesha C. Dhani[455], Jenna Eagles[241], Steven Gallinger[451,453,454], Robert C. Grant[451,454], David Hedley[455], Michael A. Hollingsworth[456], Gun Ho Jang[451], Jeremy Johns[241], Sangeetha Kalimuthu[451], Sheng-Ben Liang[457], Ilinca Lungu[451,458], Xuemei Luo[12], Faridah Mbabaali[241], Treasa A. McPherson[454], Jessica K. Miller[241], Malcolm J. Moore[455], Faiyaz Notta[451,459], Danielle Pasternack[241], Gloria M. Petersen[460], Michael H. A. Roehrl[18,451,461,462,463], Michelle Sam[241], Iris Selander[454], Stefano Serra[410], Sagedeh Shahabi[457], Sarah P. Thayer[456], Lee E. Timms[241], Gavin W. Wilson[12,451], Julie M. Wilson[451], Bradly G. Wouters[464], John D. McPherson[241,451,465], Timothy A. Beck[15,466], Vinayak Bhandari[12], Colin C. Collins[146], Neil E. Fleshner[467], Natalie S. Fox[12], Michael Fraser[12], Lawrence E. Heisler[468], Emilie Lalonde[12], Julie Livingstone[12], Alice Meng[469], Veronica Y. Sabelnykova[12], Yu-Jia Shiah[12], Theodorus Van der Kwast[470], Robert G. Bristow[18,471,472,473,474], Shuai Ding[475], Daiming Fan[476], Lin Li[179], Yongzhan Nie[476,477], Xiao Xiao[157], Rui Xing[222,478], Shanlin Yang[475], Yingyan Yu[479], Yong Zhou[179], Rosamonde E. Banks[480], Guillaume Bourque[481,482], Paul Brennan[483], Louis Letourneau[484], Yasser Riazalhosseini[482], Ghislaine Scelo[483], Naveen Vasudev[480,485], Juris Viksna[486], Mark Lathrop[482], Jörg Tost[487], Sung-Min Ahn[488], Samuel Aparicio[489], Laurent Arnould[490], M. R. Aure[491], Shriram G. Bhosle[1], Ewan Birney[7], Ake Borg[492], Sandrine Boyault[493], Arie B. Brinkman[494], Jane E. Brock[495], Annegien Broeks[496], Anne-Lise Børresen-Dale[491], Carlos Caldas[497,498], Suet-Feung Chin[497,498], Helen Davies[1,360,361], Christine Desmedt[499,500], Luc Dirix[501], Serge Dronov[1], Anna Ehinger[502], Jorunn E. Eyfjord[503], Aquila Fatima[203], John A. Foekens[504], P. Andrew Futreal[505], Øystein Garred[506,507], Dilip D. Giri[508], Dominik Glodzik[1], Dorthe Grabau[509], Holmfridur Hilmarsdottir[503], Gerrit K. Hooijer[510], Jocelyne Jacquemier[511], Se Jin Jang[512], Jon G. Jonasson[503], Jos Jonkers[513], Hyung-Yong Kim[511], Tari A. King[514,515], Stian Knappskog[1,516], Gu Kong[511], Savitri Krishnamurthy[517], Sunil R. Lakhani[518], Anita Langerød[491], Denis Larsimont[519], Hee Jin Lee[512], Jeong-Yeon Lee[520], Ming Ta Michael Lee[505], Ole Christian Lingjærde[521], Gaetan MacGrogan[522], John W. M. Martens[504], Sarah O'Meara[1], Iris Pauporté[214], Sarah Pinder[523], Xavier Pivot[524], Elena Provenzano[525], Colin A. Purdie[526], Manasa Ramakrishna[1], Kamna Ramakrishnan[1], Jorge Reis-Filho[508], Andrea L. Richardson[203], Markus Ringnér[492], Javier Bartolomé Rodriguez[40], F. Germán Rodríguez-González[261], Gilles Romieu[527], Roberto Salgado[410], Torill Sauer[521], Rebecca Shepherd[1], Anieta M. Sieuwerts[504], Peter T. Simpson[518], Marcel Smid[504], Christos Sotiriou[234], Paul N. Span[528], Ólafur Andri Stefánsson[529], Alasdair Stenhouse[530], Henk G. Stunnenberg[180,531], Fred Sweep[532], Benita Kiat Tee Tan[533], Gilles Thomas[534], Alastair M. Thompson[530], Stefania Tommasi[535], Isabelle Treilleux[536,537], Andrew Tutt[203], Naoto T. Ueno[387], Steven Van Laere[501], Gert G. Van den Eynden[501], Peter Vermeulen[501], Alain Viari[418], Anne Vincent-Salomon[531], Bernice H. Wong[538], Lucy Yates[1], Xueqing Zou[1], Carolien H. M. van Deurzen[539], Marc J. van de Vijver[410], Laura van't Veer[540], Ole Ammerpohl[541,542,543], Sietse Aukema[542,543,544], Anke K. Bergmann[545], Stephan H. Bernhart[311,312,315], Arndt Borkhardt[546], Christoph Borst[547], Birgit Burkhardt[548], Alexander Claviez[549], Maria Elisabeth Goebler[550], Andrea Haake[541], Siegfried Haas[547], Martin Hansmann[551], Jessica I. Hoell[546], Michael Hummel[552], Dennis Karsch[553], Wolfram Klapper[544], Michael Kneba[553], Markus Kreuz[554], Dieter Kube[555], Ralf Küppers[556], Dido Lenze[552], Markus Loeffler[554], Cristina López[80,541], Luisa Mantovani-Löffler[557], Peter Möller[558], German Ott[559], Bernhard Radlwimmer[399], Julia Richter[541,544], Marius Rohde[560], Philip C. Rosenstiel[561], Andreas Rosenwald[562], Markus B. Schilhabel[561], Stefan Schreiber[563], Peter F. Stadler[311,312,315], Peter Staib[564], Stephan Stilgenbauer[565], Stephanie Sungalee[8], Monika Szczepanowski[544], Umut H. Toprak[30,566], Lorenz H. P. Trümper[555], Rabea Wagener[80,541], Thorsten Zenz[149], Volker Hovestadt[399], Christof von Kalle[120], Marcel Kool[246,331], Andrey Korshunov[246], Pablo Landgraf[567,568], Hans Lehrach[569], Paul A. Northcott[570], Stefan M. Pfister[246,331,571], Guido Reifenberger[568], Hans-Jörg Warnatz[569], Stephan Wolf[572], Marie-Laure Yaspo[569], Yassen Assenov[573], Clarissa Gerhauser[320], Sarah Minner[574], Thorsten Schlomm[99,575], Ronald Simon[576], Guido Sauter[576], Holger Sültmann[149,577], Nidhan K. Biswas[578], Arindam Maitra[578], Partha P. Majumder[578], Rajiv Sarin[579], Stefano Barbi[422], Giada Bonizzato[418], Cinzia Cantù[418], Angelo P. Dei Tos[580], Matteo Fassan[581], Sonia Grimaldi[418], Claudio Luchini[417], Giuseppe Malleo[414], Giovanni Marchegiani[414], Michele Milella[229], Salvatore Paiella[414], Antonio Pea[414], Paolo Pederzoli[414], Andrea Ruzzenente[414], Roberto Salvia[414], Nicola Sperandio[418], Yasuhito Arai[226], Natsuko Hama[226], Nobuyoshi Hiraoka[582], Fumie Hosoda[226], Hiromi Nakamura[226], Hidenori Ojima[583], Takuji Okusaka[584], Yasushi Totoki[226], Tomoko Urushidate[227], Masashi Fukayama[585], Shumpei Ishikawa[586], Hitoshi Katai[587], Hiroto Katoh[586], Daisuke Komura[586], Hirofumi Rokutan[358], Mihoko Saito-Adachi[358], Akihiro Suzuki[310,588], Hirokazu Taniguchi[589], Kenji Tatsuno[310], Tetsuo Ushiku[585], Shinichi Yachida[226,590], Shogo Yamamoto[310], Hiroshi Aikata[591], Koji Arihiro[591], Shun-ichi Ariizumi[592], Kazuaki Chayama[591], Mayuko Furuta[81], Kunihito Gotoh[593], Shinya Hayami[594], Satoshi Hirano[595], Yoshiiku Kawakami[591], Kazuhiro Maejima[81], Toru Nakamura[595], Kaoru Nakano[81], Hideki Ohdan[591], Aya Sasaki-Oku[81], Hiroko Tanaka[39], Masaki Ueno[594], Masakazu Yamamoto[592], Hiroki Yamaue[594], Su Pin Choo[596], Ioana Cutcutache[195,346], Narong Khuntikeo[414,597], Choon Kiat Ong[598], Chawalit Pairojkul[410], Irinel Popescu[599], Keun Soo Ahn[600], Marta Aymerich[601], Armando Lopez-Guillermo[602], Carlos López-Otín[603], Xose S. Puente[603], Elias Campo[604,605], Fernanda Amary[606], Daniel Baumhoer[607], Sam Behjati[1], Bodil Bjerkehagen[607,608], P. A. Futreal[505], Ola Myklebost[516], Nischalan Pillay[609], Patrick Tarpey[610], Roberto Tirabosco[611], Olga Zaikova[612], Adrienne M. Flanagan[613], Jacqueline Boultwood[614], David T. Bowen[1], Mario Cazzola[615], Anthony R. Green[329], Eva Hellstrom-Lindberg[616], Luca Malcovati[615], Jyoti Nangalia[617], Elli Papaemmanuil[1], Paresh Vyas[340,618], Yeng Ang[619], Hugh Barr[620], Duncan Beardsmore[621], Matthew Eldridge[328], James Gossage[622], Nicola Grehan[361], George B. Hanna[623], Stephen J. Hayes[624,625], Ted R. Hupp[626], David Khoo[627], Jesper Lagergren[616,628], Laurence B. Lovat[188], Shona MacRae[136], Maria O'Donovan[361], J. Robert O'Neill[629], Simon L. Parsons[630], Shaun R. Preston[631], Sonia Puig[632], Tom Roques[633], Grant Sanders[24], Sharmila Sothi[634], Simon Tavaré[328], Olga Tucker[635], Richard Turkington[636], Timothy J. Underwood[637], Ian Welch[638], Rebecca C. Fitzgerald[361], Daniel M. Berney[639], Johann S. De Bono[396], Declan Cahill[640], Niedzica Camacho[396], Nening M. Dennis[640], Tim Dudderidge[640,641], Sandra E. Edwards[396], Cyril Fisher[640], Christopher S. Foster[642,643], Mohammed Ghori[1], Pelvender Gill[618], Vincent J. Gnanapragasam[379,644], Gunes Gundem[278], Freddie C. Hamdy[645], Steve Hawkins[328], Steven Hazell[640], William Howat[379], William B. Isaacs[646], Katalin Karaszi[618], Jonathan D. Kay[188], Vincent Khoo[640], Zsofia Kote-Jarai[396], Barbara Kremeyer[1], Pardeep Kumar[640], Adam Lambert[618], Daniel A. Leongamornlert[1,396], Naomi Livni[640], Yong-Jie Lu[639,647], Hayley J. Luxton[188], Luke Marsden[618], Charlie E. Massie[328], Lucy Matthews[396], Erik Mayer[640,648], Ultan McDermott[1], Sue Merson[396], David E. Neal[328,379], Anthony Ng[649], David Nicol[640], Christopher Ogden[640], Edward W. Rowe[640], Nimish C. Shah[379], Sarah Thomas[640], Alan Thompson[640], Clare Verrill[618,650], Tapio Visakorpi[126], Anne Y. Warren[379,651], Hayley C. Whitaker[188], Hongwei Zhang[647], Nicholas van As[640], Rosalind A. Eeles[396,640], Adam Abeshouse[278], Nishant Agrawal[162], Rehan Akbani[361,652], Hikmat Al-Ahmadie[278], Monique Albert[450], Kenneth Aldape[400,653], Adrian Ally[654], Elizabeth L. Appelbaum[27,188], Joshua Armenia[655], Sylvia Asa[630,656], J. Todd Auman[657], Miruna Balasundaram[654], Saianand Balu[24], Jill Barnholtz-Sloan[658,659], Oliver F. Bathe[660,661], Stephen B. Baylin[123,641], Christopher Benz[662], Andrew Berchuck[663], Mario Berrios[664], Darell Bigner[665], Michael Birrer[19], Tom Bodenheimer[24], Lori Boice[632], Moiz S. Bootwalla[664], Marcus Bosenberg[666], Reanne Bowlby[654], Jeffrey Boyd[667], Russell R. Broaddus[400], Malcolm Brock[668], Denise Brooks[654], Susan Bullman[3,167], Samantha J. Caesar-Johnson[231], Thomas E. Carey[669], Rebecca Carlsen[654], Robert Cerfolio[670], Vishal S. Chandan[671], Hsiao-Wei Chen[619,655], Andrew D. Cherniack[3,156,167], Jeremy Chien[672], Juok Cho[3], Eric Chuah[654], Carrie Cibulskis[3], Leslie Cope[673], Matthew G. Cordes[27,633], Erin Curley[674], Bogdan Czerniak[400,627], Ludmila Danilova[673], Ian J. Davis[675], Timothy Defreitas[3], John A. Demchok[231], Noreen Dhalla[654], Rajiv Dhir[676], HarshaVardhan Doddapaneni[34], Adel El-Naggar[400,627], Ina Felau[231], Martin L. Ferguson[677], Gaetano Finocchiaro[678], Kwun M. Fong[679], Scott Frazer[3], William Friedman[680], Catrina C. Fronick[27,633], Lucinda A. Fulton[27], Stacey B. Gabriel[3], Jianjiong Gao[655], Nils Gehlenborg[3,681], Jeffrey E. Gershenwald[682,683], Ronald Ghossein[508], Nasra H. Giama[684], Richard A. Gibbs[34], Carmen Gomez[685], Ramaswamy Govindan[26], D. Neil Hayes[24,686,687], Apurva M. Hegde[136,137], David I. Heiman[3], Zachary Heins[278], Austin J. Hepperla[24], Andrea Holbrook[664], Robert A. Holt[654], Alan P. Hoyle[24], Ralph H. Hruban[673], Jianhong Hu[34], Mei Huang[632], David Huntsman[688], Jason Huse[278], Christine A. Iacobuzio-Donahue[508], Michael Ittmann[689,690], Joy C. Jayaseelan[34], Stuart R. Jefferys[24], Corbin D. Jones[691], Steven J. M. Jones[692], Hartmut Juhl[693], Koo Jeong Kang[694], Beth Karlan[695], Katayoon Kasaian[692], Electron Kebebew[696,697], Hark Kyun Kim[698], Viktoriya Korchina[34], Ritika Kundra[619,655], Phillip H. Lai[664], Eric Lander[3], Xuan Le[699], Darlene Lee[654], Douglas A. Levine[278,700], Lora Lewis[34], Tim Ley[701], Haiyan Irene Li[654], Pei Lin[3], W. M. Linehan[702], Fei Fei Liu[380], Yiling Lu[137], Lisa Lype[703], Yussanne Ma[654], Dennis T. Maglinte[664,704], Elaine R. Mardis[27,667,705], Jeffrey Marks[414,706], Marco A. Marra[654], Thomas J. Matthew[37], Michael Mayo[654], Karen McCune[707], Samuel R. Meier[3], Shaowu Meng[24], Piotr A. Mieczkowski[23], Tom Mikkelsen[708], Christopher A. Miller[27], Gordon B. Mills[709], Richard A. Moore[654], Carl Morrison[410,710], Lisle E. Mose[24], Catherine D. Moser[684], Andrew J. Mungall[654], Karen Mungall[654], David Mutch[711], Donna M. Muzny[712], Jerome Myers[713], Yulia Newton[37], Michael S. Noble[3], Peter O'Donnell[714], Brian Patrick O'Neill[715], Angelica Ochoa[278], Joong-Won Park[716], Joel S. Parker[717], Harvey Pass[718], Alessandro Pastore[112], Nathan A. Pennell[719], Charles M. Perou[720], Nicholas Petrelli[721], Olga Potapova[722], Janet S. Rader[723], Suresh Ramalingam[724], W. Kimryn Rathmell[725], Victor Reuter[508], Sheila M. Reynolds[703], Matthew Ringel[726], Jeffrey Roach[727], Lewis R. Roberts[684], A. Gordon Robertson[654], Sara Sadeghi[654], Charles Saller[728], Francisco Sanchez-Vega[619,655], Dirk Schadendorf[148,729], Jacqueline E. Schein[654], Heather K. Schmidt[27], Nikolaus Schultz[655], Raja Seethala[730], Yasin Senbabaoglu[112], Troy Shelton[674], Yan Shi[24], Juliann Shih[3,167], Ilya Shmulevich[703], Craig Shriver[731], Sabina Signoretti[167,263,732], Janae V. Simons[24], Samuel Singer[414,733], Payal Sipahimalani[654], Tara J. Skelly[23], Karen Smith-McCune[707], Nicholas D. Socci[112], Matthew G. Soloway[717], Anil K. Sood[734], Angela Tam[654], Donghui Tan[23], Roy Tarnuzzer[231], Nina Thiessen[654], R. Houston Thompson[735], Leigh B. Thorne[632], Ming Tsao[630,656], Christopher Umbricht[324,621,736], David J. Van Den Berg[664], Erwin G. Van Meir[737], Umadevi Veluvolu[23], Douglas Voet[3], Linghua Wang[34], Paul Weinberger[738],

Daniel J. Weisenberger[664], Dennis Wigle[739], Matthew D. Wilkerson[23], Richard K. Wilson[27,740], Boris Winterhoff[741], Maciej Wiznerowicz[742,743], Tina Wong[27,654], Winghing Wong[744], Liu Xi[34], Christina Yau[662], Hailei Zhang[3], Hongxin Zhang[655] & Jiashan Zhang[231]

[1]Wellcome Sanger Institute, Hinxton, UK. [2]Department of Haematology, University of Cambridge, Cambridge, UK. [3]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Center for Cancer Research, Massachusetts General Hospital, Boston, MA, USA. [5]Department of Pathology, Massachusetts General Hospital, Boston, MA, USA. [6]Harvard Medical School, Boston, MA, USA. [7]European Molecular Biology Laboratory (EMBL), European Bioinformatics Institute (EMBL-EBI), Hinxton, UK. [8]European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany. [9]Biomolecular Engineering Department, University of California Santa Cruz, Santa Cruz, CA, USA. [10]Adaptive Oncology Initiative, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [11]International Cancer Genome Consortium (ICGC)/ICGC Accelerating Research in Genomic Oncology (ICGC-ARGO) Secretariat, Toronto, Ontario, Canada. [12]Computational Biology Program, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [13]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [14]Department of Radiation Oncology, University of California San Francisco, San Francisco, CA, USA. [15]Genome Informatics Program, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [16]Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada. [17]Genome Informatics, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [18]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. [19]Massachusetts General Hospital, Boston, MA, USA. [20]Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada. [21]University of California Los Angeles, Los Angeles, CA, USA. [22]Department of Pathology, Department of Genomic Medicine and Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [23]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [24]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [25]The Hospital for Sick Children, Toronto, Ontario, Canada. [26]Alvin J. Siteman Cancer Center, Washington University School of Medicine, St Louis, MO, USA. [27]The McDonnell Genome Institute, Washington University, St Louis, MO, USA. [28]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [29]Heidelberg Center for Personalized Oncology (DKFZ-HIPO), German Cancer Research Center, Heidelberg, Germany. [30]Institute of Pharmacy and Molecular Biotechnology, and BioQuant, Heidelberg University, Heidelberg, Germany. [31]Bioinformatics and Omics Data Analytics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [32]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [33]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. [34]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. [35]Department of Genetics and Department of Medicine, Washington University in St Louis, St Louis, MO, USA. [36]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. [37]University of California Santa Cruz, Santa Cruz, CA, USA. [38]Computational Biology Program, Oregon Health & Science University, Portland, OR, USA. [39]The Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [40]Barcelona Supercomputing Center (BSC), Barcelona, Spain. [41]Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway. [42]Centre for Research in Molecular Medicine and Chronic Diseases (CiMUS), Universidade de Santiago de Compostela, Santiago de Compostela, Spain. [43]Department of Zoology, Genetics and Physical Anthropology, Centre for Research in Molecular Medicine and Chronic Diseases (CiMUS), Universidade de Santiago de Compostela, Santiago de Compostela, Spain. [44]The Biomedical Research Centre (CINBIO), Universidade de Vigo, Vigo, Spain. [45]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [46]Annai Systems, Carlsbad, CA, USA. [47]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [48]Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany. [49]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [50]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [51]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland. [52]Swiss Institute of Bioinformatics, University of Geneva, Geneva, Switzerland. [53]Department of Ophthalmology, Ocular Genomics Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA, USA. [54]Department of Experimental and Health Sciences, Institute of Evolutionary Biology (UPF-CSIC), Universitat Pompeu Fabra (UPF), Barcelona, Spain. [55]Department of Veterinary Medicine, Transmissible Cancer Group, University of Cambridge, Cambridge, UK. [56]Department of Biochemistry, College of Medicine, Ewha Womans University, Seoul, South Korea. [57]Division of Oncology, Washington University School of Medicine, St Louis, MO, USA. [58]School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China. [59]The First Affiliated Hospital, Xi'an Jiaotong University, Xi'an, China. [60]Independent Consultant, Wellesley, MA, USA. [61]Icahn School of Medicine at Mount Sinai, New York, NY, USA. [62]Biobyte Solutions, Heidelberg, Germany. [63]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA. [64]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA. [65]Big Data Institute, Li Ka Shing Centre, University of Oxford, Oxford, UK. [66]Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford, UK. [67]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [68]The McDonnell Genome Institute at Washington University School of Medicine, and Department of Genetics and Department of Medicine, Siteman Cancer Center, Washington University in St Louis, St Louis, MO, USA. [69]Department of Computer Science, Yale University, New Haven, CT, USA. [70]Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA. [71]Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA. [72]Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA. [73]Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA. [74]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [75]Department of Experimental and Health Sciences, Institute of Evolutionary Biology (UPF-CSIC), Universitat Pompeu Fabra (UPF), Barcelona, Spain. [76]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. [77]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Barcelona, Spain. [78]Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA, USA. [79]Human Genetics, University of Kiel, Kiel, Germany. [80]Institute of Human Genetics, Ulm University and Ulm University Medical Center, Ulm, Germany. [81]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [82]Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK. [83]Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK. [84]Quantitative Genomics Laboratories (qGenomics), Barcelona, Spain. [85]Sage Bionetworks, Seattle, WA, USA. [86]Department of Biochemistry and Molecular Medicine, University of Montreal, Montreal, Quebec, Canada. [87]Institute for Research in Biomedicine (IRB Barcelona), Barcelona, Spain. [88]National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India. [89]Research Program on Biomedical Informatics, Universitat Pompeu Fabra (UPF), Barcelona, Spain. [90]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [91]The Francis Crick Institute, London, UK. [92]University of Leuven, Leuven, Belgium. [93]Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Cambridge, UK. [94]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [95]Ludwig Center at Harvard Medical School, Boston, MA, USA. [96]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [97]Genome Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences (NIEHS), Durham, NC, USA. [98]Integrative Bioinformatics Support Group, National Institute of Environmental Health Sciences (NIEHS), Durham, NC, USA. [99]Department of Urology, Charité Universitätsmedizin Berlin, Berlin, Germany. [100]Finsen Laboratory and Biotech Research & Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark. [101]Department of Bioengineering and Department of Cellular and Molecular Medicine, Moores Cancer Center, University of California San Diego, La Jolla, CA, USA. [102]Department of Genetics, Microbiology and Statistics, University of Barcelona, IRSJD, IBUB, Barcelona, Spain. [103]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. [104]Research Group on Statistics, Econometrics and Health (GRECS), UdG, Barcelona, Spain. [105]Oxford Nanopore Technologies, New York, NY, USA. [106]Applications Department, Oxford Nanopore Technologies, Oxford, UK. [107]School of Molecular Biosciences and Center for Reproductive Biology, Washington State University, Pullman, WA, USA. [108]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [109]Department of Medical and Clinical Genetics, Genome-Scale Biology Research Program, University of Helsinki, Helsinki, Finland. [110]Integrated Graduate Program in Physical and Engineering Biology, Yale University, New Haven, CT, USA. [111]Applied Tumor Genomics Research Program, Research Programs Unit, University of Helsinki, Helsinki, Finland. [112]Computational Biology Center, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [113]Department of Biology, ETH Zurich, Zurich, Switzerland. [114]Department of Computer Science, ETH Zurich, Zurich, Switzerland. [115]SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland. [116]University Hospital Zurich, Zurich, Switzerland. [117]Weill Cornell Medical College, New York, NY, USA. [118]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany. [119]German Cancer Consortium (DKTK), Partner site Berlin, Berlin, Germany. [120]German Cancer Research Center (DKFZ), Heidelberg, Germany. [121]Bakar Computational Health Sciences Institute and Department of Pediatrics, University of California, San Francisco, CA, USA. [122]Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. [123]Department of Oncology, The Johns Hopkins School of Medicine, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, MD, USA. [124]Division of Computational Genomics and Systems Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [125]Department of Medicine and Moores Cancer Center, Division of Biomedical Informatics, UC San Diego School of Medicine, San Diego, CA, USA. [126]Faculty of Medicine and Health Technology, Tampere University and Tays Cancer Center, Tampere University Hospital, Tampere, Finland. [127]Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [128]Faculty of Biosciences, Heidelberg University, Heidelberg, Germany. [129]Centre for Law and Genetics, University of Tasmania, Hobart, Tasmania, Australia. [130]Centre of Genomics and Policy, McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada. [131]Heidelberg Academy of Sciences and Humanities, Heidelberg, Germany. [132]UC Santa Cruz Genomics Institute, University of California Santa Cruz, Santa Cruz, CA, USA. [133]CIBIO/InBIO, Research Center in Biodiversity and Genetic Resources, Universidade do Porto, Vairão, Portugal. [134]Bioinformatics Unit, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. [135]Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA, USA. [136]Cancer Unit, MRC University of Cambridge, Cambridge, UK. [137]Department of Bioinformatics and Computational Biology and Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [138]Center for Digital Health, Berlin Institute of Health (BIH) and Charitè–Universitätsmedizin Berlin, Berlin, Germany. [139]Heidelberg Center for Personalized Oncology (DKFZ-HIPO), German Cancer

Research Center (DKFZ), Heidelberg, Germany. [140]Department of Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center, Houston, TX, USA. [141]Department of Genetics and Informatics Institute, University of Alabama at Birmingham, Birmingham, AL, USA. [142]Heidelberg University, Heidelberg, Germany. [143]New BIH Digital Health Center, Berlin Institute of Health (BIH) and Charité–Universitätsmedizin Berlin, Berlin, Germany. [144]Department of Biochemistry and Molecular Biomedicine, University of Barcelona, Barcelona, Spain. [145]Department of Urologic Sciences, University of British Columbia, Vancouver, British Columbia, Canada. [146]Vancouver Prostate Centre, Vancouver, British Columbia, Canada. [147]Division of Life Science and Applied Genomics Center, Hong Kong University of Science and Technology, Hong Kong, China. [148]German Cancer Consortium (DKTK), Heidelberg, Germany. [149]National Center for Tumor Diseases (NCT) Heidelberg, Heidelberg, Germany. [150]Genome Integration Data Center, Syntekabio, Daejon, South Korea. [151]Massachusetts General Hospital Center for Cancer Research, Charlestown, MA, USA. [152]Department of Molecular Medicine (MOMA), Aarhus University Hospital, Aarhus, Denmark. [153]Bioinformatics Research Centre (BiRC), Aarhus University, Aarhus, Denmark. [154]Indiana University, Bloomington, IN, USA. [155]Simon Fraser University, Burnaby, British Columbia, Canada. [156]Dana-Farber Cancer Institute, Boston, MA, USA. [157]School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China. [158]Department of Genetics, Washington University School of Medicine, St Louis, MO, USA. [159]Department of Mathematics, Washington University in St Louis, St Louis, MO, USA. [160]Department of Biological Oceanography, Leibniz Institute of Baltic Sea Research, Rostock, Germany. [161]Seven Bridges Genomics, Charlestown, MA, USA. [162]University of Chicago, Chicago, IL, USA. [163]Department of Health Sciences and Technology, Sungkyunkwan University School of Medicine, Seoul, South Korea. [164]Samsung Genome Institute, Seoul, South Korea. [165]New York Genome Center, New York, NY, USA. [166]Weill Cornell Medicine, New York, NY, USA. [167]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [168]Rigshospitalet, Copenhagen, Denmark. [169]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada. [170]The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada. [171]Vector Institute, Toronto, Ontario, Canada. [172]Department of Medical Genetics, College of Medicine, Hallym University, Chuncheon, South Korea. [173]Department of Biology, ETH Zurich, Zurich, Switzerland. [174]University Hospital Zurich, Zurich, Switzerland. [175]Peking University, Beijing, China. [176]School of Life Sciences, Peking University, Beijing, China. [177]Computational and Systems Biology, Genome Institute of Singapore, Singapore, Singapore. [178]School of Computing, National University of Singapore, Singapore, Singapore. [179]BGI-Shenzhen, Shenzhen, China. [180]China National GeneBank-Shenzhen, Shenzhen, China. [181]Computational & Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [182]Korea University, Seoul, South Korea. [183]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [184]Quantitative & Computational Biosciences Graduate Program, Baylor College of Medicine, Houston, TX, USA. [185]The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA. [186]Wolfson Wohl Cancer Research Centre, Institute of Cancer Sciences, University of Glasgow, Bearsden, UK. [187]The Azrieli Faculty of Medicine, Bar-Ilan University, Safed, Israel. [188]University College London, London, UK. [189]Genome Institute of Singapore, Singapore, Singapore. [190]Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA. [191]HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA. [192]O'Neal Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, AL, USA. [193]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden. [194]Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore. [195]Programme in Cancer & Stem Cell Biology, Duke-NUS Medical School, Singapore, Singapore. [196]SingHealth, Duke-NUS Institute of Precision Medicine, National Heart Centre Singapore, Singapore, Singapore. [197]Institute of Molecular and Cell Biology, Singapore, Singapore. [198]Laboratory of Cancer Epigenome, Division of Medical Science, National Cancer Centre Singapore, Singapore, Singapore. [199]Department of Medicine, Baylor College of Medicine, Houston, TX, USA. [200]National Cancer Centre Singapore, Singapore, Singapore. [201]BIOPIC, ICG and College of Life Sciences, Peking University, Beijing, China. [202]Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain. [203]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA. [204]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [205]Department of Mathematics, Aarhus University, Aarhus, Denmark. [206]Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain. [207]Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [208]King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia. [209]DLR Project Management Agency, Bonn, Germany. [210]Genome Canada, Ottawa, Ontario, Canada. [211]Instituto Carlos Slim de la Salud, Mexico City, Mexico. [212]Federal Ministry of Education and Research, Berlin, Germany. [213]Institut Gustave Roussy, Villejuif, France. [214]Institut National du Cancer (INCA), Boulogne-Billancourt, France. [215]The Wellcome Trust, London, UK. [216]Prostate Cancer Canada, Toronto, Ontario, Canada. [217]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA. [218]Department of Biotechnology, Ministry of Science & Technology, Government of India, New Delhi, Delhi, India. [219]Science Writer, Garrett Park, MD, USA. [220]Cancer Research UK, London, UK. [221]Chinese Cancer Genome Consortium, Shenzhen, China. [222]Laboratory of Molecular Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital & Institute, Beijing, China. [223]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital & Institute, Beijing, China. [224]National Cancer Center, Tokyo, Japan. [225]German Cancer Aid, Bonn, Germany. [226]Division of Cancer Genomics, National Cancer Center Research Institute, National Cancer Center, Tokyo, Japan. [227]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan. [228]Japan Agency for Medical Research and Development, Chiyoda-ku, Tokyo, Japan. [229]Medical Oncology, University and Hospital Trust of Verona, Verona, Italy. [230]University of Verona, Verona, Italy. [231]National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [232]CAPHRI Research School, Maastricht University, Maastricht, The Netherlands. [233]Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [234]University of California San Diego, San Diego, CA, USA. [235]PDXen Biosystems, Seoul, South Korea. [236]Electronics and Telecommunications Research Institute, Daejeon, South Korea. [237]Children's Hospital of Philadelphia, Philadelphia, PA, USA. [238]University of Melbourne Centre for Cancer Research, Melbourne, Victoria, Australia. [239]Syntekabio, Daejon, South Korea. [240]AbbVie, North Chicago, IL, USA. [241]Genomics Research Program, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [242]Department of Pediatric Immunology, Hematology and Oncology, University Hospital, Heidelberg, Germany. [243]Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM), Heidelberg, Germany. [244]Seven Bridges, Charlestown, MA, USA. [245]Health Sciences Department of Biomedical Informatics, University of California San Diego, La Jolla, CA, USA. [246]Functional and Structural Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [247]Leidos Biomedical Research, McLean, VA, USA. [248]CSRA Incorporated, Fairfax, VA, USA. [249]Department of Internal Medicine, Stanford University, Stanford, CA, USA. [250]Clinical Bioinformatics, Swiss Institute of Bioinformatics, Geneva, Switzerland. [251]Institute for Pathology and Molecular Pathology, University Hospital Zurich, Zurich, Switzerland. [252]Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland. [253]MIT Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. [254]Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland. [255]Office of Cancer Genomics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [256]Computer Network Information Center, Chinese Academy of Sciences, Beijing, China. [257]Geneplus-Shenzhen, Shenzhen, China. [258]Dana-Farber/Boston Children's Cancer and Blood Disorders Center, Boston, MA, USA. [259]Department of Pediatrics, Harvard Medical School, Boston, MA, USA. [260]Technical University of Denmark, Lyngby, Denmark. [261]University of Copenhagen, Copenhagen, Denmark. [262]Department for BioMedical Research, University of Bern, Bern, Switzerland. [263]Department of Medical Oncology, Inselspital, University Hospital and University of Bern, Bern, Switzerland. [264]Graduate School for Cellular and Biomedical Sciences, University of Bern, Bern, Switzerland. [265]Department of Genitourinary Medical Oncology - Research, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [266]Department of Urology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [267]Korea Advanced Institute of Science and Technology, Daejeon, South Korea. [268]Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden. [269]Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. [270]University of Milano Bicocca, Monza, Italy. [271]Sir Peter MacCallum Department of Oncology, Peter MacCallum Cancer Centre, University of Melbourne, Melbourne, Victoria, Australia. [272]Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center, Houston, TX, USA. [273]Health Data Science Unit, University Clinics, Heidelberg, Germany. [274]Department for Biomedical Research, University of Bern, Bern, Switzerland. [275]Research Core Center, National Cancer Centre Korea, Goyang-si, South Korea. [276]Institute of Computer Science, Polish Academy of Sciences, Warsawa, Poland. [277]Harvard University, Cambridge, MA, USA. [278]Memorial Sloan Kettering Cancer Center, New York, NY, USA. [279]Department of Information Technology, Ghent University, Ghent, Belgium. [280]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium. [281]Yale School of Medicine, Yale University, New Haven, CT, USA. [282]Division of Hematology-Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea. [283]Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University School of Medicine, Seoul, South Korea. [284]Cheonan Industry-Academic Collaboration Foundation, Sangmyung University, Cheonan, South Korea. [285]Spanish National Cancer Research Centre, Madrid, Spain. [286]Department of Computer Science, Princeton University, Princeton, NJ, USA. [287]Bern Center for Precision Medicine, University Hospital of Bern, University of Bern, Bern, Switzerland. [288]Englander Institute for Precision Medicine, Weill Cornell Medicine and New York Presbyterian Hospital, New York, NY, USA. [289]Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA. [290]Pathology and Laboratory, Weill Cornell Medical College, New York, NY, USA. [291]cBio Center, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA. [292]Department of Cell Biology, Harvard Medical School, Boston, MA, USA. [293]cBio Center, Dana-Farber Cancer Institute, Boston, MA, USA. [294]CREST, Japan Science and Technology Agency, Tokyo, Japan. [295]Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Bunkyo-ku, Tokyo, Japan. [296]Laboratory for Medical Science Mathematics, Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo, Japan. [297]Science for Life Laboratory, Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden. [298]Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia. [299]Genetics & Genome Biology Program, SickKids Research Institute, The Hospital for Sick Children, Toronto, Ontario, Canada. [300]Department of Information Technology, Ghent University, Interuniversitair Micro-Electronica Centrum (IMEC), Ghent, Belgium. [301]Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. [302]Oregon Health & Sciences University, Portland, OR, USA. [303]Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Shatin, Hong Kong, China. [304]The University of Texas Health Science Center at Houston, Houston, TX, USA. [305]Department of Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, OH, USA.

[306]The Ohio State University Comprehensive Cancer Center (OSUCCC – James), Columbus, OH, USA. [307]The University of Texas School of Biomedical Informatics (SBMI) at Houston, Houston, TX, USA. [308]Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. [309]Physics Division, Optimization and Systems Biology Lab, Massachusetts General Hospital, Boston, MA, USA. [310]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan. [311]Bioinformatics Group, Department of Computer Science, University of Leipzig, Leipzig, Germany. [312]Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig, Germany. [313]Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [314]Computational Biology, Leibniz Institute on Aging - Fritz Lipmann Institute (FLI), Jena, Germany. [315]Transcriptome Bioinformatics, LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany. [316]Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI, USA. [317]Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain. [318]Research Center for Advanced Science and Technology, The University of Tokyo, Minato-ku, Tokyo, Japan. [319]Van Andel Research Institute, Grand Rapids, MI, USA. [320]Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [321]Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [322]The Hebrew University Faculty of Medicine, Jerusalem, Israel. [323]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. [324]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [325]McKusick-Nathans Institute of Genetic Medicine, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [326]Foundation Medicine, Cambridge, MA, USA. [327]Department of Biochemistry, Microbiology and Immunology, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada. [328]Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK. [329]University of Cambridge, Cambridge, UK. [330]Brandeis University, Waltham, MA, USA. [331]Hopp Children's Cancer Center (KiTZ), Heidelberg, Germany. [332]Pediatric Glioma Research Group, German Cancer Research Center (DKFZ), Heidelberg, Germany. [333]A. A. Kharkevich Institute of Information Transmission Problems, Moscow, Russia. [334]Oncology and Immunology, Dmitry Rogachev National Research Center of Pediatric Hematology, Moscow, Russia. [335]Skolkovo Institute of Science and Technology, Moscow, Russia. [336]Center for Medical Innovation, Seoul National University Hospital, Seoul, South Korea. [337]Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea. [338]Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. [339]School of Medicine/School of Mathematics and Statistics, University of St Andrews, St Andrews, UK. [340]Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [341]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [342]Cancer Research Institute, Beth Israel Deaconess Medical Center, Boston, MA, USA. [343]Ben May Department for Cancer Research, Department of Human Genetics, The University of Chicago, Chicago, IL, USA. [344]Tri-Institutional PhD Program in Computational Biology and Medicine, Weill Cornell Medicine, New York, NY, USA. [345]Department of Bioengineering, and Department of Cellular and Molecular Medicine, Moores Cancer Center, University of California, San Diego, La Jolla, CA, USA. [346]Centre for Computational Biology, Duke-NUS Medical School, Singapore, Singapore. [347]Department of Computer Science, University of Helsinki, Helsinki, Finland. [348]Institute of Biotechnology, University of Helsinki, Helsinki, Finland. [349]Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland. [350]Programme in Cancer & Stem Cell Biology, Centre for Computational Biology, Duke-NUS Medical School, Singapore, Singapore. [351]Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Cambridge, UK. [352]Department of Statistics, Columbia University, New York, NY, USA. [353]Duke-NUS Medical School, Singapore, Singapore. [354]School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China. [355]The Kinghorn Cancer Centre, Cancer Division, Garvan Institute of Medical Research, University of New South Wales, Sydney, New South Wales, Australia. [356]MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Edinburgh, UK. [357]Bioinformatics Group, Division of Molecular Biology, Department of Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia. [358]Department of Bioinformatics, Division of Cancer Genomics, National Cancer Center Research Institute, National Cancer Center, Tokyo, Japan. [359]University of Glasgow, Glasgow, UK. [360]Academic Department of Medical Genetics, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK. [361]MRC Cancer Unit, University of Cambridge, Cambridge, UK. [362]The University of Cambridge School of Clinical Medicine, Cambridge, UK. [363]MRC-University of Glasgow Centre for Virus Research, Glasgow, UK. [364]Wolfson Wohl Cancer Research Centre, Institute of Cancer Sciences, University of Glasgow, Bearsden, UK. [365]School of Computing Science, University of Glasgow, Glasgow, UK. [366]South Western Sydney Clinical School, Faculty of Medicine, University of New South Wales, Liverpool, New South Wales, Australia. [367]West of Scotland Pancreatic Unit, Glasgow Royal Infirmary, Glasgow, UK. [368]University of Melbourne Centre for Cancer Research, Melbourne, Victoria, Australia. [369]Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR, USA. [370]Department of Surgery, University of Melbourne, Parkville, Victoria, Australia. [371]The Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia. [372]Walter + Eliza Hall Institute, Parkville, Victoria, Australia. [373]University of Cologne, Cologne, Germany. [374]The Edward S. Rogers Sr Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada. [375]University of Ljubljana, Ljubljana, Slovenia. [376]Department of Public Health Sciences, The University of Chicago, Chicago, IL, USA. [377]Research Institute, NorthShore University HealthSystem, Evanston, IL, USA. [378]Department of Statistics, University of California Santa Cruz, Santa Cruz, CA, USA. [379]Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [380]University of Toronto, Toronto, Ontario, Canada. [381]Department of Computer Science, Carleton College, Northfield, MN, USA. [382]Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR, USA. [383]Center for Psychiatric Genetics, NorthShore University HealthSystem, Evanston, IL, USA. [384]Argmix Consulting, North Vancouver, British Columbia, Canada. [385]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [386]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [387]The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [388]Molecular and Medical Genetics, Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA. [389]Department of Health Sciences, Faculty of Medical Sciences, Kyushu University, Fukuoka, Japan. [390]Baylor College of Medicine, Houston, TX, USA. [391]Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA. [392]Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Hamburg, Germany. [393]University Medical Center Hamburg-Eppendorf, Bioinformatics Core, Hamburg, Germany. [394]Earlham Institute, Norwich, UK. [395]Norwich Medical School, University of East Anglia, Norwich, UK. [396]The Institute of Cancer Research, London, UK. [397]University of East Anglia, Norwich, UK. [398]German Center for Infection Research (DZIF), Partner Site Hamburg-Borstel-Lübeck-Riems, Hamburg, Germany. [399]Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [400]Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [401]Victorian Institute of Forensic Medicine, Southbank, Victoria, Australia. [402]Peter MacCallum Cancer Centre, University of Melbourne, Melbourne, Victoria, Australia. [403]University of Pennsylvania, Philadelphia, PA, USA. [404]Centre for Cancer Research, The Westmead Institute for Medical Research, Sydney, New South Wales, Australia. [405]Department of Gynaecological Oncology, Westmead Hospital, Sydney, New South Wales, Australia. [406]Genetics and Molecular Pathology, SA Pathology, Adelaide, South Australia, Australia. [407]Centre for Cancer Research, The Westmead Institute for Medical Research, The University of Sydney, Sydney, New South Wales, Australia. [408]Department of Gynaecological Oncology, Westmead Hospital, Sydney, New South Wales, Australia. [409]Garvan Institute of Medical Research, Darlinghurst, New South Wales, Australia. [410]Department of Clinical Pathology, University of Melbourne, Melbourne, Victoria, Australia. [411]Centre for Cancer Research, The Westmead Institute for Medical Research, The University of Sydney, Sydney, New South Wales, Australia. [412]Department of Gynaecological Oncology, Westmead Hospital, Sydney, New South Wales, Australia. [413]Westmead Clinical School, The Westmead Institute for Medical Research, Sydney, New South Wales, Australia. [414]Department of Surgery, Pancreas Institute, University and Hospital Trust of Verona, Verona, Italy. [415]Department of Surgery, Princess Alexandra Hospital, Brisbane, Queensland, Australia. [416]Surgical Oncology Group, Diamantina Institute, The University of Queensland, Brisbane, Queensland, Australia. [417]Department of Diagnostics and Public Health, University and Hospital Trust of Verona, Verona, Italy. [418]ARC-Net Centre for Applied Research on Cancer, University and Hospital Trust of Verona, Verona, Italy. [419]Illawarra Shoalhaven Local Health District L3 Illawarra Cancer Care Centre, Wollongong Hospital, Wollongong, New South Wales, Australia. [420]Department of Pathology, University of Sydney, Sydney, New South Wales, Australia. [421]School of Biological Sciences, The University of Auckland, Auckland, New Zealand. [422]Department of Pathology and Diagnostics, University and Hospital Trust of Verona, Verona, Italy. [423]Department of Medicine, Section of Endocrinology, University and Hospital Trust of Verona, Verona, Italy. [424]Department of Pathology, Queen Elizabeth University Hospital, Glasgow, UK. [425]Department of Medical Oncology, Beatson West of Scotland Cancer Centre, Glasgow, UK. [426]Academic Unit of Surgery, School of Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow Royal Infirmary, Glasgow, UK. [427]Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown, New South Wales, Australia. [428]Discipline of Surgery, Western Sydney University, Penrith, New South Wales, Australia. [429]Institute of Cancer Sciences, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. [430]The Kinghorn Cancer Centre, Cancer Division, Garvan Institute of Medical Research, University of New South Wales, Sydney, New South Wales, Australia. [431]School of Environmental and Life Sciences, Faculty of Science, The University of Newcastle, Ourimbah, New South Wales, Australia. [432]Eastern Clinical School, Monash University, Melbourne, Victoria, Australia. [433]Epworth HealthCare, Richmond, Victoria, Australia. [434]Olivia Newton-John Cancer Research Institute, La Trobe University, Heidelberg, Victoria, Australia. [435]Melanoma Institute Australia, The University of Sydney, Wollstonecraft, New South Wales, Australia. [436]Children's Hospital at Westmead, The University of Sydney, Sydney, New South Wales, Australia. [437]Melanoma Institute Australia, The University of Sydney, Sydney, New South Wales, Australia. [438]Australian Institute of Tropical Health and Medicine, James Cook University, Douglas, Queensland, Australia. [439]Bioplatforms Australia, North Ryde, New South Wales, Australia. [440]Melanoma Institute Australia, Macquarie University, Wollstonecraft, New South Wales, Australia. [441]Children's Medical Research Institute, Sydney, New South Wales, Australia. [442]Melanoma Institute Australia, The University of Sydney, Wollstonecraft, New South Wales, Australia. [443]Centre for Cancer Research, The Westmead Millennium Institute for Medical Research, University of Sydney, Westmead Hospital, Sydney, New South Wales, Australia. [444]Translational Cancer Research Centre, The University of Sydney at the Westmead Institute, Sydney, New South Wales, Australia. [445]Discipline of Pathology, Sydney Medical School, The University of Sydney, Sydney, New South Wales, Australia. [446]School of Mathematics and Statistics, The University of Sydney, Sydney, New South Wales, Australia. [447]Melanoma Institute Australia, The University of Sydney, Wollstonecraft, New South Wales, Australia. [448]Royal Prince Alfred Hospital, Sydney, New South Wales, Australia. [449]Diagnostic Development, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [450]Ontario

Tumour Bank, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [451]PanCuRx Translational Research Initiative, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [452]BioSpecimen Sciences Program, University Health Network, Toronto, Ontario, Canada. [453]Hepatobiliary/Pancreatic Surgical Oncology Program, University Health Network, Toronto, Ontario, Canada. [454]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [455]Division of Medical Oncology, Princess Margaret Cancer Centre, Toronto, Ontario, Canada. [456]University of Nebraska Medical Center, Omaha, NE, USA. [457]BioSpecimen Sciences Program, University Health Network, Toronto, Ontario, Canada. [458]Transformative Pathology, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [459]University Health Network, Princess Margaret Cancer Centre, Toronto, Ontario, Canada. [460]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA. [461]BioSpecimen Sciences, Laboratory Medicine (Toronto), Medical Biophysics, PanCuRX, Toronto, Ontario, Canada. [462]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. [463]Department of Pathology, Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [464]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. [465]Department of Biochemistry and Molecular Medicine, University California at Davis, Sacramento, CA, USA. [466]Human Longevity, San Diego, CA, USA. [467]Department of Surgical Oncology, Princess Margaret Cancer Centre, Toronto, Ontario, Canada. [468]Genome Informatics Program, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [469]STTARR Innovation Facility, Princess Margaret Cancer Centre, Toronto, Ontario, Canada. [470]Department of Pathology, Toronto General Hospital, Toronto, Ontario, Canada. [471]CRUK Manchester Institute and Centre, Manchester, UK. [472]Department of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada. [473]Manchester Cancer Research Centre, Cancer Division, FBMH, University of Manchester, Manchester, UK. [474]Radiation Medicine Program, Princess Margaret Cancer Centre, Toronto, Ontario, Canada. [475]Hefei University of Technology, Anhui, China. [476]State Key Laboratory of Cancer Biology and Xijing Hospital of Digestive Diseases, Fourth Military Medical University, Shaanxi, China. [477]Fourth Military Medical University, Shaanxi, China. [478]Laboratory of Molecular Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital & Institute, Beijing, China. [479]Department of Surgery, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China. [480]Leeds Institute of Medical Research, University of Leeds, St James's University Hospital, Leeds, UK. [481]Canadian Center for Computational Genomics, McGill University, Montreal, Quebec, Canada. [482]Department of Human Genetics, McGill University, Montreal, Quebec, Canada. [483]International Agency for Research on Cancer, Lyon, France. [484]McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada. [485]St James Institute of Oncology, University of Leeds, St James's University Hospital, Leeds, UK. [486]Institute of Mathematics and Computer Science, University of Latvia, Riga, Latvia. [487]Centre National de Génotypage, CEA - Institute de Génomique, Evry, France. [488]Department of Oncology, Gil Medical Center, Gachon University, Incheon, South Korea. [489]Department of Molecular Oncology, BC Cancer Agency, Vancouver, British Columbia, Canada. [490]Los Alamos National Laboratory, Los Alamos, NM, USA. [491]Department of Genetics, Institute for Cancer Research, Oslo University Hospital, The Norwegian Radium Hospital, Oslo, Norway. [492]Lund University, Lund, Sweden. [493]Translational Research Lab, Centre Léon Bérard, Lyon, France. [494]Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University, Nijmegen, The Netherlands. [495]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [496]Department of Molecular Pathology, The Netherlands Cancer Institute, Amsterdam, The Netherlands. [497]Li Ka Shing Centre, Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK. [498]Department of Oncology, University of Cambridge, Cambridge, UK. [499]Breast Cancer Translational Research Laboratory J. C. Heuson, Institut Jules Bordet, Brussels, Belgium. [500]Laboratory for Translational Breast Cancer Research, Department of Oncology, KU Leuven, Leuven, Belgium. [501]Translational Cancer Research Unit, GZA Hospitals St-Augustinus, Center for Oncological Research, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium. [502]Department of Gynecology & Obstetrics and Department of Clinical Sciences, Skåne University Hospital, Lund University, Lund, Sweden. [503]Icelandic Cancer Registry, Icelandic Cancer Society, Reykjavik, Iceland. [504]Department of Medical Oncology, Josephine Nefkens Institute and Cancer Genomics Centre, Erasmus Medical Center, Rotterdam, The Netherlands. [505]National Genotyping Center, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan. [506]Department of Pathology, Oslo University Hospital Ulleval, Oslo, Norway. [507]Faculty of Medicine and Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [508]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [509]Department of Pathology, Skåne University Hospital, Lund University, Lund, Sweden. [510]Department of Pathology, Academic Medical Center, Amsterdam, The Netherlands. [511]Department of Pathology, College of Medicine, Hanyang University, Seoul, South Korea. [512]Department of Pathology, Asan Medical Center, College of Medicine, Ulsan University, Songpa-gu, Seoul, South Korea. [513]The Netherlands Cancer Institute, Amsterdam, The Netherlands. [514]Department of Surgery, Dana-Farber Cancer Institute, Brigham and Women's Hospital, Boston, MA, USA. [515]Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [516]Department of Clinical Science, University of Bergen, Bergen, Norway. [517]Morgan Welch Inflammatory Breast Cancer Research Program and Clinic, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [518]The University of Queensland Centre for Clinical Research, The Royal Brisbane & Women's Hospital, Herston, Queensland, Australia. [519]Department of Pathology, Institut Jules Bordet, Brussels, Belgium. [520]Institute for Bioengineering and Biopharmaceutical Research (IBBR), Hanyang University, Seoul, South Korea. [521]University of Oslo, Oslo, Norway. [522]Institut Bergonié, Bordeaux, France. [523]Department of Research Oncology, Guy's Hospital, King's Health Partners AHSC, King's College London School of Medicine, London, UK. [524]University Hospital of Minjoz, INSERM UMR 1098, Besançon, France. [525]Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge, UK. [526]East of Scotland Breast Service, Ninewells Hospital, Aberdeen, UK. [527]Oncologie Sénologie, ICM Institut Régional du Cancer, Montpellier, France. [528]Department of Radiation Oncology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [529]University of Iceland, Reykjavik, Iceland. [530]Dundee Cancer Centre, Ninewells Hospital, Dundee, UK. [531]Institut Curie, INSERM Unit 830, Paris, France. [532]Department of Laboratory Medicine, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [533]Department of General Surgery, Singapore General Hospital, Singapore, Singapore. [534]INCa-Synergie, Centre Léon Bérard, Universite Lyon, Lyon, France. [535]Giovanni Paolo II/I.R.C.C.S. Cancer Institute, Bari, Italy. [536]Department of Biopathology, Centre Léon Bérard, Lyon, France. [537]Université Claude Bernard Lyon 1, Villeurbanne, France. [538]NCCS-VARI Translational Research Laboratory, National Cancer Centre Singapore, Singapore, Singapore. [539]Department of Pathology, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands. [540]Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Amsterdam, The Netherlands. [541]Institute of Human Genetics, Christian-Albrechts-University, Kiel, Germany. [542]Institute of Human Genetics, University of Ulm, Ulm, Germany. [543]University Hospital of Ulm, Ulm, Germany. [544]Hematopathology Section, Institute of Pathology, Christian-Albrechts-University, Kiel, Germany. [545]Department of Human Genetics, Hannover Medical School, Hannover, Germany. [546]Department of Pediatric Oncology, Hematology and Clinical Immunology, Heinrich-Heine-University, Düsseldorf, Germany. [547]Department of Internal Medicine/Hematology, Friedrich-Ebert-Hospital, Neumünster, Germany. [548]Pediatric Hematology and Oncology, University Hospital Muenster, Muenster, Germany. [549]Department of Pediatrics, University Hospital Schleswig-Holstein, Kiel, Germany. [550]Department of Medicine II, University of Würzburg, Würzburg, Germany. [551]Senckenberg Institute of Pathology, University of Frankfurt Medical School, Frankfurt, Germany. [552]Institute of Pathology, Charité–University Medicine Berlin, Berlin, Germany. [553]Department for Internal Medicine II, University Hospital Schleswig-Holstein, Kiel, Germany. [554]Institute for Medical Informatics Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. [555]Department of Hematology and Oncology, Georg-Augusts-University of Göttingen, Göttingen, Germany. [556]Institute of Cell Biology (Cancer Research), University of Duisburg-Essen, Essen, Germany. [557]MVZ Department of Oncology, PraxisClinic am Johannisplatz, Leipzig, Germany. [558]Institute of Pathology, Ulm University and University Hospital of Ulm, Ulm, Germany. [559]Department of Pathology, Robert-Bosch-Hospital, Stuttgart, Germany. [560]Pediatric Hematology and Oncology, University Hospital Giessen, Giessen, Germany. [561]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany. [562]Institute of Pathology, University of Wuerzburg, Wuerzburg, Germany. [563]Department of General Internal Medicine, University Kiel, Kiel, Germany. [564]Clinic for Hematology and Oncology, St-Antonius-Hospital, Eschweiler, Germany. [565]Department for Internal Medicine III, University of Ulm and University Hospital of Ulm, Ulm, Germany. [566]Neuroblastoma Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany. [567]Department of Pediatric Oncology and Hematology, University of Cologne, Cologne, Germany. [568]University of Düsseldorf, Düsseldorf, Germany. [569]Department of Vertebrate Genomics/Otto Warburg Laboratory Gene Regulation and Systems Biology of Cancer, Max Planck Institute for Molecular Genetics, Berlin, Germany. [570]St Jude Children's Research Hospital, Memphis, TN, USA. [571]Heidelberg University Hospital, Heidelberg, Germany. [572]Genomics and Proteomics Core Facility High Throughput Sequencing Unit, German Cancer Research Center (DKFZ), Heidelberg, Germany. [573]Epigenomics and Cancer Risk Factors, German Cancer Research Center (DKFZ), Heidelberg, Germany. [574]University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [575]Martini-Clinic, Prostate Cancer Center, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [576]Institute of Pathology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [577]Division of Cancer Genome Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. [578]National Institute of Biomedical Genomics, Kalyani, India. [579]Advanced Centre for Treatment Research & Education in Cancer, Tata Memorial Centre, Navi Mumbai, India. [580]Department of Pathology, General Hospital of Treviso, Department of Medicine, University of Padua, Treviso, Italy. [581]Department of Medicine (DIMED), Surgical Pathology Unit, University of Padua, Padua, Italy. [582]Department of Hepatobiliary and Pancreatic Oncology, Hepatobiliary and Pancreatic Surgery Division, Division of Pathology and Clinical Laboratories, National Cancer Center Hospital, Chuo-ku, Tokyo, Japan. [583]Department of Pathology, Keio University School of Medicine, Tokyo, Japan. [584]Department of Hepatobiliary and Pancreatic Oncology, National Cancer Center Hospital, Tokyo, Japan. [585]Department of Pathology, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Tokyo, Japan. [586]Preventive Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [587]Gastric Surgery Division, Division of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo, Japan. [588]Department of Gastroenterology and Hepatology, Yokohama City University Graduate School of Medicine, Kanagawa, Japan. [589]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, University of Tokyo, Tokyo, Japan. [590]Department of Cancer Genome Informatics, Graduate School of Medicine, Osaka University, Osaka, Japan. [591]Hiroshima University, Hiroshima, Japan. [592]Tokyo Women's Medical University, Tokyo, Japan. [593]Osaka International Cancer Center, Osaka, Japan. [594]Wakayama Medical University, Wakayama, Japan. [595]Hokkaido University, Sapporo, Japan. [596]Division of Medical Oncology, National Cancer Centre, Singapore, Singapore. [597]Cholangiocarcinoma Screening and Care Program and Liver Fluke and Cholangiocarcinoma Research Centre, Faculty of Medicine, Khon Kaen University,

Khon Kaen, Thailand. [598]Lymphoma Genomic Translational Research Laboratory, National Cancer Centre, Singapore, Singapore. [599]Center of Digestive Diseases and Liver Transplantation, Fundeni Clinical Institute, Bucharest, Romania. [600]Division of Hepatobiliary and Pancreatic Surgery, Department of Surgery, School of Medicine, Keimyung University Dongsan Medical Center, Daegu, South Korea. [601]Pathology, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain. [602]Hematology, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain. [603]Department of Biochemistry and Molecular Biology, Faculty of Medicine, University Institute of Oncology-IUOPA, Oviedo, Spain. [604]Anatomia Patológica, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain. [605]Spanish Ministry of Science and Innovation, Madrid, Spain. [606]Royal National Orthopaedic Hospital (Bolsover), London, UK. [607]Department of Pathology, Oslo University Hospital, The Norwegian Radium Hospital, Oslo, Norway. [608]Institute of Clinical Medicine and Institute of Oral Biology, University of Oslo, Oslo, Norway. [609]Research Department of Pathology, University College London Cancer Institute, London, UK. [610]East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [611]Royal National Orthopaedic Hospital (Stanmore), London, UK. [612]Division of Orthopaedic Surgery, Oslo University Hospital, Oslo, Norway. [613]Department of Pathology (Research), University College London Cancer Institute, London, UK. [614]Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [615]University of Pavia, Pavia, Italy. [616]Karolinska Institute, Stockholm, Sweden. [617]Wellcome Sanger Institute, Hinxton, UK. [618]University of Oxford, Oxford, UK. [619]Salford Royal NHS Foundation Trust, Salford, UK. [620]Gloucester Royal Hospital, Gloucester, UK. [621]Royal Stoke University Hospital, Stoke-on-Trent, UK. [622]St Thomas's Hospital, London, UK. [623]Imperial College NHS Trust, Imperial College London, London, UK. [624]Department of Histopathology, Salford Royal NHS Foundation Trust, Salford, UK. [625]Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK. [626]Edinburgh Royal Infirmary, Edinburgh, UK. [627]Barking Havering and Redbridge University Hospitals NHS Trust, Romford, UK. [628]King's College London and Guy's and St Thomas' NHS Foundation Trust, London, UK. [629]Cambridge Oesophagogastric Centre, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [630]Nottingham University Hospitals NHS Trust, Nottingham, UK. [631]St Luke's Cancer Centre, Royal Surrey County Hospital NHS Foundation Trust, Guildford, UK. [632]University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [633]Norfolk and Norwich University Hospital NHS Trust, Norwich, UK. [634]University Hospitals Coventry and Warwickshire NHS Trust, Coventry, UK. [635]University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK. [636]Centre for Cancer Research and Cell Biology, Queen's University, Belfast, UK. [637]School of Cancer Sciences, Faculty of Medicine, University of Southampton, Southampton, UK. [638]Wythenshawe Hospital, Manchester, UK. [639]Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [640]Royal Marsden NHS Foundation Trust, London and Sutton, London, UK. [641]University Hospital Southampton NHS Foundation Trust, Southampton, UK. [642]HCA Laboratories, London, UK. [643]University of Liverpool, Liverpool, UK. [644]Academic Urology Group, Department of Surgery, University of Cambridge, Cambridge, UK. [645]University of Oxford, Oxford, Oxford, UK. [646]Department of Urology, James Buchanan Brady Urological Institute, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [647]Second Military Medical University, Shanghai, China. [648]Department of Surgery and Cancer, Imperial College London, London, UK. [649]The Chinese University of Hong Kong, Shatin, Hong Kong, China. [650]Nuffield Department of Surgical Sciences, John Radcliffe Hospital, University of Oxford, Headington, Oxford, UK. [651]Department of Histopathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [652]Department of Bioinformatics and Computational Biology and Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [653]Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA. [654]Canada's Michael Smith Genome Sciences Center, BC Cancer Agency, Vancouver, British Columbia, Canada. [655]Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [656]University Health Network, Toronto, Ontario, Canada. [657]Department of Pathology and Laboratory Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [658]Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA. [659]Research Health Analytics and Informatics, University Hospitals Cleveland Medical Center, Cleveland, OH, USA. [660]Arnie Charbonneau Cancer Institute, University of Calgary, Calgary, Alberta, Canada. [661]Department of Surgery and Department of Oncology, University of Calgary, Calgary, Alberta, Canada. [662]Buck Institute for Research on Aging, Novato, CA, USA. [663]Duke University Medical Center, Durham, NC, USA. [664]USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA. [665]The Preston Robert Tisch Brain Tumor Center, Duke University Medical Center, Durham, NC, USA. [666]Department of Dermatology and Department of Pathology, Yale University, New Haven, CT, USA. [667]Fox Chase Cancer Center, Philadelphia, PA, USA. [668]Department of Surgery, Division of Thoracic Surgery, The Johns Hopkins University School of Medicine, Baltimore, MD, USA. [669]University of Michigan Comprehensive Cancer Center, Ann Arbor, MI, USA. [670]University of Alabama at Birmingham, Birmingham, AL, USA. [671]Division of Anatomic Pathology, Mayo Clinic, Rochester, MN, USA. [672]Division of Experimental Pathology, Mayo Clinic, Rochester, MN, USA. [673]Department of Oncology, The Johns Hopkins School of Medicine, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, MD, USA. [674]International Genomics Consortium, Phoenix, AZ, USA. [675]Department of Pediatrics and Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [676]Department of Pathology, UPMC Shadyside, Pittsburgh, PA, USA. [677]Center for Cancer Genomics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [678]Department of Neuro-Oncology, Istituto Neurologico Besta, Milan, Italy. [679]University of Queensland Thoracic Research Centre, The Prince Charles Hospital, Brisbane, Queensland, Australia. [680]Department of Neurosurgery, University of Florida, Gainesville, FL, USA. [681]Center for Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [682]Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [683]Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [684]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA. [685]Sylvester Comprehensive Cancer Center, University of Miami, Miami, FL, USA. [686]Department of Internal Medicine, Division of Medical Oncology, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [687]University of Tennessee Health Science Center for Cancer Research, Memphis, TN, USA. [688]Centre for Translational and Applied Genomics, British Columbia Cancer Agency, Vancouver, British Columbia, Canada. [689]Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX, USA. [690]Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, USA. [691]Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [692]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada. [693]Indivumed, Hamburg, Germany. [694]Division of Hepatobiliary and Pancreatic Surgery, Department of Surgery, School of Medicine, Keimyung University Dong-san Medical Center, Daegu, South Korea. [695]Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [696]Department of Surgery, School of Medicine and Health Science, The George Washington University, Washington, DC, USA. [697]Endocrine Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [698]National Cancer Center, Gyeonggi, South Korea. [699]ILSbio, LLC Biobank, Chestertown, MD, USA. [700]Gynecologic Oncology, NYU Laura and Isaac Perlmutter Cancer Center, New York University, New York, NY, USA. [701]Division of Oncology, Stem Cell Biology Section, Washington University School of Medicine, St Louis, MO, USA. [702]Urologic Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [703]Institute for Systems Biology, Seattle, WA, USA. [704]Center for Personalized Medicine, Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Los Angeles, CA, USA. [705]Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH, USA. [706]Department of Surgery, Duke University, Durham, NC, USA. [707]Department of Obstetrics, Gynecology and Reproductive Services, University of California San Francisco, San Francisco, CA, USA. [708]Department of Neurology and Department of Neurosurgery, Henry Ford Hospital, Detroit, MI, USA. [709]Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA. [710]Department of Pathology, Roswell Park Cancer Institute, Buffalo, NY, USA. [711]Department of Obstetrics and Gynecology, Division of Gynecologic Oncology, Washington University School of Medicine, St Louis, MO, USA. [712]Department of Palliative, Rehabilitation and Integrative Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [713]Penrose St Francis Health Services, Colorado Springs, CO, USA. [714]The University of Chicago, Chicago, IL, USA. [715]Department of Neurology, Mayo Clinic, Rochester, MN, USA. [716]Center for Liver Cancer, Research Institute and Hospital, National Cancer Center, Gyeonggi, South Korea. [717]Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [718]NYU Langone Medical Center, New York, NY, USA. [719]Department of Hematology and Medical Oncology, Cleveland Clinic, Cleveland, OH, USA. [720]Department of Genetics, Department of Pathology and Laboratory Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [721]Helen F. Graham Cancer Center at Christiana Care Health Systems, Newark, DE, USA. [722]Cureline, South San Francisco, CA, USA. [723]Department of Obstetrics and Gynecology, Medical College of Wisconsin, Milwaukee, WI, USA. [724]Hematology and Medical Oncology, Winship Cancer Institute of Emory University, Atlanta, GA, USA. [725]Vanderbilt Ingram Cancer Center, Vanderbilt University, Nashville, TN, USA. [726]Ohio State University College of Medicine and Arthur G. James Comprehensive Cancer Center, Columbus, OH, USA. [727]Research Computing Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [728]Analytical Biological Services, Wilmington, DE, USA. [729]Department of Dermatology, University Hospital Essen, Westdeutsches Tumorzentrum and German Cancer Consortium, Essen, Germany. [730]University of Pittsburgh, Pittsburgh, PA, USA. [731]Murtha Cancer Center, Walter Reed National Military Medical Center, Bethesda, MD, USA. [732]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [733]Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [734]Department of Gynecologic Oncology and Reproductive Medicine, and Center for RNA Interference and Non-Coding RNA, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [735]Department of Urology, Mayo Clinic, Rochester, MN, USA. [736]Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [737]Department of Neurosurgery, Department of Hematology and Department of Medical Oncology, Winship Cancer Institute and School of Medicine, Emory University, Atlanta, GA, USA. [738]Georgia Regents University Cancer Center, Augusta, GA, USA. [739]Thoracic Oncology Laboratory, Mayo Clinic, Rochester, MN, USA. [740]Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH, USA. [741]Department of Obstetrics & Gynecology, Division of Gynecologic Oncology, Mayo Clinic, Rochester, MN, USA. [742]International Institute for Molecular Oncology, Poznań, Poland. [743]Poznan University of Medical Sciences, Poznań, Poland. [744]Edison Family Center for Genome Sciences and Systems Biology, Washington University, St Louis, MO, USA. [745]These authors jointly supervised this work: Peter J. Campbell, Gad Getz, Jan O. Korbel, Joshua M. Stuart, Lincoln D. Stein. *e-mail: pc8@sanger.ac.uk; gadgetz@broadinstitute.org; korbel@embl.de; jstuart@ucsc.edu; lincoln.stein@gmail.com

Methods

Samples

We compiled an inventory of matched tumour–normal whole-cancer genomes in the ICGC Data Coordinating Centre. Most samples came from treatment-naive, primary cancers, although a small number of donors had multiple samples of primary, metastatic and/or recurrent tumours. Our inclusion criteria were: (1) matched tumour and normal specimen pair; (2) a minimal set of clinical fields; and (3) characterization of tumour and normal whole genomes using Illumina HiSeq paired-end sequencing reads.

We collected genome data from 2,834 donors, representing all ICGC and TCGA donors that met these criteria at the time of the final data freeze in autumn 2014 (Extended Data Table 1). After quality assurance (Supplementary Methods 2.5), data from 176 donors were excluded as unusable, 75 had minor issues that could affect some analyses (grey-listed donors) and 2,583 had data of optimal quality (white-listed donors) (Supplementary Table 1). Across the 2,658 white- and grey-listed donors, whole-genome sequences were available from 2,605 primary tumours and 173 metastases or local recurrences. Matching normal samples were obtained from blood (2,064 donors), tissue adjacent to the primary tumour (87 donors) or from distant sites (507 donors). Whole-genome sequencing data were available for tumour and normal DNA for the entire cohort. The mean read coverage was 39× for normal samples, whereas tumours had a bimodal coverage distribution with modes at 38× and 60× (Supplementary Fig. 1). The majority of specimens (65.3%) were sequenced using 101-bp paired-end reads. An additional 28% were sequenced with 100-bp paired-end reads. Of the remaining specimens, 4.7% were sequenced with read lengths longer than 101 bp, and 1.9% with read lengths shorter than 100 bp. The distribution of read lengths by tumour cohort is shown in Supplementary Fig. 11. Median read length for whole-genome sequencing paired-end reads was 101 bp (mean = 106.2, s.d. = 16.7; minimum–maximum = 50–151). RNA-sequencing data were collected and re-analysed centrally for 1,222 donors, including 1,178 primary tumours, 67 metastases or local recurrences and 153 matched normal tissue samples adjacent to the primary tumour.

Demographically, the cohort included 1,469 men (55%) and 1,189 women (45%), with a mean age of 56 years (range, 1–90 years) (Supplementary Table 1). Using population ancestry-differentiated single nucleotide polymorphisms, the ancestry distribution was heavily weighted towards donors of European descent (77% of total) followed by East Asians (16%), as expected for large contributions from European, North American and Australian projects (Supplementary Table 1).

We consolidated histopathology descriptions of the tumour samples, using the ICD-0-3 tumour site controlled vocabulary[89]. Overall, the PCAWG dataset comprises 38 distinct tumour types (Extended Data Table 1 and Supplementary Table 1). Although the most common tumour types are included in the dataset, their distribution does not match the relative population incidences, largely owing to differences among contributing ICGC/TCGA groups in the numbers of sequenced samples.

Uniform processing and somatic variant calling

To generate a consistent set of somatic mutation calls that could be used for cross-tumour analyses, we analysed all 6,835 samples using a uniform set of algorithms for alignment, variant calling and quality control (Extended Data Fig. 1, Supplementary Fig. 2, Supplementary Table 3 and Supplementary Methods 2). We used the BWA-MEM algorithm[90] to align each tumour and normal sample to human reference build hs37d5 (as used in the 1000 Genomes Project[91]). Somatic mutations were identified in the aligned data using three established pipelines, which were run independently on each tumour–normal pair. Each of the three pipelines—labelled 'Sanger'[92–95], 'EMBL/DKFZ'[96,97] and 'Broad'[98–101] after the computational biology groups that created or assembled them—consisted of multiple software packages for calling somatic SNVs, small indels, CNAs and somatic SVs (with intrachromosomal SVs defined as those >100 bp). Two additional variant algorithms[102,103] were included to further improve accuracy across a broad range of clonal and subclonal mutations. We tested different merging strategies using validation data, and choses the optimal method for each variant type to generate a final consensus set of mutation calls (Supplementary Methods S2.4).

Somatic retrotransposition events, including Alu and LINE-1 insertions[72], L1-mediated transductions[73] and pseudogene formation[104], were called using a dedicated pipeline[73]. We removed these retrotransposition events from the somatic SV call-set. Mitochondrial DNA mutations were called using a published algorithm[105]. RNA-sequencing data were uniformly processed to quantify normalized gene-level expression, splicing variation and allele-specific expression, and to identify fusion transcripts, alternative promoter usage and sites of RNA editing[8].

Integration, phasing and validation of germline variant call-sets

Calls of common (≥1% frequency in PCAWG) and rare (<1%) germline variants including single-nucleotide polymorphisms, indels, SVs and mobile-element insertions (MEIs) were generated using a population-scale genetic polymorphism-detection approach[91,106]. The uniform germline data-processing workflow comprised variant identification using six different variant-calling algorithms[96,107,108] and was orchestrated using the Butler workflow system[109].

We performed call-set benchmarking, merging, variant genotyping and statistical haplotype-block phasing[91] (Supplementary Methods 3.4). Using this strategy, we identified 80.1 million germline single-nucleotide polymorphisms, 5.9 million germline indels, 1.8 million multi-allelic short (<50 bp) germline variants, as well as germline SVs ≥ 50 bp in size including 29,492 biallelic deletions and 27,254 MEIs (Supplementary Table 2). We statistically phased this germline variant set using haplotypes from the 1000 Genomes Project[91] as a reference panel, yielding an N50-phased block length of 265 kb based on haploid chromosomes from donor-matched tumour genomes. Precision estimates for germline SNVs and indels were >99% for the phased merged call-set, and sensitivity estimates ranged from 92% to 98%.

Core alignment and variant calling by cloud computing

The requirement to uniformly realign and call variants on nearly 5,800 whole genomes (tumour plus normal) presented considerable computational challenges, and raised ethical issues owing to the use of data from different jurisdictions (Extended Data Table 2). To process the data, we adopted a cloud-computing architecture[26] in which the alignment and variant calling was spread across 13 data centres on 3 continents, representing a mixture of commercial, infrastructure-as-a-service, academic cloud compute and traditional academic high-performance computer clusters (Supplementary Table 3). Together, the effort used 10 million CPU-core hours.

To generate reproducible variant calling across the 13 data centres, we built the core pipelines into Docker containers[28], in which the workflow description, required code and all associated dependencies were packaged together in stand-alone packages. These heavily tested, extensively validated workflows are available for download (Box 1).

Validation, benchmarking and merging of somatic variant calls

To evaluate the performance of each of the mutation-calling pipelines and determine an integration strategy, we performed a large-scale deep-sequencing validation experiment (Supplementary Notes 1). We selected a pilot set of 63 representative tumour–normal pairs, on which we ran the 3 core pipelines, together with a set of 10 additional somatic variant-calling pipelines contributed by members of the PCAWG SNV Calling Methods Working Group. Sufficient DNA remained for 50 of the 63 cases for validation, which was performed by hybridization of tumour and matched normal DNA to a custom RNA bait set, followed

by deep sequencing, as previously described[29]. Although performed using the same sequencing chemistry as the original whole-genome sequencing analyses, the considerably greater depth achieved in the validation experiment enabled accurate assessment of sensitivity and precision of variant calls. Variant calls in repeat-masked regions were not tested, owing to the challenge of designing reliable validation probes in these areas.

The 3 core pipelines had individual estimates of sensitivity of 80–90% to detect a true somatic SNV called by any of the 13 pipelines; with >95% of SNV calls made by each of the core pipelines being genuine somatic variants (Fig. 1a). For indels—a more-challenging class of variants to identify in short-read sequencing data—the 3 core algorithms had individual sensitivity estimates in the range of 40–50%, with precision 70–95% (Fig. 1b). Validation of SV calls is inherently more difficult, as methods based on PCR or hybridization to RNA baits often fail to isolate DNA that spans the breakpoint. To assess the accuracy of SV calls, we therefore used the property that an SV must either generate a copy-number change or be balanced, whereas artefactual calls will not respect this property. For individual SV-calling algorithms, we estimated precision to be in the range of 80–95% for samples in the 63-sample pilot dataset.

Next, we examined multiple methods for merging calls made by several algorithms into a single definitive call-set to be used for downstream analysis. The final consensus calls for SNVs were based on a simple approach that required two or more methods to agree on a call. For indels, because methods were less concordant, we used stacked logistic regression[110,111] to integrate the calls. The merged SV set includes all calls made by two or more of the four primary SV-calling algorithms[96,100,112,113]. Consensus CNA calls were obtained by joining the outputs of six individual CNA-calling algorithms with SV consensus breakpoints to obtain base-pair resolution CNAs (Supplementary Methods 2.4.3). Consensus purity and ploidy were derived, and a multitier system was developed for consensus copy-number calls (Supplementary Methods 2.4.3, and described in detail elsewhere[7]).

Overall, the sensitivity and precision of the consensus somatic variant calls were 95% (90% confidence interval, 88–98%) and 95% (90% confidence interval, 71–99%), respectively, for SNVs (Extended Data Fig. 2). For somatic indels, sensitivity and precision were 60% (90% confidence interval, 34–72%) and 91% (90% confidence interval, 73–96%), respectively. Regarding SVs, we estimate the sensitivity of the merging algorithm to be 90% for true calls generated by any one calling pipeline; precision was estimated to be 97.5%. That is, 97.5% of SVs in the merged SV call-set had an associated copy-number change or balanced partner rearrangement. The improvement in calling accuracy from combining different pipelines was most noticeable in variants that had low variant allele fractions, which are likely to originate from subclonal populations of the tumour (Fig. 1c, d). There remains much work to be done to improve indel calling software; we still lack sensitivity for calling even fully clonal complex indels from short-read sequencing data.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The PCAWG-generated alignments, somatic variant calls, annotations and derived datasets are available for general research use for browsing and download at http://dcc.icgc.org/pcawg/ (Box 1 and Supplementary Table 4). In accordance with the data access policies of the ICGC and TCGA projects, most molecular, clinical and specimen data are in an open tier which does not require access approval. To access potentially identifying information, such as germline alleles and underlying read data, researchers will need to apply to the TCGA Data Access Committee (DAC) via dbGaP (https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login) for access to the TCGA portion of the dataset, and to the ICGC Data Access Compliance Office (DACO; http://icgc.org/daco) for the ICGC portion. In addition, to access somatic single nucleotide variants derived from TCGA donors, researchers will also need to obtain dbGaP authorization.

Beyond the core sequence data and variant call-sets, the analyses in this paper used a number of datasets that were derived from the variant calls (Supplementary Table 4). The individual datasets are available at Synapse (https://www.synapse.org/), and are denoted with synXXXXX accession numbers; all these datasets are also mirrored at https://dcc.icgc.org, with full links, filenames, accession numbers and descriptions detailed in Supplementary Table 4. The datasets encompass: clinical data from each patient including demographics, tumour stage and vital status (syn10389158); harmonized tumour histopathology annotations using a standardised hierarchical ontology (syn1038916); inferred purity and ploidy values for each tumour sample (syn8272483); driver mutations for each patient from their cancer genome spanning all classes of variant, and coding versus non-coding drivers (syn11639581); mutational signatures inferred from PCAWG donors (syn11804065), including APOBEC mutagenesis (syn7437313); and transcriptional data from RNA sequencing, including gene expression levels (syn5553985, syn5553991, syn8105922) and gene fusions (syn10003873, syn7221157).

Code availability

Computational pipelines for calling somatic mutations are available to the public at https://dockstore.org/organizations/PCAWG/collections/PCAWG. A range of data-visualization and -exploration tools are also available for the PCAWG data (Box 1).

89. NCI SEER. *ICD-O-3 Coding Materials* (2018).
90. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
91. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
92. Raine, K. M. et al. ascatNgs: identifying somatically acquired copy-number alterations from whole-genome sequencing data. *Curr. Protoc. Bioinformatics* **56**, 15.9.1–15.9.17 (2016).
93. Jones, D. et al. cgpCaVEManWrapper: simple execution of CaVEMan in order to detect somatic single nucleotide variants in NGS data. *Curr. Protoc. Bioinformatics* **56**, 15.10.1–15.10.18 (2016).
94. Raine, K. M. et al. cgpPindel: identifying somatically acquired insertion and deletion events from paired end sequencing. *Curr. Protoc. Bioinformatics* **52**, 15.7.1–15.7.12 (2015).
95. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
96. Rausch, T. et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339 (2012).
97. Rimmer, A. et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014).
98. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
99. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
100. Drier, Y. et al. Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res.* **23**, 228–235 (2013).
101. Ramos, A. H. et al. Oncotator: cancer variant annotation tool. *Hum. Mutat.* **36**, E2423–E2429 (2015).
102. Moncunill, V. et al. Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. *Nat. Biotechnol.* **32**, 1106–1112 (2014).
103. Fan, Y. et al. MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. *Genome Biol.* **17**, 178 (2016).
104. Cooke, S. L. et al. Processed pseudogenes acquired somatically during cancer development. *Nat. Commun.* **5**, 3644 (2014).
105. Ju, Y. S. et al. Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. *eLife* **3**, e02935 (2014).
106. Sudmant, P. H. et al. An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015).
107. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at https://arxiv.org/abs/1207.3907 (2012).
108. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

109. Yakneen, S., Waszak, S. M., Gertz, M. & Korbel, J. O. & PCAWG Consortium. Butler enables rapid cloud-based analysis of thousands of human genomes. *Nat. Biotechnol*. https://doi.org/10.1038/s41587-019-0360-3 (2020).
110. Kim, S. Y., Jacob, L. & Speed, T. P. Combining calls from multiple somatic mutation-callers. *BMC Bioinformatics* **15**, 154 (2014).
111. Breiman, L. Stacked regressions. *Mach. Learn*. **24**, 49–64 (1996).
112. Campbell, P. J. et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet*. **40**, 722–729 (2008).
113. Wala, J. A. et al. SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res*. **28**, 581–591 (2018).

Acknowledgements We thank research participants who donated samples and data, the physicians and clinical staff who contributed to sample annotation and collection, and the numerous funding agencies that contributed to the collection and analysis of this dataset.

Author contributions Writing committee leads: Peter J. Campbell, Gad Getz, Jan O. Korbel, Joshua M. Stuart, Jennifer L. Jennings, Lincoln D. Stein. Head of project management: Jennifer L. Jennings. Sample collection: major contributions from Marc D. Perry, Hardeep K. Nahal-Bose; led by B. F. Francis Ouellette. Histopathology harmonization: major contribution from Constance H. Li; further contributions from Esther Rheinbay, G. Petur Nielsen, Dennis C. Sgroi, Chin-Lee Wu, William C. Faquin, Vikram Deshpande, Paul C. Boutros, Alexander J. Lazar, Katherine A. Hoadley; led by Lincoln D. Stein, David N. Louis. Uniform processing, somatic, germline variant calling: major contribution from L. Jonathan Dursi; further contributions from Christina K. Yung, Matthew H. Bailey, Gordon Saksena, Keiran M. Raine, Ivo Buchhalter, Kortine Kleinheinz, Matthias Schlesner, Junjun Zhang, Wenyi Wang, David A. Wheeler; led by Li Ding, Jared T. Simpson. Core alignment, variant calling by cloud computing: major contributions from Christina K. Yung, Brian D. O'Connor, Sergei Yakneen, Junjun Zhang; further contributions from Kyle Ellrott, Kortine Kleinheinz, Naoki Miyoshi, Keiran M. Raine, Adam P. Butler, Romina Royo, Gordon Saksena, Matthias Schlesner, Solomon I. Shorser, Miguel Vazquez. Integration, phasing, validation of germline variant callsets: major contributions from Tobias Rausch, Grace Tiao, Sebastian M. Waszak, Bernardo Rodriguez-Martin, Suyash Shringarpure, Dai-Ying Wu; further contributions from Sergei Yakneen, German M. Demidov, Olivier Delaneau, Shuto Hayashi, Seiya Imoto, Nina Habermann, Ayellet V. Segre, Erik Garrison, Andy Cafferkey, Eva G. Alvarez, José María Heredia-Genestar, Francesc Muyas, Oliver Drechsel, Alicia L. Bruzos, Javier Temes, Jorge Zamora, L. Jonathan Dursi, Adrian Baez-Ortega, Hyung-Lae Kim, Matthew H. Bailey, R. Jay Mashl, Kai Ye, Ivo Buchhalter, Anthony DiBiase, Kuan-lin Huang, Ivica Letunic, Michael D. McLellan, Steven J. Newhouse, Matthias Schlesner, Tal Shmaya, Sushant Kumar, David C. Wedge, Mark H. Wright, Venkata D. Yellapantula, Mark Gerstein, Ekta Khurana, Tomas Marques-Bonet, Arcadi Navarro, Carlos D. Bustamante, Jared T. Simpson, Li Ding, Reiner Siebert, Hidewaki Nakagawa, Douglas F. Easton; led by Stephan Ossowski, Jose M. C. Tubio, Gad Getz, Francisco M. De La Vega, Xavier Estivill, Jan O. Korbel. Validation, benchmarking, merging of somatic variant calls: major contribution from L. Jonathan Dursi; further contributions from David A. Wheeler, Christina K. Yung; led by Li Ding, Jared T. Simpson. Data and code availability: major contribution from Junjun Zhang; further contributions from Christina K. Yung, Sergei Yakneen, Denis Yuen, George L. Mihaiescu, Larsson Omberg; led by Vincent Ferretti. Pan-cancer burden of somatic mutations: major contribution from Junjun Zhang; led by Peter J. Campbell. Panorama of driver mutations in human cancer: led by Radhakrishnan Sabarinathan, Oriol Pich, Abel Gonzalez-Perez. PCAWG tumours with no apparent driver mutations: major contribution from Esther Rheinbay; further contributions from Amaro Taylor-Weiner, Radhakrishnan Sabarinathan; led by Peter J. Campbell, Gad Getz. Patterns, oncogenicity of kataegis, chromoplexy: major contributions from Matthew W. Fittall, Jonas Demeulemeester, Maxime Tarabichi; further contributions from Nicola D. Roberts, Peter J. Campbell, Jan O. Korbel; led by Peter Van Loo. Patterns, oncogenicity of chromothripsis: major contributions from Maxime Tarabichi, Jonas Demeulemeester, Matthew W. Fittall; further contributions from Isidro Cortes-Ciriano, Lara Urban, Peter J. Park, Peter J. Campbell, Jan O. Korbel; led by Peter Van Loo. Timing-clustered mutational processes during tumour evolution: major contributions from Jonas Demeulemeester, Maxime Tarabichi, Matthew W. Fittall; further contributions from Jan O. Korbel, Peter J. Campbell; led by Peter Van Loo. Germline effects on somatic mutation: major contributions from Sebastian M. Waszak, Bin Zhu, Bernardo Rodriguez-Martin, Esa Pitkanen, Tobias Rausch; further contributions from Yilong Li, Natalie Saini, Leszek J. Klimczak, Joachim Weischenfeldt, Nikos Sidiropoulos, Ludmil B. Alexandrov, Francesc Muyas, Raquel Rabionet, Georgia Escaramis, Adrian Baez-Ortega, Mattia Bosio, Aliaksei Z. Holik, Hana Susak, Eva G. Alvarez, Alicia L. Bruzos, Javier Temes, Aparna Prasad, Nina Habermann, Serap Erkek, Lara Urban, Claudia Calabrese, Benjamin Raeder, Eoghan Harrington, Simon Mayes, Daniel Turner, Sissel Juul, Steven A. Roberts, Lei Song, Roelof Koster, Lisa Mirabello, Xing Hua, Tomas J. Tanskanen, Marta Tojo, David C. Wedge, Jorge Zamora, Jieming Chen, Lauri A. Aaltonen, Gunnar Ratsch, Roland F. Schwarz, Atul J. Butte, Alvis Brazma, Peter J. Campbell, Stephen J. Chanock, Nilanjan Chatterjee, Oliver Stegle, Olivier Harismendy; led by G. Steven Bova, Dmitry A. Gordenin, Jose M. C. Tubio, Douglas F. Easton, Xavier Estivill, Jan O. Korbel. Replicative immortality: major contribution from David Haan; further contributions from Lina Sieverling, Lars Feuerbach; led by Lincoln D. Stein, Joshua M. Stuart. Ethical considerations of genomic cloud computing: led by Don Chalmers, Yann Joly, Bartha Knoppers, Fruzsina Molnar-Gabor, Jan O. Korbel, Mark Phillips, Adrian Thorogood, David Townend. Online resources for data access, visualization, exploration and analysis: major contributions from Mary Goldman, Junjun Zhang, Nuno A. Fonseca; further contributions from Qian Xiang, Brian Craft, Elena Pineiro-Yanez, Alfonso Munoz, Robert Petryszak, Anja Fullgrabe, Fatima Al-Shahrour, Maria Keays, David Haussler, John Weinstein, Wolfgang Huber, Alfonso Valencia, Irene Papatheodorou, Jingchun Zhu; led by Brian D. O'Connor, Lincoln D. Stein, Alvis Brazma, Vincent Ferretti, Miguel Vazquez. The 63-sample pilot-analysis validation process: major contribution from L. Jonathan Dursi; further contributions from Christina K. Yung, Matthew H. Bailey, Gordon Saksena, Keiran M. Raine, Ivo Buchhalter, Kortine Kleinheinz, Matthias Schlesner, Yu Fan, David Torrents, Matthias Bieg, Paul C. Boutros, Ken Chen, Zechen Chong, Kristian Cibulskis, Oliver Drechsel, Roland Eils, Robert S. Fulton, Josep Gelpi, Mark Gerstein, Santiago Gonzalez, Gad Getz, Ivo G. Gut, Faraz Hach, Michael Heinold, Taobo Hu, Vincent Huang, Barbara Hutter, Hyung-Lae Kim, Natalie Jager, Jongsun Jung, Sushant Kumar, Yogesh Kumar, Christopher Lalansingh, Ignaty Leshchiner, Ivica Letunic, Dimitri Livitz, Eric Z. Ma, Yosef E. Maruvka, R. Jay Mashl, Michael D. McLellan, Ana Milovanovic, Morten Muhlig Nielsen, Brian D. O'Connor, Stephan Ossowski, Nagarajan Paramasivam, Jakob Skou Pedersen, Marc D. Perry, Montserrat Puiggros, Romina Royo, Esther Rheinbay, S. Cenk Sahinalp, Iman Sarrafi, Chip Stewart, Miranda D. Stobbe, Grace Tiao, Jeremiah A. Wala, Jiayin Wang, Wenyi Wang, Sebastian M. Waszak, Joachim Weischenfeldt, Michael Wendl, Johannes Werner, Zhenggang Wu, Hong Xue, Sergei Yakneen, Takafumi N. Yamaguchi, Kai Ye, Venkata Yellapantula, Junjun Zhang, David A. Wheeler; led by Li Ding, Jared T. Simpson. Processing of validation data: major contributions from Christina K. Yung, Brian D. O'Connor, Sergei Yakneen, Junjun Zhang; further contributions from Kyle Ellrott, Kortine Kleinheinz, Naoki Miyoshi, Keiran M. Raine, Romina Royo, Gordon Saksena, Matthias Schlesner, Solomon I. Shorser, Miguel Vazquez, Joachim Weischenfeldt, Denis Yuen, Adam P. Butler, Brandi N. Davis-Dusenbery, Roland Eils, Vincent Ferretti, Robert L. Grossman, Olivier Harismendy, Youngwook Kim, Hidewaki Nakagawa, Steven J. Newhouse, David Torrents; led by Lincoln D. Stein. Whole-genome sequencing somatic variant calling: major contribution from Junjun Zhang; further contributions from Christina K. Yung, Solomon I. Shorser. Whole-genome alignment: Keiran M. Raine, Junjun Zhang, Brian D. O'Connor. DKFZ pipeline: Kortine Kleinheinz, Tobias Rausch, Jan O. Korbel, Ivo Buchhalter, Michael C. Heinold, Barbara Hutter, Natalie Jager, Nagarajan Paramasivam, Matthias Schlesner. EMBL pipeline: Joachim Weischenfeldt. Sanger pipeline: Keiran M. Raine, Jonathan Hinton, David R. Jones, Andrew Menzies, Lucy Stebbings, Adam P. Butler. Broad pipeline: Gordon Saksena, Dimitri Livitz, Esther Rheinbay, Julian M. Hess, Ignaty Leshchiner, Chip Stewart, Grace Tiao, Jeremiah A. Wala, Amaro Taylor-Weiner, Mara Rosenberg, Andrew J. Dunford, Manaswi Gupta, Marcin Imielinski, Matthew Meyerson, Rameen Beroukhim, Gad Getz. MuSE Pipeline: Yu Fan, Wenyi Wang. Consensus somatic SNV/indel annotation: Andrew Menzies, Matthias Schlesner, Juri Reimand, Priyanka Dhingra, Ekta Khurana. Somatic SNV, indel merging: major contribution from L. Jonathan Dursi; further contributions from Christina K. Yung, Matthew H. Bailey, Gordon Saksena, Keiran M. Raine, Ivo Buchhalter, Kortine Kleinheinz, Matthias Schlesner, Yu Fan, David Torrents, Matthias Bieg, Paul C. Boutros, Ken Chen, Zechen Chong, Kristian Cibulskis, Oliver Drechsel, Roland Eils, Robert S. Fulton, Josep L. Gelpi, Mark Gerstein, Santiago Gonzalez, Gad Getz, Ivo G. Gut, Faraz Hach, Michael Heinold, Taobo Hu, Vincent Huang, Barbara Hutter, Hyung-Lae Kim, Natalie Jager, Jongsun Jung, Sushant Kumar, Christopher Lalansingh, Ignaty Leshchiner, Ivica Letunic, Dimitri Livitz, Eric Z. Ma, Yosef E. Maruvka, R. Jay Mashl, Michael D. McLellan, Ana Milovanovic, Morten Muhlig Nielsen, Brian D. O'Connor, Stephan Ossowski, Nagarajan Paramasivam, Jakob Skou Pedersen, Marc D. Perry, Montserrat Puiggros, Romina Royo, Esther Rheinbay, S. Cenk Sahinalp, Iman Sarrafi, Chip Stewart, Miranda D. Stobbe, Grace Tiao, Jeremiah A. Wala, Jiayin Wang, Wenyi Wang, Sebastian M. Waszak, Joachim Weischenfeldt, Michael Wendl, Johannes Werner, Zhenggang Wu, Hong Xue, Sergei Yakneen, Takafumi N. Yamaguchi, Kai Ye, Venkata Yellapantula, Junjun Zhang, David A. Wheeler; major contributions from Li Ding, Jared T. Simpson. Somatic SV merging: Joachim Weischenfeldt, Francesco Favero, Yilong Li. Somatic CNA merging: Stefan Dentro, Jeff Wintersinger, Ignaty Leshchiner. Oxidative artefact filtration: Dimitri Livitz, Ignaty Leshchiner, Chip Stewart, Esther Rheinbay, Gordon Saksena, Gad Getz. Strand bias filtration: Matthias Bieg, Ivo Buchhalter, Johannes Werner, Matthias Schlesner. miniBAM generation: Jeremiah A. Wala, Gordon Saksena, Rameen Beroukhim, Gad Getz. Germline variant identification from whole-genome sequencing: major contributions from Tobias Rausch, Grace Tiao, Sebastian M. Waszak, Bernardo Rodriguez-Martin, Suyash Shringarpure, Dai-Ying Wu; further contributions from Sergei Yakneen, German M. Demidov, Olivier Delaneau, Shuto Hayashi, Seiya Imoto, Nina Habermann, Ayellet V. Segre, Erik Garrison, Andy Cafferkey, Eva G. Alvarez, Alicia L. Bruzos, Jorge Zamora, José María Heredia-Genestar, Francesc Muyas, Oliver Drechsel, L. Jonathan Dursi, Adrian Baez-Ortega, Hyung-Lae Kim, Matthew H. Bailey, R. Jay Mashl, Kai Ye, Ivo Buchhalter, Vasilisa Rudneva, Ji Wan Park, Eun Pyo Hong, Seong Gu Heo, Anthony DiBiase, Kuan-lin Huang, Ivica Letunic, Michael D. McLellan, Steven J. Newhouse, Matthias Schlesner, Tal Shmaya, Sushant Kumar, David C. Wedge, Mark H. Wright, Venkata D. Yellapantula, Mark Gerstein, Ekta Khurana, Tomas Marques-Bonet, Arcadi Navarro, Carlos D. Bustamante, Jared T. Simpson, Li Ding, Reiner Siebert, Hidewaki Nakagawa, Douglas F. Easton; led by Stephan Ossowski, Jose M. C. Tubio, Gad Getz, Francisco M. De La Vega, Xavier Estivill, Jan O. Korbel. RNA-sequencing analysis: major contributions from Nuno A. Fonseca, Andre Kahles, Kjong-Van Lehmann, Lara Urban, Cameron M. Soulette, Yuichi Shiraishi, Fenglin Liu, Yao He, Deniz Demircioglu, Natalie R. Davidson, Claudia Calabrese, Junjun Zhang, Marc D. Perry, Qian Xiang; further contributions from Liliana Greger, Siliang Li, Dongbing Liu, Stefan G. Stark, Fan Zhang, Samirkumar B. Amin, Peter Bailey, Aurelien Chateigner, Isidro Cortes-Ciriano, Brian Craft, Serap Erkek, Milana Frenkel-Morgenstern, Mary Goldman, Katherine A. Hoadley, Yong Hou, Matthew R. Huska, Ekta Khurana, Helena Kilpinen, Jan O. Korbel, Fabien C. Lamaze, Chang Li, Xiaobo Li, Xinyue Li, Xingmin Liu, Maximillian G. Marin, Julia Markowski, Tannistha Nandi, Morten Muhlig Nielsen, Akinyemi I. Ojesina, Qiang Pan-Hammarstrom, Peter J. Park, Chandra Sekhar Pedamallu, Jakob Skou Pedersen, Reiner Siebert, Hong Su, Patrick Tan, Bin Tean Teh, Jian Wang, Sebastian M. Waszak, Heng Xiong, Sergei Yakneen, Chen Ye, Christina Yung, Xiuqing Zhang, Liangtao Zheng, Jingchun Zhu, Shida Zhu, Philip Awadalla, Chad J. Creighton, Matthew Meyerson, B. F. Francis Ouellette, Kui Wu, Huanming Yang; led by Jonathan Goke, Roland F. Schwarz, Oliver Stegle, Zemin Zhang, Alvis Brazma, Gunnar Ratsch, Angela N. Brooks. Clustering of tumour genomes based on telomere maintenance-related features: major contribution from David Haan; led by Lincoln D. Stein, Joshua M. Stuart. Clustered mutational processes in PCAWG: major contributions from Jonas Demeulemeester, Maxime Tarabichi, Matthew W. Fittall; led by Peter J. Campbell, Jan O. Korbel, Peter Van Loo. Tumours without detected driver mutations: Esther Rheinbay, Amaro Taylor-Weiner, Radhakrishnan Sabarinathan, Peter J. Campbell, Gad Getz. Panorama of driver mutations in human cancer: major contributions from Radhakrishnan Sabarinathan, Oriol Pich; further contributions from Inigo Martincorena, Carlota Rubio-Perez, Malene Juul, Jeremiah A. Wala, Steven Schumacher, Ofer Shapira, Nikos Sidiropoulos, Sebastian M. Waszak, David Tamborero, Loris Mularoni, Esther Rheinbay, Henrik Hornshoj, Jordi Deu-Pons, Ferran Muinos, Johanna Bertl, Qianyun Guo, Chad J. Creighton, Joachim Weischenfeldt, Jan O. Korbel, Gad Getz, Peter J. Campbell, Jakob Skou Pedersen, Rameen Beroukhim; led by Abel Gonzalez-Perez. Pilot benchmarking, variant consensus development and validation: major contribution from L. Jonathan Dursi; further contributions from Christina K. Yung, Matthew H. Bailey, Gordon Saksena, Keiran M. Raine, Ivo Buchhalter, Kortine Kleinheinz, Matthias Schlesner, Yu Fan, David Torrents, Matthias Bieg, Paul C. Boutros, Ken Chen, Zechen Chong, Kristian Cibulskis, Oliver Drechsel, Roland Eils, Robert S. Fulton, Josep

Gelpi, Mark Gerstein, Santiago Gonzalez, Gad Getz, Ivo G. Gut, Faraz Hach, Michael Heinold, Taobo Hu, Vincent Huang, Barbara Hutter, Hyung-Lae Kim, Natalie Jager, Jongsun Jung, Sushant Kumar, Yogesh Kumar, Christopher Lalansingh, Ignaty Leshchiner, Ivica Letunic, Dimitri Livitz, Eric Z. Ma, Yosef E. Maruvka, R. Jay Mashl, Michael D. McLellan, Ana Milovanovic, Morten Muhlig Nielsen, Brian D. O'Connor, Stephan Ossowski, Nagarajan Paramasivam, Jakob Skou Pedersen, Marc D. Perry, Montserrat Puiggros, Romina Royo, Esther Rheinbay, S. Cenk Sahinalp, Iman Sarrafi, Chip Stewart, Miranda D. Stobbe, Grace Tiao, Jeremiah A. Wala, Jiayin Wang, Wenyi Wang, Sebastian M. Waszak, Joachim Weischenfeldt, Michael Wendl, Johannes Werner, Zhenggang Wu, Hong Xue, Sergei Yakneen, Takafumi N. Yamaguchi, Kai Ye, Venkata Yellapantula, Junjun Zhang, David A. Wheeler; led by Li Ding, Jared T. Simpson. Production somatic variant calling on the PCAWG compute cloud: major contributions from Christina K. Yung, Brian D. O'Connor, Sergei Yakneen, Junjun Zhang; further contributions from Kyle Ellrott, Kortine Kleinheinz, Naoki Miyoshi, Keiran M. Raine, Romina Royo, Gordon Saksena, Matthias Schlesner, Solomon I. Shorser, Miguel Vazquez, Joachim Weischenfeldt, Denis Yuen, Adam P. Butler, Brandi N. Davis-Dusenbery, Roland Eils, Vincent Ferretti, Robert L. Grossman, Olivier Harismendy, Youngwook Kim, Hidewaki Nakagawa, Steven J Newhouse, David Torrents; led by Lincoln D. Stein. PCAWG data portals: major contributions from Mary Goldman, Junjun Zhang, Nuno A. Fonseca, Isidro Cortes-Ciriano; further contributions from Qian Xiang, Brian Craft, Elena Pineiro-Yanez, Brian D O'Connor, Wojciech Bazant, Elisabet Barrera, Alfonso Munoz, Robert Petryszak, Anja Fullgrabe, Fatima Al-Shahrour, Maria Keays, David Haussler, John Weinstein, Wolfgang Huber, Alfonso Valencia, Irene Papatheodorou, Jingchun Zhu; led by Vincent Ferretti, Miguel Vazquez.

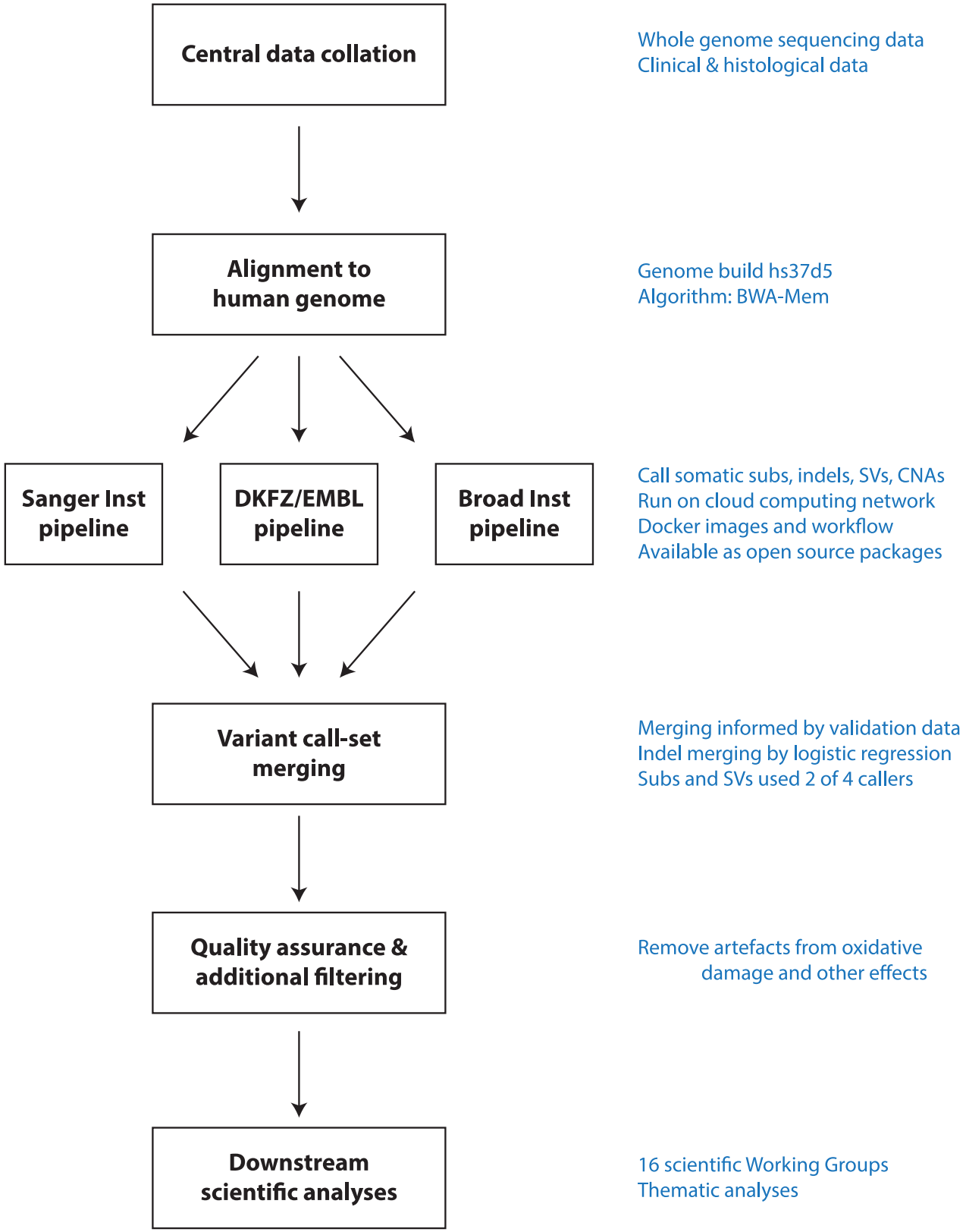
Competing interests Gad Getz receives research funds from IBM and Pharmacyclics and is an inventor on patent applications related to MuTect, ABSOLUTE, MutSig, MSMuTect, MSMutSig and POLYSOLVER. Hikmat Al-Ahmadie is consultant for AstraZeneca and Bristol-Myers Squibb. Samuel Aparicio is a founder and shareholder of Contextual Genomics. Pratiti Bandopadhayay receives grant funding from Novartis for an unrelated project. Rameen Beroukhim owns equity in Ampressa Therapeutics. Andrew Biankin receives grant funding from Celgene, AstraZeneca and is a consultant for or on advisory boards of AstraZeneca, Celgene, Elstar Therapeutics, Clovis Oncology and Roche. Ewan Birney is a consultant for Oxford Nanopore, Dovetail and GSK. Marcus Bosenberg is a consultant for Eli Lilly. Atul Butte is a cofounder of and consultant for Personalis, NuMedii, a consultant for Samsung, Geisinger Health, Mango Tree Corporation, Regenstrief Institute and in the recent past a consultant for 10x Genomics and Helix, a shareholder in Personalis, a minor shareholder in Apple, Twitter, Facebook, Google, Microsoft, Sarepta, 10x Genomics, Amazon, Biogen, CVS, Illumina, Snap and Sutro and has received honoraria and travel reimbursement for invited talks from Genentech, Roche, Pfizer, Optum, AbbVie and many academic institutions and health systems. Carlos Caldas has served on the Scientific Advisory Board of Illumina. Lorraine Chantrill acted on an advisory board for AMGEN Australia in the past 2 years. Andrew D. Cherniack receives research funding from Bayer. Helen Davies is an inventor on a number of patent applications that encompass the use of mutational signatures. Francisco De La Vega was employed at Annai Systems during part of the project. Ronny Drapkin serves on the scientific advisory board of Repare Therapeutics and Siamab Therapeutics. Rosalind Eeles has received an honorarium for the GU-ASCO meeting in San Francisco in January 2016 as a speaker, a honorarium and support from Janssen for the RMH FR meeting in November 2017 as a speaker (title: genetics and prostate cancer), a honorarium for an University of Chicago invited talk in May 2018 as speaker and an educational honorarium paid by Bayer & Ipsen to attend GU Connect 'Treatment sequencing for mCRPC patients within the changing landscape of mHSPC' at a venue at ESMO, Barcelona, on 28 September 2019. Paul Flicek is a member of the scientific advisory boards of Fabric Genomics and Eagle Genomics. Ronald Ghossein is a consultant for Veracyte. Dominik Glodzik is an inventor on a number of patent applications that encompass the use of mutational signatures. Eoghan Harrington is a full-time employee of Oxford Nanopore Technologies and is a stock holder. Yann Joly is responsible for the Data Access Compliance Office (DACO) of ICGC 2009-2018. Sissel Juul is a full-time employee of Oxford Nanopore Technologies and is a stock holder. Vincent Khoo has received personal fees and non-financial support from Accuray, Astellas, Bayer, Boston Scientific and Janssen. Stian Knappskog is a coprincipal investigator on a clinical trial that receives research funding from AstraZeneca and Pfizer. Ignaty Leshchiner is a consultant for PACT Pharma. Carlos López-Otín has ownership interest (including stock and patents) in DREAMgenics. Matthew Meyerson is a scientific advisory board chair of, and consultant for, OrigiMed, has obtained research funding from Bayer and Ono Pharma and receives patent royalties from LabCorp. Serena Nik-Zainal is an inventor on a number of patent applications that encompass the use of mutational signatures. Nathan Pennell has done consulting work with Merck, Astrazeneca, Eli Lilly and Bristol-Myers Squibb. Xose S. Puente has ownership interest (including stock and patents in DREAMgenics. Benjamin J. Raphael is a consultant for and has ownership interest (including stock and patents) in Medley Genomics. Jorge Reis-Filho is a consultant for Goldman Sachs and REPARE Therapeutics, member of the scientific advisory board of Volition RX and Paige.AI and an ad hoc member of the scientific advisory board of Ventana Medical Systems, Roche Tissue Diagnostics, InVicro, Roche, Genentech and Novartis. Lewis R. Roberts has received grant support from ARIAD Pharmaceuticals, Bayer, BTG International, Exact Sciences, Gilead Sciences, Glycotest, RedHill Biopharma, Target PharmaSolutions and Wako Diagnostics and has provided advisory services to Bayer, Exact Sciences, Gilead Sciences, GRAIL, QED Therapeutics and TAVEC Pharmaceuticals. Richard A. Scolyer has received fees for professional services from Merck Sharp & Dohme, GlaxoSmithKline Australia, Bristol-Myers Squibb, Dermpedia, Novartis Pharmaceuticals Australia, Myriad, NeraCare GmbH and Amgen. Tal Shmaya is employed at Annai Systems. Reiner Siebert has received speaker honoraria from Roche and AstraZeneca. Sabina Signoretti is a consultant for Bristol-Myers Squibb, AstraZeneca, Merck, AACR and NCI and has received funding from Bristol-Myers Squibb, AstraZeneca, Exelixis and royalties from Biogenex. Jared Simpson has received research funding and travel support from Oxford Nanopore Technologies. Anil K. Sood is a consultant for Merck and Kiyatec, has received research funding from M-Trap and is a shareholder in BioPath. Simon Tavaré is on the scientific advisory board of Ipsen and a consultant for Kallyope. John F. Thompson has received honoraria and travel support for attending advisory board meetings of GlaxoSmithKline and Provectus and has received honoraria for participation in advisory boards for MSD Australia and BMS Australia. Daniel Turner is a full-time employee of Oxford Nanopore Technologies and is a stock holder. Naveen Vasudev has received speaker honoraria and/or consultancy fees from Bristol-Myers Squibb, Pfizer, EUSA pharma, MSD and Novartis. Jeremiah A. Wala is a consultant for Nference. Daniel J. Weisenberger is a consultant for Zymo Research. Dai-Ying Wu is employed at Annai Systems. Cheng-Zhong Zhang is a cofounder and equity holder of Pillar Biosciences, a for-profit company that specializes in the development of targeted sequencing assays. The other authors declare no competing interests.

Additional information

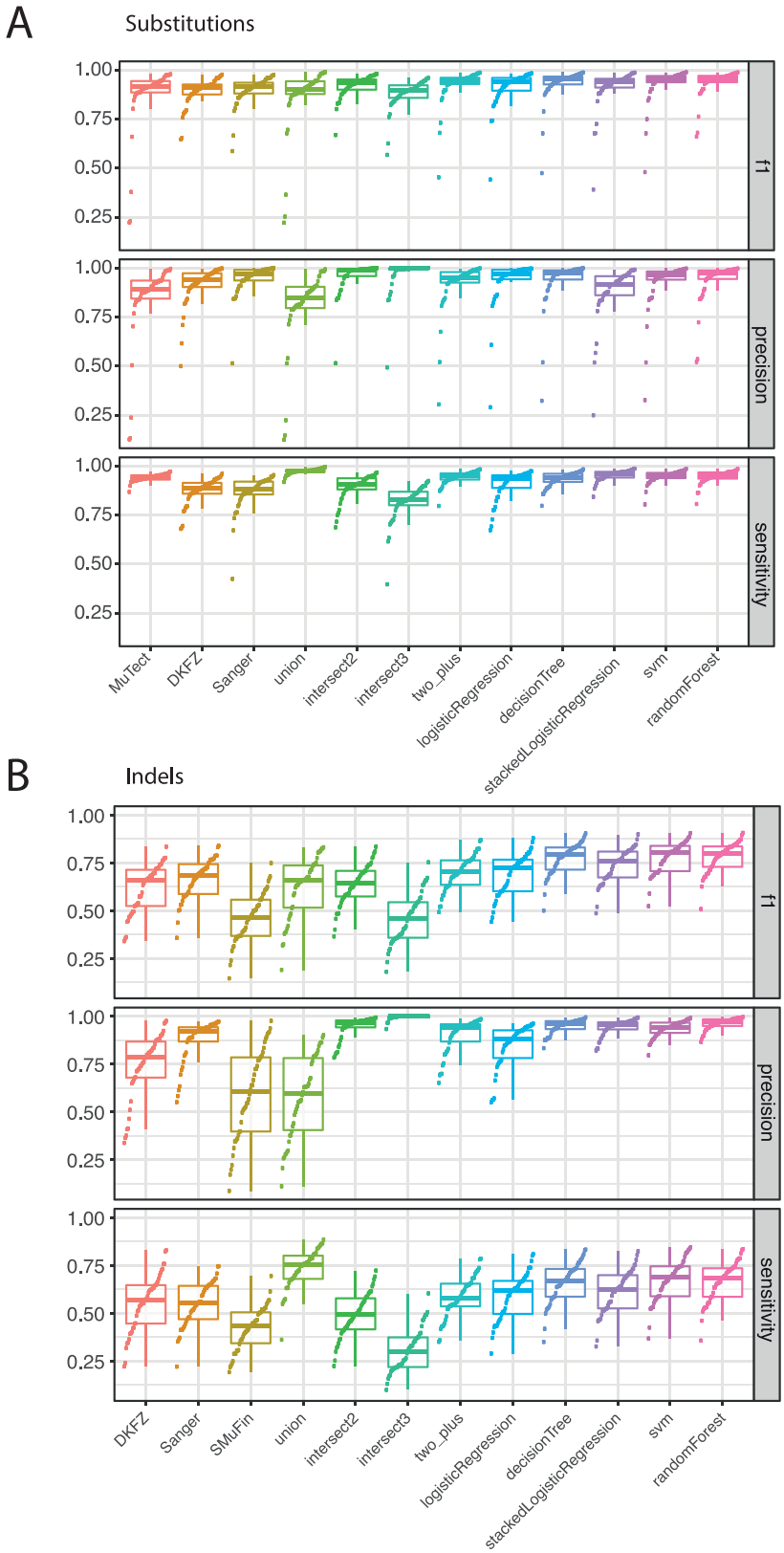
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-1969-6.

Correspondence and requests for materials should be addressed to P.J.C., G.G., J.O.K., J.M.S. or L.D.S.

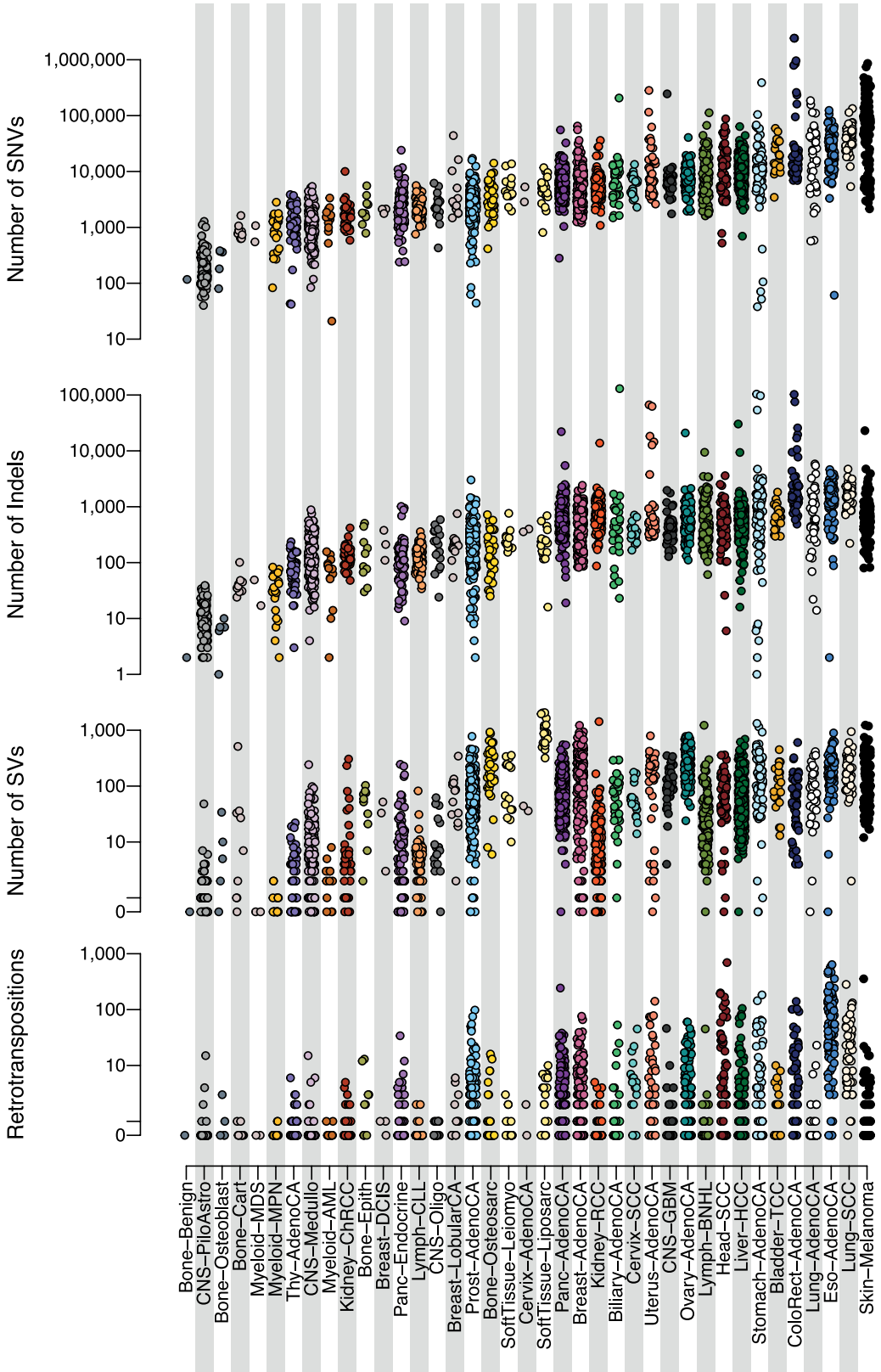
Peer review information *Nature* thanks Arul Chinnaiyan, Ben Lehner, Nicolas Robine and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



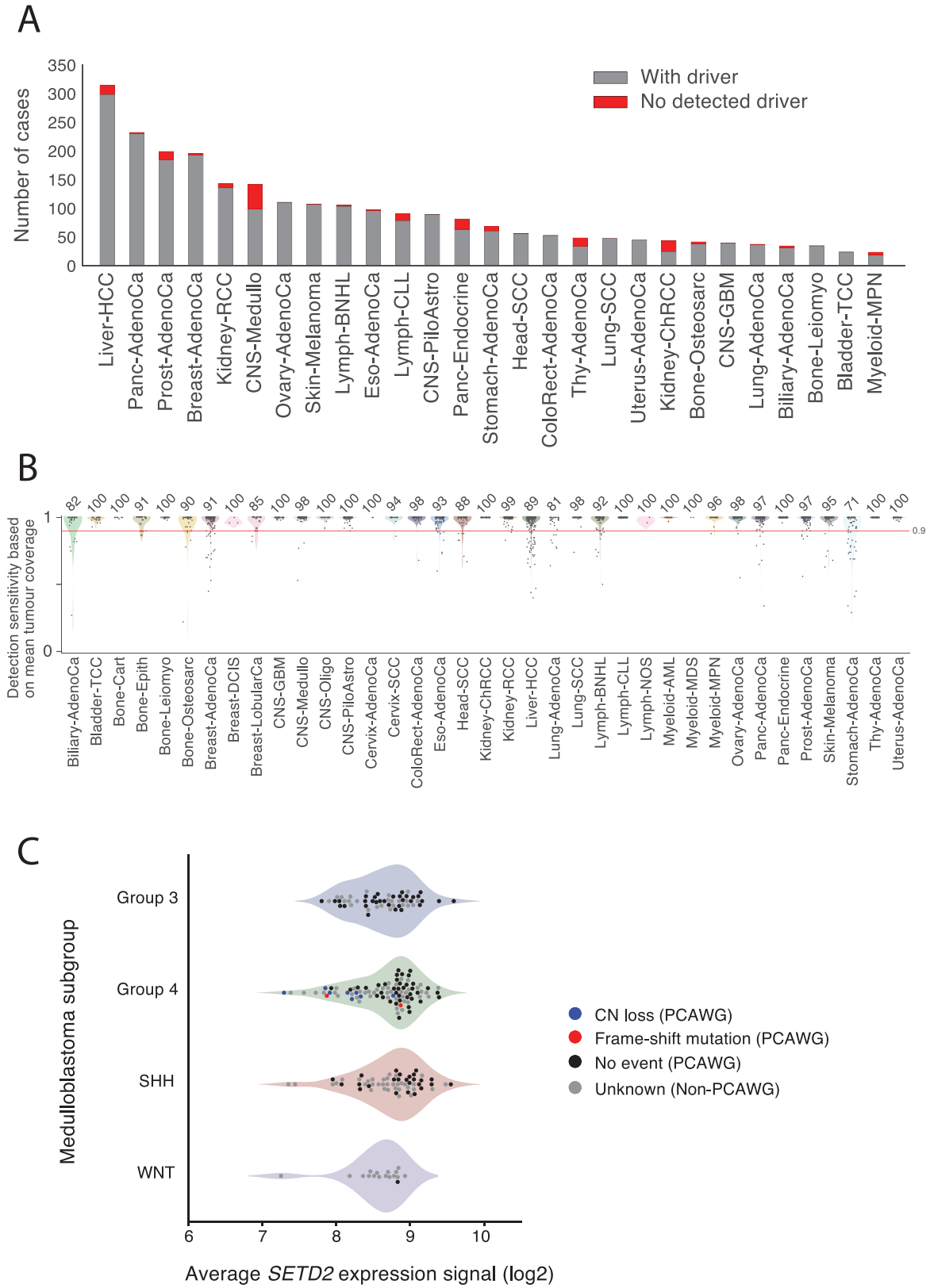
Extended Data Fig. 1 | Flow-chart showing key steps in the analysis of PCAWG genomes. After alignment to the genome, somatic mutations were identified by three pipelines, with subsequent merging into a consensus variant set used for downstream scientific analyses. Subs, substitutions; DKFZ/EMBL, the German Cancer Research Centre (DKFZ) and Europen Molecular Biology Laboratory (EMBL).



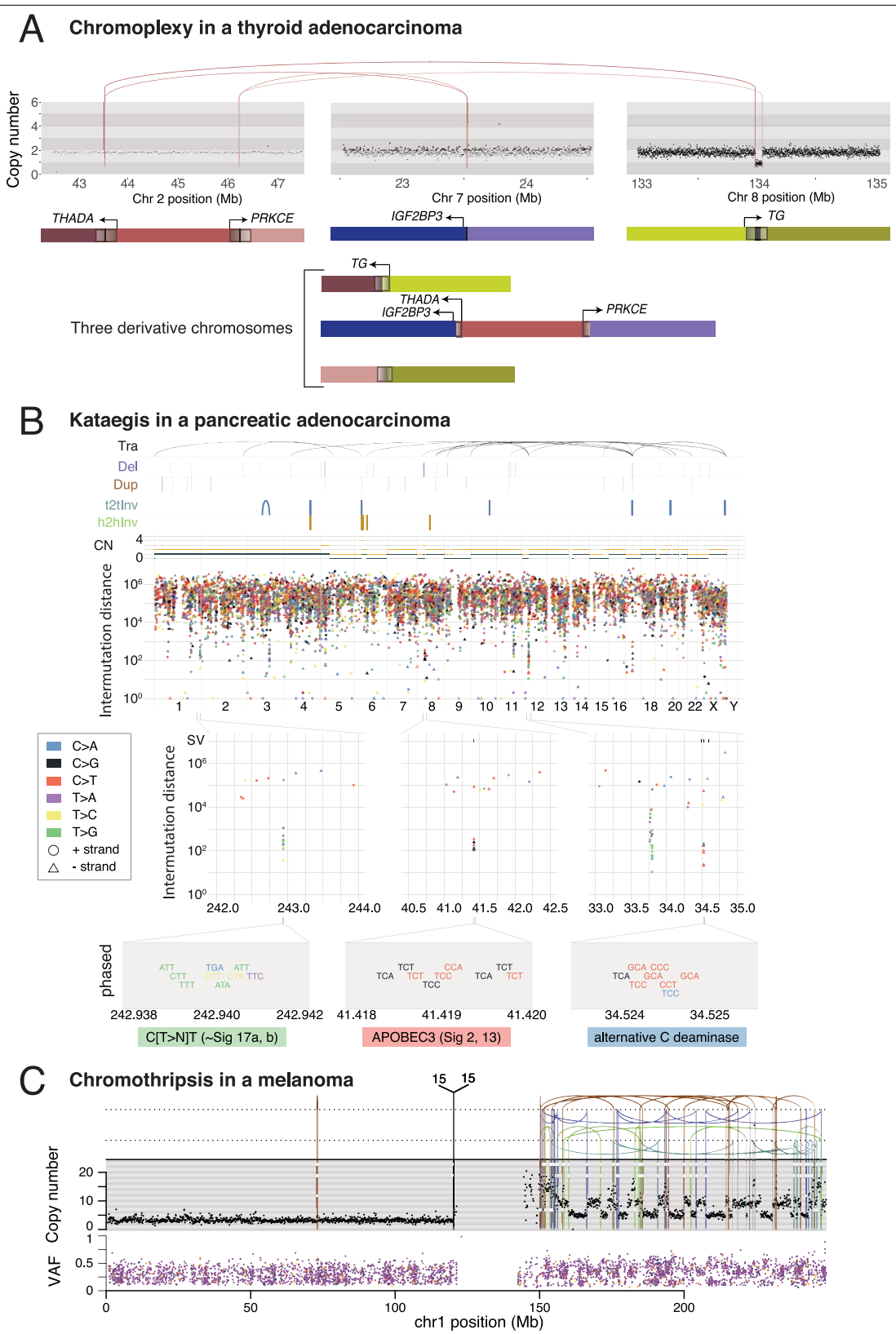
Extended Data Fig. 2 | Distribution of accuracy estimates across algorithms and samples from validation data. a, F_1 accuracy, precision and sensitivity estimates for somatic SNVs across the core algorithms and different approaches to merging the call-sets. The box plots demarcate the interquartile range and median of estimates across the $n = 50$ samples in the validation dataset. **b**, F_1 accuracy, precision and sensitivity estimates for somatic indels ($n = 50$ samples). SVM, support vector machine; union, calls made by all variant-calling algorithms; intersect2, calls made by any combination of two variant-calling algorithms; intersect3, calls made by any three variant-calling algorithms.



Extended Data Fig. 3 | Distribution of numbers of somatic mutations of different classes across tumour types. The *y* axis is on a log scale. The 2,583 donors with the highest quality metrics (white-listed donors) are plotted. SNVs indicate substitutions; indels are taken as insertions or deletions <100 bp in size; retrotranspositions are the combined counts of somatic retrotransposon insertions, transductions and somatic pseudogene insertions.

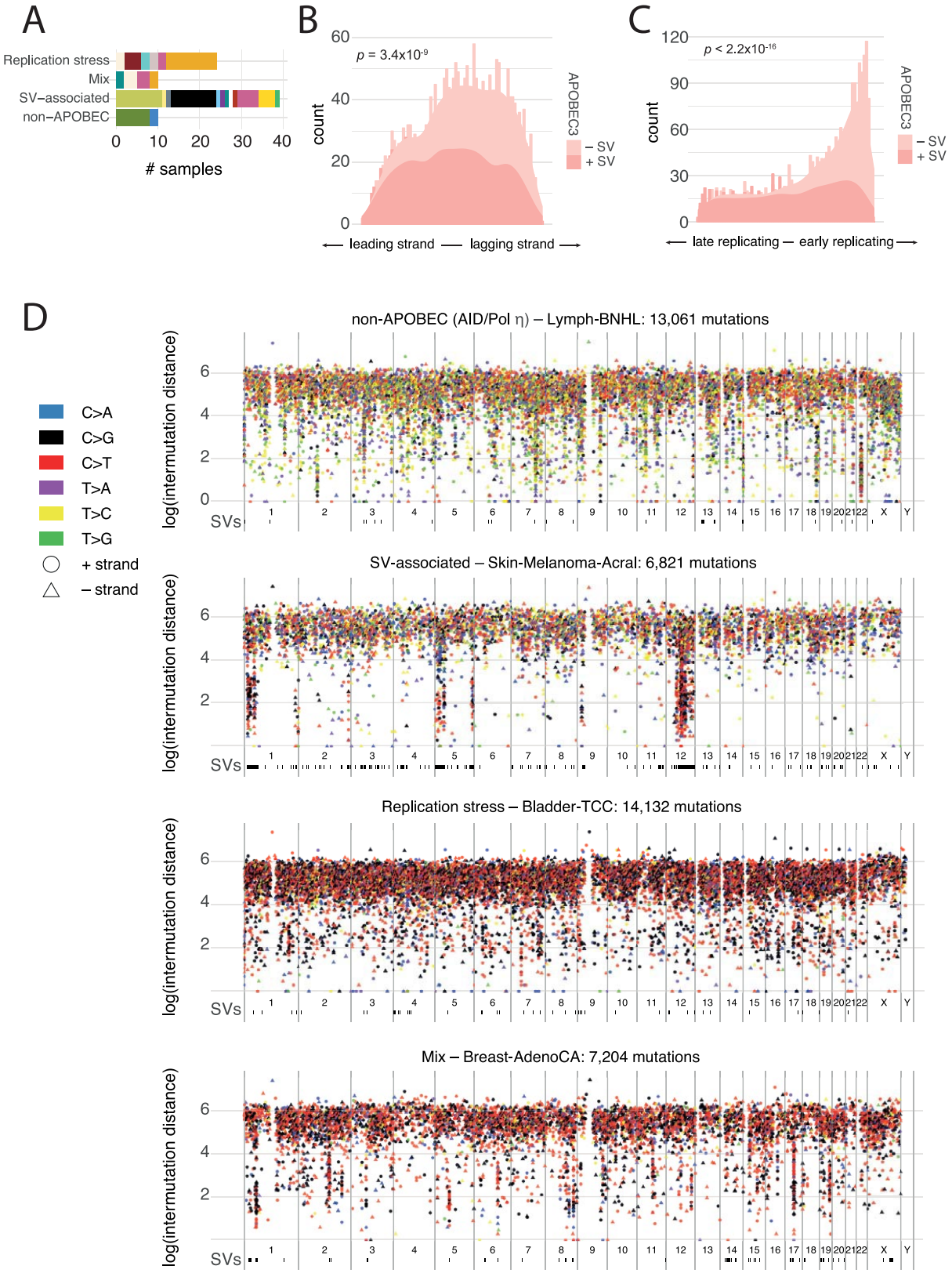


Extended Data Fig. 4 | Patients with no detected driver mutations in PCAWG. **a**, Number (red) of patients without detected driver mutations distributed across the different tumour types studied. **b**, Estimated sensitivity for detecting somatic point mutations genome-wide across tumour types (total sample size: $n = 2,583$ patients). Each point represents the estimate for a single patient, layered on violin plots that show the estimated density distribution of sensitivity values for that tumour type (the width proportional is to density). **c**, *SETD2* expression levels across different medulloblastoma subtypes. Points represent individual patients, coloured by whether the gene exhibited focal copy number (CN) loss or a truncating point mutation, or was the wild-type gene. The coloured areas are violin plots showing the estimated density distribution of expression values for that medulloblastoma subtype.

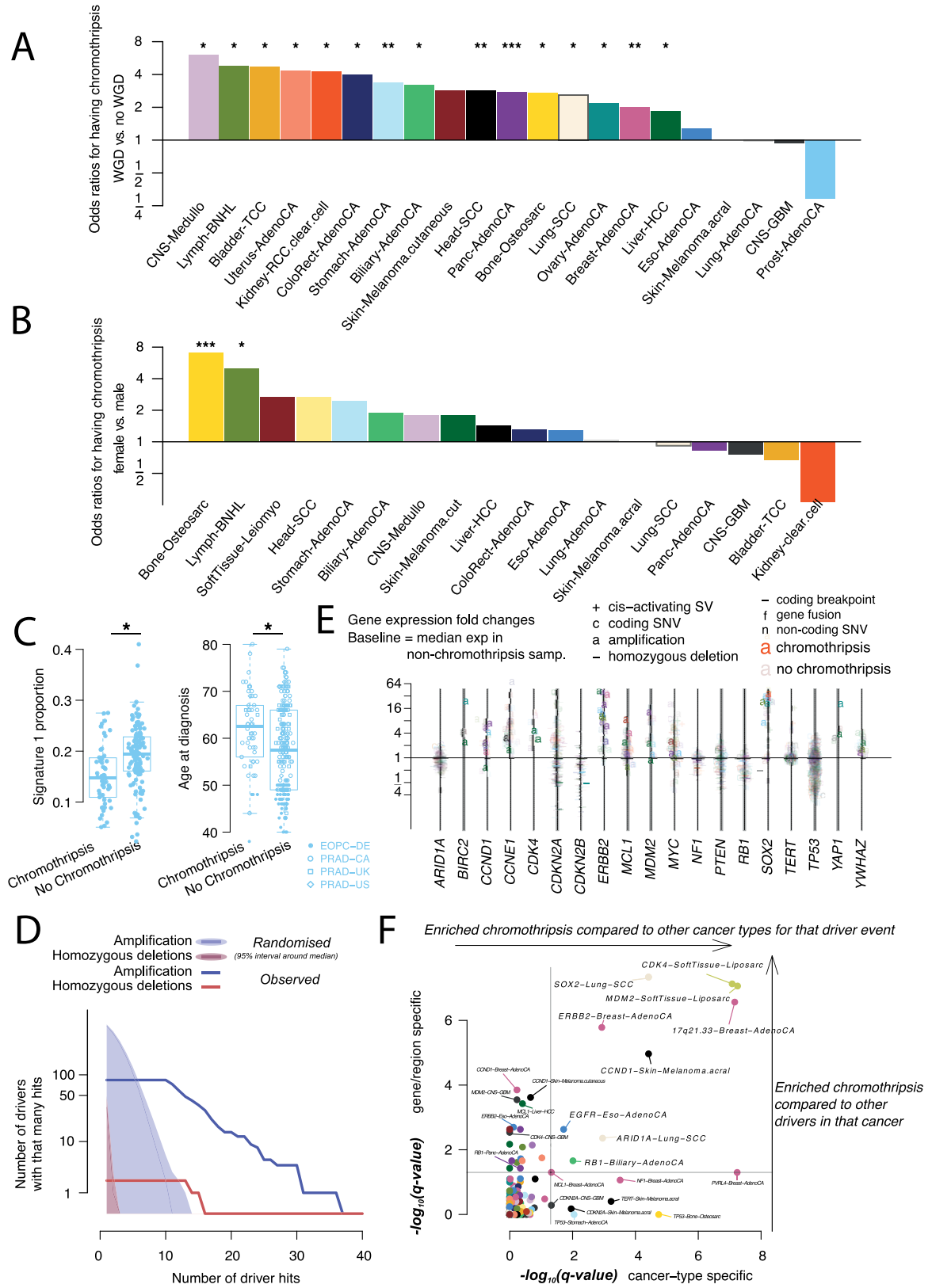


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Examples of clustered mutational processes. **a**, Chromoplexy example in a thyroid adenocarcinoma. Genes at the breakpoints are schematically depicted in their normal genomic context and again in the reconstructed derivative chromosomes below. **b**, Distinct kataegis signatures in the genome of a pancreatic adenocarcinoma sample. SVs and their classification are shown above the main rainfall plot, as well as the total and minor allele copy number. Tra, translocation; del, deletion; dup, duplication; t2tInv, tail-to-tail inversion; h2hInv, head-to-head inversion. Magnifications of the three foci on chromosomes 1, 8 and 12, respectively, highlight distinct manifestations of kataegis. Left, a novel process similar to signature 17 with T > N mutations at CT or TT dinucleotides. Middle, the prototypical APOBEC3A/B type with C > T (signature 2) and/or C>G/A (signature 13) substitutions at TpC. Right, an alternative cytidine deaminase(s) with a preference for substitutions at C/GpC. Most of the SNVs in each of these foci can be phased to the same allele and no evidence of anti-phasing is observed. **c**, Example of a chromothripsis event in a melanoma. The black points (top) represent copy-number estimates from individual genomic bins, with SVs shown as coloured arcs (translocation in black, deletion in purple, duplication in brown, tail-to-tail inversion in cyan, head-to-head inversion in green) that mostly demarcate copy-number changes. The mate chromosomes are displayed above translocations. Bottom, the variant allele fractions of somatic mutations distributed along the relevant chromosomal region.

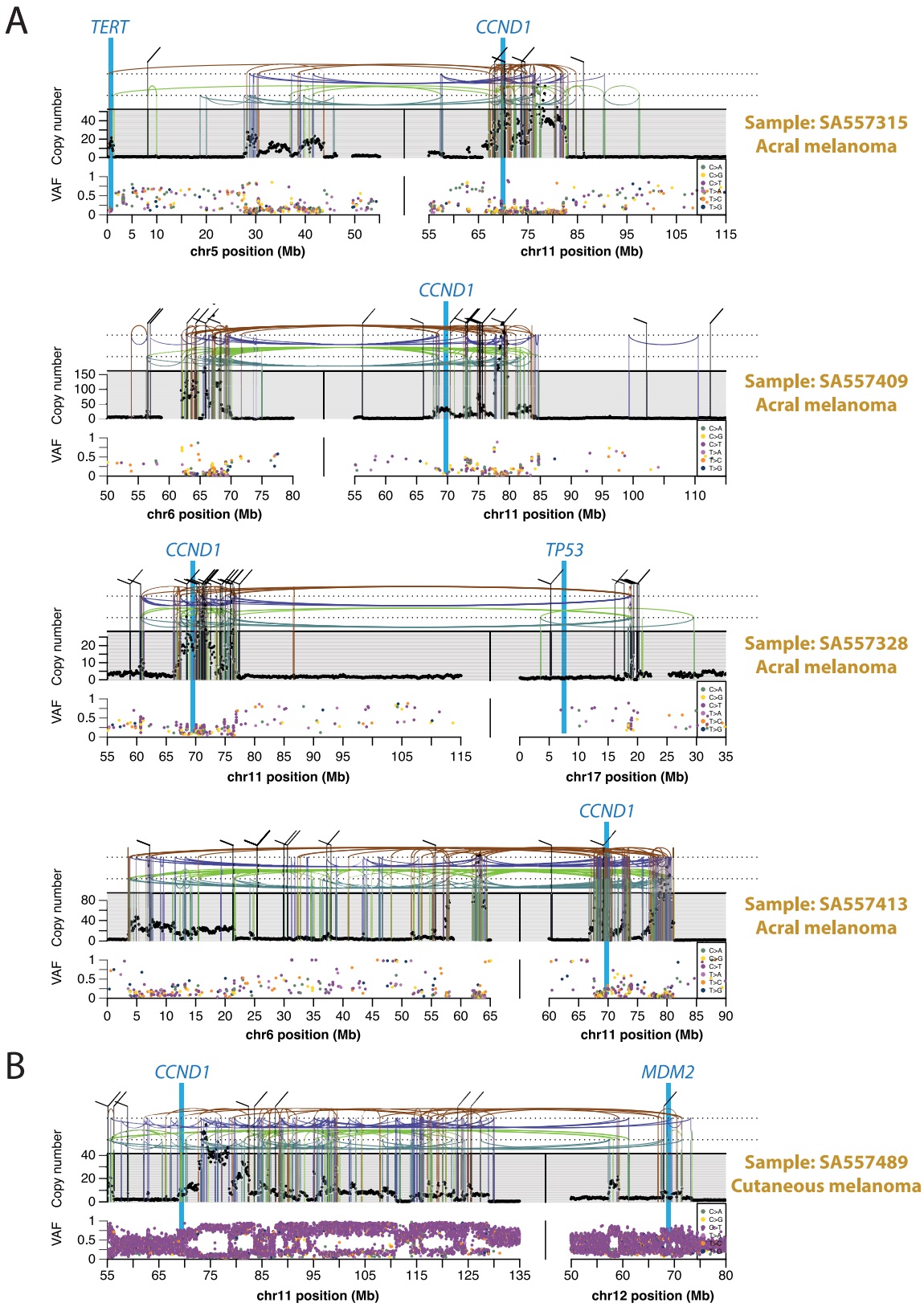


Extended Data Fig. 6 | Patterns of intense kataegis. a, Distribution of the tumour types (colour-coded as in Extended Data Fig. 3) of the samples in the top 5% of kataegis intensity in each of the four identified genome-wide patterns: non-APOBEC, replication stress, rearrangement-associated and the combination of the last two. **b**, **c**, Distribution of leading/lagging strand (**b**) and replication timing bias (**c**) for rearrangement-(in)dependent APOBEC kataegis, based on $n = 2{,}583$ tumours. *P* values were derived using a two-sided Mann–Whitney *U*-test. **d**, Example rainfall plots for each of the four identified kataegis patterns.



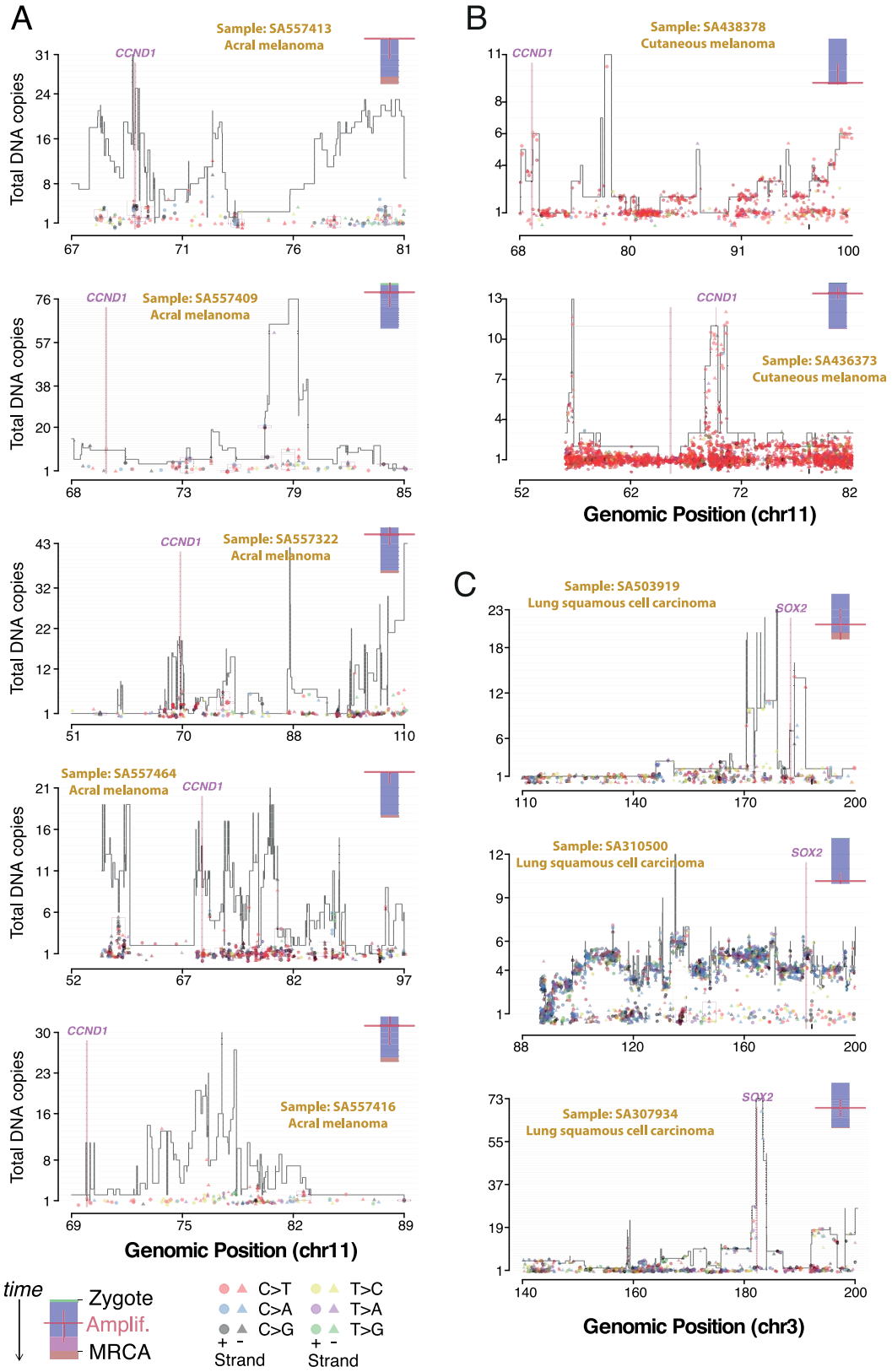
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Association of chromothripsis with covariates and driver events. a, Odds ratios per cancer type of containing chromothripsis in whole-genome duplicated versus diploid samples ($n = 2,583$ patients). ***$q < 0.001$; **$q < 0.01$; *$q < 0.05$. Two-sided hypothesis testing was performed using Fisher–Boschloo tests, corrected for multiple-hypothesis testing. **b**, Same as **a** for female versus male. **c**, Proportion of mutations explained by single-base substitution signature 1 and age at diagnosis in prostate cancer samples ($n = 210$ patients) with or without chromothripsis ($q < 0.05$). The early-onset prostate cancer project drives the signal and was sequenced at lower depth. For the box-and-whisker plots, the box denotes the interquartile range, with the median marked as a horizontal line. The whiskers extend as far as the range or 1.5× the interquartile range, whichever is less. Two-sided hypothesis testing was performed using Mann–Whitney *U*-tests. **d**, Counts of co-occurrence of chromothripsis with amplification (blue) and homozygous deletions (red) in driver regions: observed (thick line) versus randomized (shaded area and thin line). The cumulative number of drivers that were hit is plotted as a function of the number of times those drivers were hit. **e**, For each sample in which chromothripsis coincided with a driver event in those genes, we show the fold change in gene expression compared to the median expression of the gene in non-chromothripsis samples of the same cancer type, coloured by cancer type and shaped by the type of driver event. We show with added transparency the fold changes calculated the same way for samples with driver mutations hitting the same driver genes, but that had no evidence of chromothripsis. Analysis is based on $n = 1,222$ patients with RNA-sequencing data. **f**, Enrichment of co-occurrence of chromothripsis with driver events. The *x* axis shows the association of chromothripsis with a driver in a given cancer type compared with its rate of association with that driver in all other cancer types. The *y* axis shows the association of chromothripsis with a driver in a given cancer type compared with its rate of association with all other drivers in that type. Exact binomial tests are used and *P* values are corrected for multiple testing according to the Benjamini–Hochberg method.



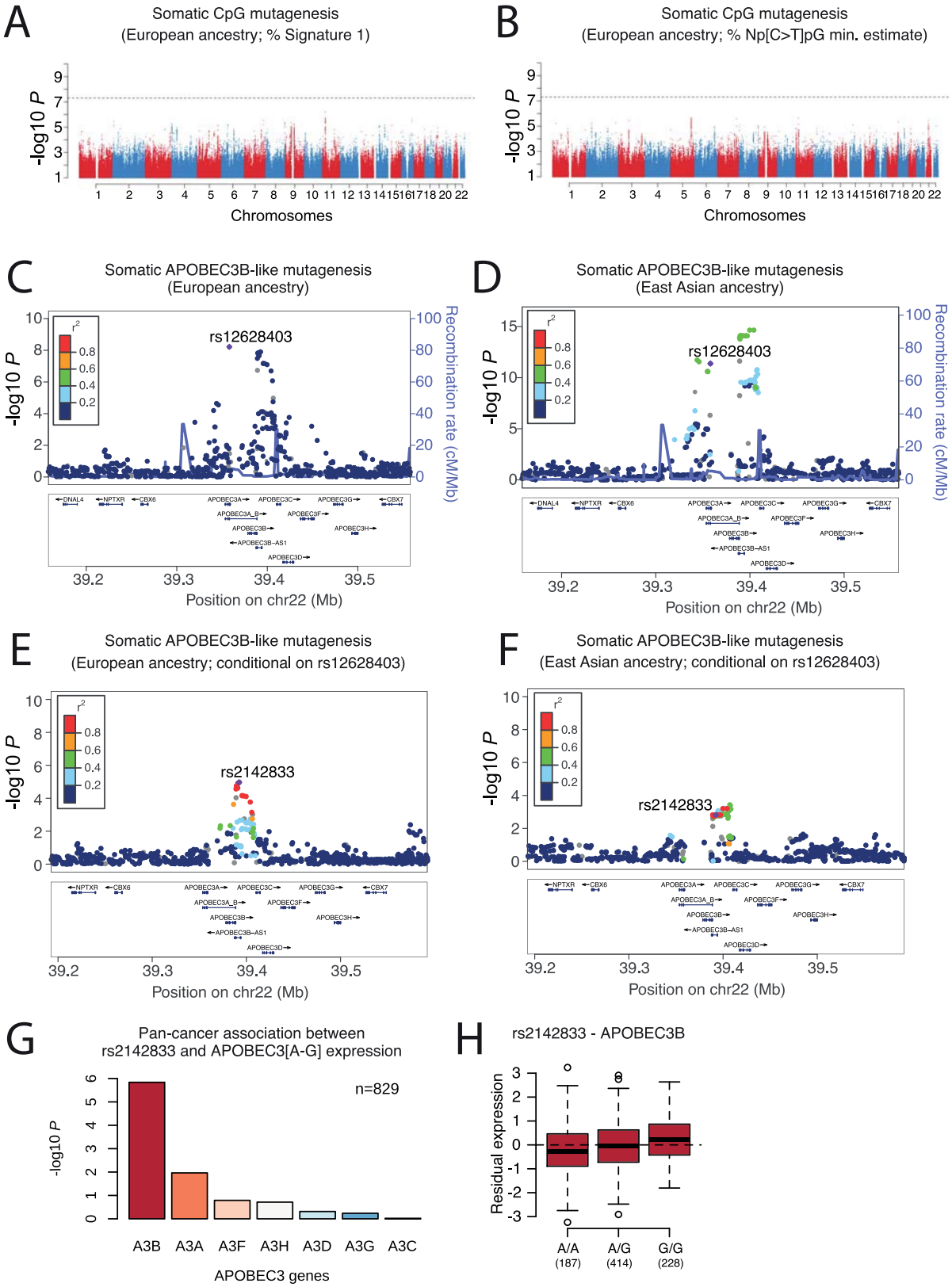
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Further examples of chromothripsis-induced amplification targeting multiple cancer-associated genes simultaneously in melanoma. a, Examples of amplifications that occurred early in the development of melanoma. The black points (top) represent copy-number estimates from individual genomic bins, with SVs shown as coloured arcs (translocation in black, deletion in purple, duplication in brown, tail-to-tail inversion in cyan and head-to-head inversion in green) that mostly demarcate copy-number changes. Bottom, the variant allele fractions of SNVs distributed along the relevant chromosomal region. The paucity of somatic mutations at high variant allele fractions in the most-heavily amplified regions indicates that these amplifications began very early in tumour evolution, before the lineage had had opportunity to acquire many SNVs. **b**, Example of an amplification that occurred late in melanoma development. The large numbers of somatic mutations at high variant allele fractions in the most-heavily amplified regions indicate that these amplifications began late in tumour evolution, after the lineage had already acquired many SNVs.



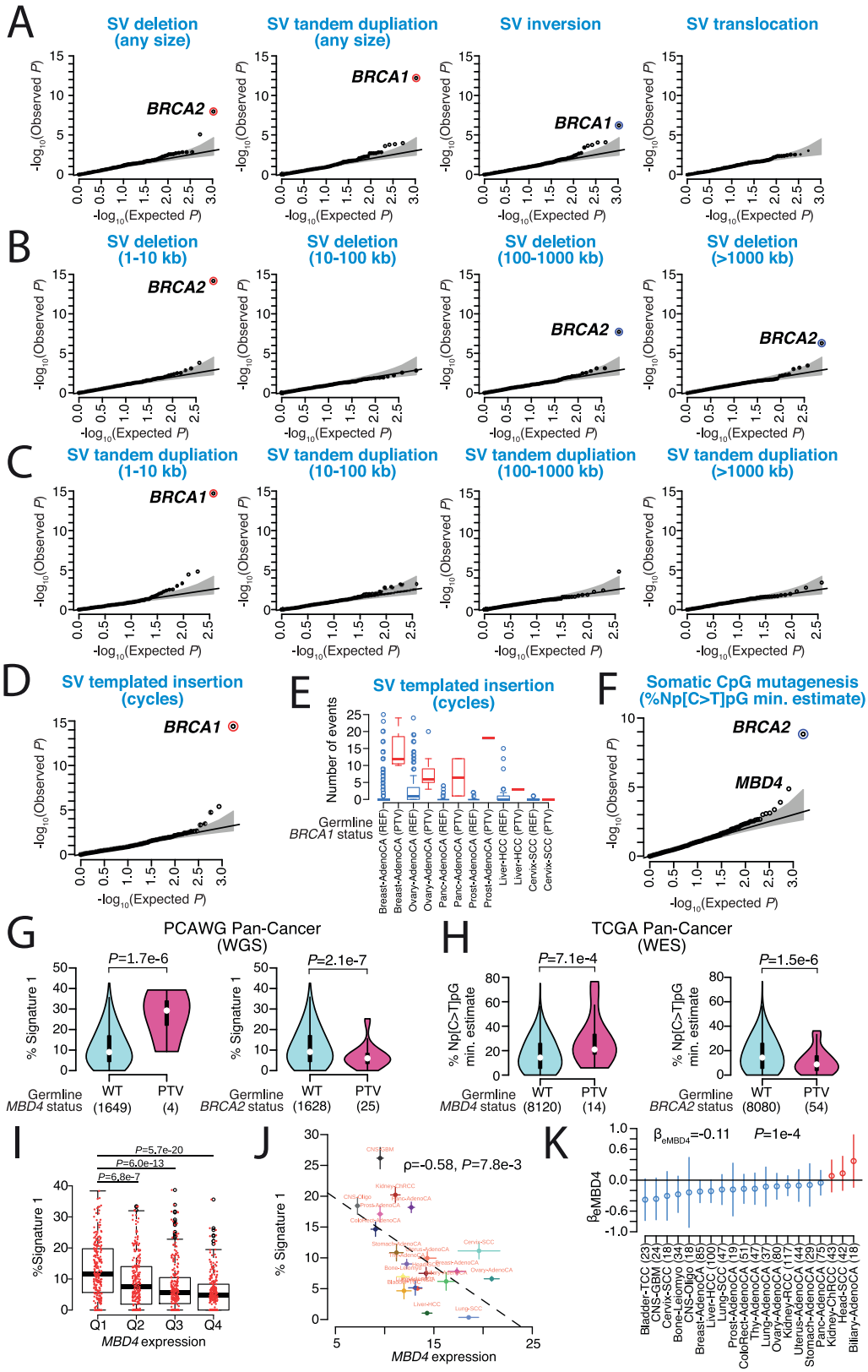
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Timing the amplifications after chromothripsis in molecular time for 10 representative cases. a, Copy-number plot of chromothriptic regions categorized as 'liposarc-like' in five acral melanomas with *CCND1* amplification. Segments indicate the copy number of the major allele. Points represent SNV multiplicities, that is, the estimated number of copies carrying each SNV, coloured by base change and shaped by strand. Small vertical arrows link SNVs to their corresponding copy-number segment. Kataegis foci are shown within black boxes and show typical strand specificities (all triangles or all circles), similar multiplicities and base changes of signatures 2 and 13 (red and black, respectively). A coloured bar (top right) represents the molecular timing of the amplification (red bar; high is early, low is late) and is coloured by the fraction of total SNVs assigned to the following timing categories: clonal [early], clonal mutations that occurred before duplications involving the relevant chromosome (including whole-genome duplications); clonal [late], clonal mutations that occurred after such duplications; and clonal [NA], mutations that occurred when no duplication was observed. **b**, Same as **a** in two cutaneous melanomas, one shows early amplification, the other late amplification. **c**, Same as **a**, **b**, for three lung squamous cell carcinomas and late amplification of *SOX2*.



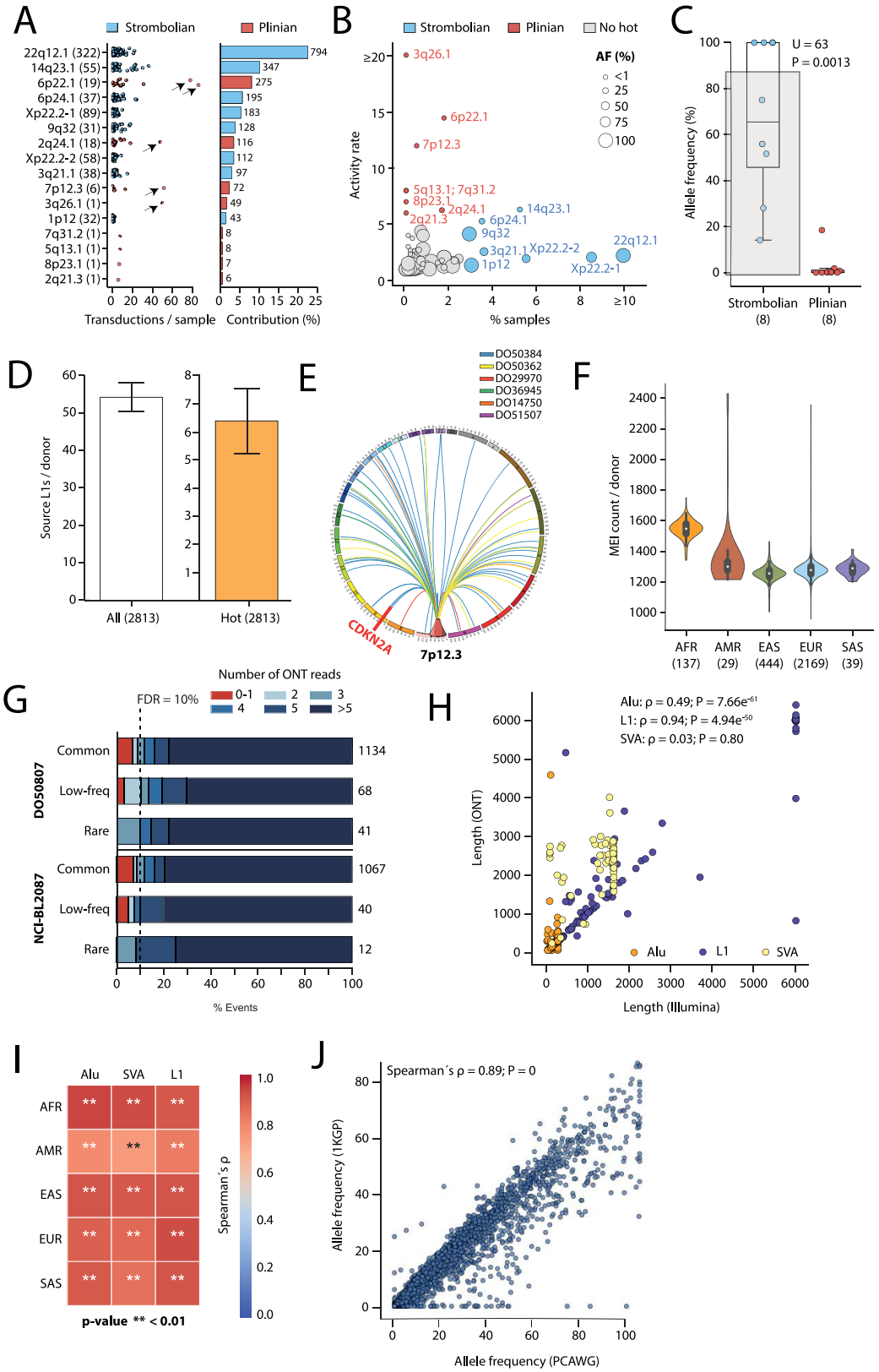
Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Association between common germline variants and endogenous mutational processes. Genome-wide association of somatic CpG mutagenesis in individuals of European ancestry ($n = 1,201$ patients) based on mutational signature analysis (**a**) and NpCpG motif analysis (**b**). Two-sided hypothesis testing was performed using PLINK v.1.9. To mitigate multiple-hypothesis testing, the significance threshold was set to genome-wide significance ($P < 5 \times 10^{-8}$). **c**, **d**, Locuszoom plot for somatic APOBEC3B-like mutagenesis association results, linkage disequilibrium and recombination rates around the genome-wide significant 22q13.1 locus in individuals with European (**c**) and East Asian (**d**) ancestry ($n = 1,201$ and 318 patients, respectively). Locuszoom plot for somatic APOBEC3B-like mutagenesis association results around the 22q13.1 locus in individuals with European (**e**) and East Asian (**f**) ancestry after conditioning on rs12628403. **g**, **h**, Association between rs2142833 and expression of *APOBEC3* genes in PCAWG tumour samples (adjusted for sex, age at diagnosis, histology and population structure in linear-regression models with two-sided hypothesis testing not corrected for multiple tests). For the box-and-whisker plot, the box denotes the interquartile range, with the median marked as a horizontal line. The whiskers extend as far as the range or 1.5× the interquartile range, whichever is less. Outliers are shown as points.



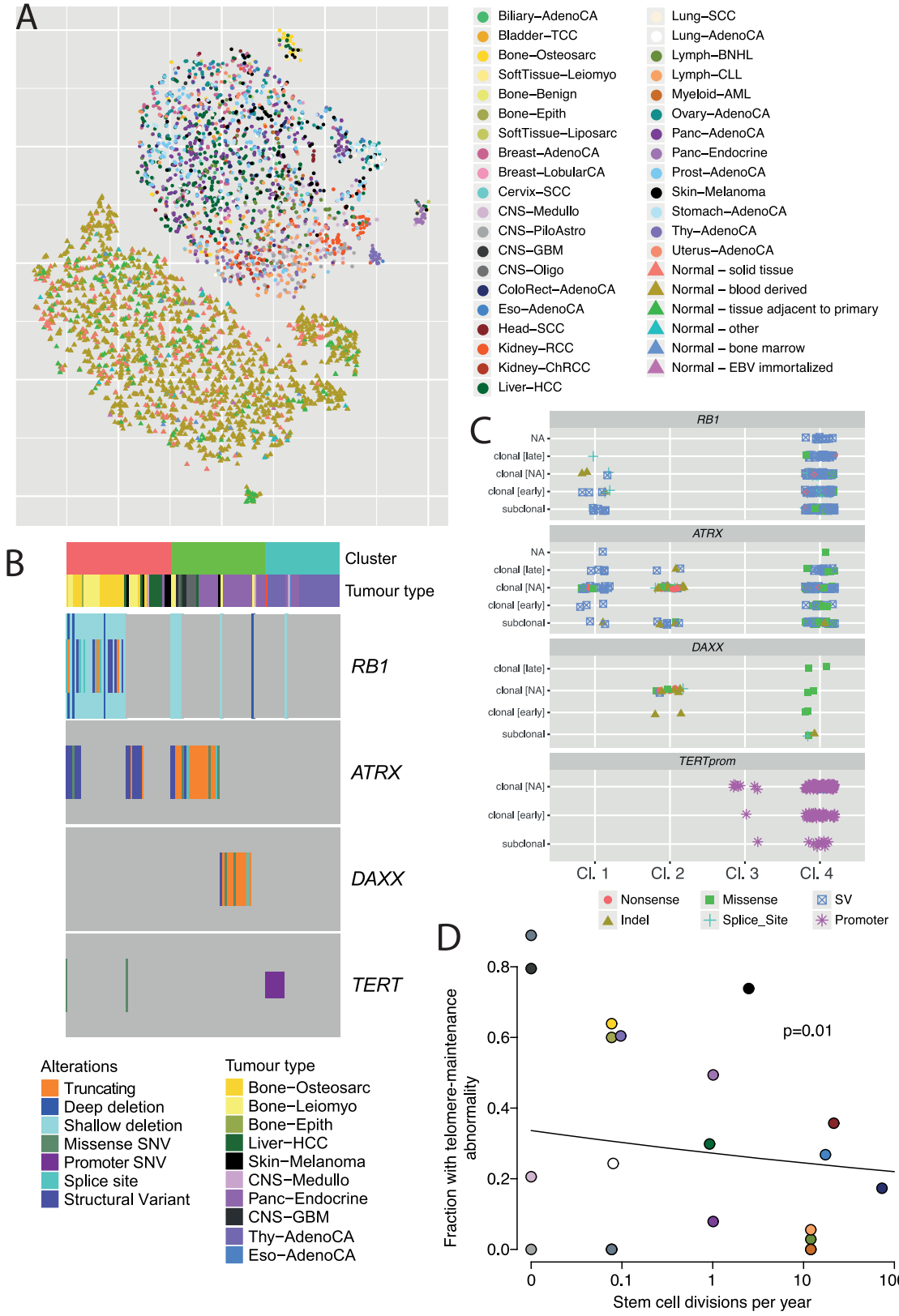
Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Association between rare germline PTVs in protein-coding genes and somatic mutational phenotypes. a–**d**, **f**, Data are based on two-sided rare-variant association testing across $n = 2{,}583$ patients, with a stringent P value threshold of $P < 2.5 \times 10^{-6}$ used to mitigate multiple-hypothesis testing (significant genes marked with coloured circles). Blue/red circles mark genes that decrease/increase somatic mutation rates. The black line represents the identity line that would be followed if the observed P values followed the null expectation, with the shaded area showing the 95% confidence intervals. **a**, *QQ* plots for the proportion of somatic SV deletions, tandem duplications, inversions and translocation in cancer genomes. **b**, *QQ* plots for the proportion of somatic SV deletions in cancer genomes stratified by four size groups (1–10 kb, 10–100 kb, 100–1,000 kb and >1,000 kb). **c**, *QQ* plots for the proportion of somatic SV tandem duplications in cancer genomes stratified by four size groups (1–10 kb, 10–100 kb, 100–1,000 kb and >1,000 kb). **d**, *QQ* plot for the presence or absence of somatic SV templated insertion (cycles) in cancer genomes. **e**, Number of SV-templated insertion cycles in PCAWG tumours with germline *BRCA1* PTVs. Only histological samples with at least one germline *BRCA1* PTV carrier are shown ($n = 1{,}095$ patients combined). The box denotes the interquartile range, with the median marked as a horizontal line. The whiskers extend as far as the range or 1.5× the interquartile range, whichever is less. Outliers are shown as points. **f**, *QQ* plot for somatic CpG mutagenesis in cancer genomes based on NpCpG motif analysis. **g**, Violin plots show estimated densities of the proportion of somatic CpG mutations in PCAWG donors with germline *MBD4* and *BRCA2* PTVs. The box denotes the interquartile range, with the median marked as a white point. The whiskers extend as far as the range or 1.5× the interquartile range, whichever is less. Two-sided hypothesis testing, not corrected for multiple testing, was performed using linear regression models. **h**, Replication of germline *MBD4* and *BRCA2* PTV associations with somatic CpG mutagenesis in TCGA whole-exome sequencing donors. Violin plots show the estimated density of the proportion of somatic CpG mutations in TCGA exomes with germline *MBD4* and *BRCA2* PTVs. The box denotes the interquartile range, with the median marked as a white point. The whiskers extend as far as the range or 1.5× the interquartile range, whichever is less. Two-sided hypothesis testing, not corrected for multiple testing, was performed using linear-regression models. **i**, Correlation between *MBD4* expression and somatic CpG mutagenesis in primary solid PCAWG tumours. Hypothesis testing was two-sided and not corrected for multiple testing, using linear-regression models. The box denotes the interquartile range, with the median marked as a horizontal line. The whiskers extend as far as the range or 1.5× the interquartile range, whichever is less. **j**, Data are mean ± s.e.m. across $n = 20$ tumour types. The dashed black line shows the fitted line to the data, estimated using linear-regression models. Hypothesis testing was two-sided and not corrected for multiple testing, using Spearman's rank correlations. **k**, *MBD4* effect sizes (open circles) with 95% confidence intervals (error bars) for individual cancer types were estimated using linear-regression analysis after (if available) accounting for sex, age at diagnosis (young/old) and ICGC project. Hypothesis testing was two-sided and not corrected for multiple testing.



Extended Data Fig. 12 | See next page for caption.

Extended Data Fig. 12 | Germline MEI call set. a, Left, dots show the number of transductions promoted by each hot element in individual samples. Arrows highlight retrotransposition burst. Right, the contribution of each hot locus is represented. The total number of transductions mediated by each source element is shown on the right. **b**, Source L1 activity rate (that is, measured as the average number of transductions mediated by an element) versus the percentage of samples with retrotransposition activity in which the germline element is active. For visualization purposes, extreme points observed for a source L1 with an activity rate of 49 and for a L1 active in 31% of the samples are shown at ≥20 and ≥10, respectively. **c**, Contrasting allele frequencies for Strombolian and Plinian source loci (sample sizes shown under each axis label). The box denotes the interquartile range, with the median marked as a white point. The whiskers extend as far as the range or 1.5× the interquartile range, whichever is less. Hypothesis testing was performed using two-sided Mann–Whitney *U*-tests without correction for multiple tests. **d**, Numbers of active and hot source L1 elements per donor. Data are mean ± s.d. number of elements per donor. **e**, The novel Plinian source element on 7p12.3 mediates 72 transductions among only 6 cancer samples. This generates a transduction that induces the deletion of the tumour-suppressor gene *CDKN2A*. **f**, Violin plots show the estimated number of distinct germline MEI alleles per PCAWG donor. The box denotes the interquartile range, with the median marked as a white point. The whiskers extend as far as the range or 1.5× the interquartile range, whichever is less. Donors are grouped according to their genetic ancestry: AFR, African; AMR, admixed American; EAS, East Asian; EUR, European; SAS, South Asian. Sample sizes are shown under each axis label. **g**, For each type of MEI (L1, Alu and SVA) identified both in PCAWG and in the 1000 Genomes Project (1KGP), the correlations between allele frequency estimates per ancestry derived from both projects are displayed in a blue (0) to red (1) coloured gradient. $n = 2,583$ PCAWG patients. Two-sided hypothesis testing was performed using Spearman's rank correlations without correction for multiple tests. **h**, Example correlation between MEI allele frequencies derived from PCAWG and the 1000 Genomes Project for individuals with European ancestry ($n = 1,201$ patients in PCAWG). Two-sided hypothesis testing was performed using Spearman's rank correlations without correction for multiple tests. **i**, Evaluation of TraFiC-mem false-discovery rate on a liver hepatocellular carcinoma sample (DO50807) and a cell line (NCI-BL2087) sequenced using single-molecule sequencing with MinION (Oxford Nanopore). For each allele frequency bin (common, >5%; low frequency, 1–5%; rare, <1%), the percentage of events supported by *N* long reads is represented (*N* ranges from 0–1 to more than 5). MEIs supported by at least two Nanopore reads were considered to be true positives (blue palette) and were classified as false positives (red) otherwise. The total number of germline MEIs per allele frequency bin is shown on the right. **j**, Correlation between predicted MEI lengths from Illumina and Nanopore data. Two-sided hypothesis testing was performed using Spearman's rank correlations without correction for multiple testing.



Extended Data Fig. 13 | See next page for caption.

Extended Data Fig. 13 | Different mechanisms of telomere lengthening in cancer. a, Scatter plot showing the four clusters of tumour-specific telomere patterns identified across PCAWG samples, together with the clusters of matched normal samples, generated by *t*-distributed stochastic neighbour embedding. Circles represent tumour samples and triangles represent matched normal samples. Points are coloured by tissue of origin. Data are based on $n = 2,518$ tumour samples and their matched normal samples. **b**, Patterns of comutation of the relevant driver mutations across individual patients. Columns in plot represent individual patients, coloured by type of abnormality observed. **c**, Distribution of clonality of driver mutations in genes relevant to telomere maintenance across clusters. Clonal [early], clonal mutations that occurred before duplications involving the relevant chromosome (including whole-genome duplications); clonal [late], clonal mutations that occurred after such duplications; and clonal [NA], mutations that occurred when no duplication was observed. **d**, Relationship between the estimated number of stem cell divisions per year and rate of telomere maintenance abnormalities across tumour types. The analysis uses data on estimated rates of stem cell division per year across $n = 19$ tissue types previously collated from the literature[82]. Tumour types are coloured according to the scheme shown in Extended Data Fig. 3. Two-sided hypothesis testing was performed using likelihood ratio tests on Poisson regression models with no correction for multiple tests.

Extended Data Table 1 | Overview of the tumour types included in PCAWG project

Organ	Abbreviation	Included Subtypes	Cases	Sex		Age	
Neural Crest			**Num.**	**F**	**M**	**Med.**	**10th-90th**
CNS	CNS-GBM	Glioblastoma	41	13	28	60	43-72
CNS	CNS-Medullo	Medulloblastoma and variants	146	67	79	9	3-28
CNS	CNS-Oligo	Oligodendroglioma	18	9	9	41	21-62
CNS	CNS-PiloAstro	Pilocytic astrocytoma	89	47	42	8	2-17
Skin	Skin-Melanoma	Malignant melanoma	107	38	69	57	37-78
Endoderm							
Biliary	Biliary-AdenoCA	Papillary cholangiocarcinoma	34	15	19	64	53-76
Bladder	Bladder-TCC	Transitional cell carcinoma	23	8	15	65	52-80
Colon/Rectum	ColoRect-AdenoCA	Adenocarcinoma; Mucinous adeno.	60	30	30	67	46-81
Oesophagus	Eso-AdenoCA	Adenocarcinoma	98	14	84	70	56-79
Liver	Liver-HCC	Hepatocellular carcinoma; Comb. HCC/cholangio	317	89	228	67	50-78
Lung	Lung-AdenoCA	Adenocarcinoma; Adenocarcinoma *in situ*	38	20	18	66	47-77
Lung	Lung-SCC	Squamous cell carcinoma; Basaloid SCC	48	10	38	68	54-77
Pancreas	Panc-AdenoCA	Adeno.; Acinar cell Ca.; Mucinous adeno.	239	119	120	67	50-79
Pancreas	Panc-Endocrine	Neuroendocrine carcinoma	85	30	55	59	38-75
Prostate	Prost-AdenoCA	Adenocarcinoma	210	0	210	59	47-71
Stomach	Stomach-AdenoCA	Adenocarcinoma; Mucinous; Papillary; Tubular	75	18	57	65	47-79
Thyroid	Thy-AdenoCA	Adenocarcinoma; Columnar cell; Follicular type	48	37	11	51	26-75
Mesoderm							
Bone/Soft Tissue	Bone-Benign	Osteoblastoma; Osteofibrous dysplasia	7	4	3	18	12-30
Bone/Soft Tissue	Bone-Benign	Chondroblastoma; Chrondromyxoid fibroma	9	2	7	16	14-38
Bone/Soft Tissue	Bone-Epith	Adamantinoma; Chordoma	10	4	6	60	37-67
Bone/Soft Tissue	Bone-Osteosarc	Osteosarcoma	38	20	18	20	9-58
Bone/Soft Tissue	SoftTissue-Leiomyo	Leiomyosarcoma	15	10	5	61	51-78
Bone/Soft Tissue	SoftTissue-Liposarc	Liposarcoma	19	5	14	n/a	n/a
Cervix	Cervix-AdenoCA	Adenocarcinoma	2	2	0	39	33-46
Cervix	Cervix-SCC	Squamous cell carcinoma	18	18	0	39	25-58
Head/Neck	Head-SCC	Squamous cell carcinoma	57	10	47	53	34-71
Kidney	Kidney-ChRCC	Adenocarcinoma, chromophobe type	45	19	26	47	34-69
Kidney	Kidney-RCC	Clear cell adenocarcinoma; papillary type	144	54	90	60	48-75
Lymphoid	Lymph-BNHL	Burkitt; Diffuse large B-cell; Follicular; Marginal	107	51	56	57	10-74
Lymphoid	Lymph-CLL	Chronic lymphocytic leukaemia	95	31	64	62	46-78
Myeloid	Myeloid-AML	Acute myeloid leukaemia	10	3	7	50	35-56
Myeloid	Myeloid-MDS	Myelodysplastic syndrome	2	1	1	76	74-77
Myeloid	Myeloid-MPN	Myeloproliferative neoplasm	26	14	12	56	38-75
Ovary	Ovary-AdenoCA	Adenocarcinoma; Serous cystadenocarcinoma	113	113	0	60	48-74
Uterus	Uterus-AdenoCA	Adeno., endometrioid; Serous cystadeno.	51	51	0	69	57-81
Ectoderm							
Breast	Breast-AdenoCA	Infiltrating duct carcinoma; Medullary; Mucinous	198	197	1	56	39-76
Breast	Breast-DCIS	Duct micropapillary carcinoma	3	3	0	55	43-60
Breast	Breast-LobularCA	Lobular carcinoma	13	13	0	53	42-69
Total			**2658**	**1189**	**1469**	**59**	**21-76**

Adeno., adenocarcinoma; Ca., carcinoma; Comb., combined; F, female; HCC, hepatocellular carcinoma; M, male; Med, median; 10–90th, 10–90th centiles; SCC, squamous cell carcinoma.

Extended Data Table 2 | Ethical considerations of genomic cloud computing

Ethical Considerations of Genomic Cloud Computing

The PCAWG project represents the first large-scale use of distributed cloud computing in genomics. The project involved the movement of large quantities of personal health information across multiple legal jurisdictions and responsible use of this data by several hundred international researchers. Donor consents were written to explicitly allow for broad research use of the data and for international data sharing. PCAWG was granted permission by the leads of each of the tumour data providers to store, analyse and distribute the data on academic and/or commercial compute clouds.

To ensure that the PCAWG personal data were handled in a manner consistent with the donor consents, authorised representatives of each of the academic clouds and high-performance computing facilities signed a commitment not to access controlled tier data beyond the minimum needed to administer it. We negotiated similar contractual terms with commercial cloud partners. Prior to accessing the data, each PCAWG researcher was required to obtain local Institutional Review Board approval for their proposed analytic projects, and obtained controlled tier authorisation from dbGaP (National Center for Biotechnology Information) and the ICGC DACO (Centre of Genomics and Policy at McGill University). To handle the data securely, we encrypted it while in motion and at rest. We used a central authentication and digital token generating system to enforce a strong data access protocol that required researchers to provide their TCGA and/or ICGC credentials prior to accessing controlled tier data. No data breach or other compromise of donor confidentiality is known to have occurred over the course of the PCAWG project, despite its extensive use of cloud computing.

Extended Data Table 3 | Scientific output using PCAWG data, in bite-size chunks

Scientific area	Key findings	Citation
Driver mutations		
Discovery of non-coding drivers	● Estimated ~10-fold more coding than non-coding driver point mutations. ● Variation in point mutation density in non-coding regions influenced more by mutational processes than selection.	4
Drivers by pathways and networks	● Both coding and non-coding alterations contribute to cancer pathways. ● Some pathways, such as RNA splicing, are primarily driven by non-coding mutations.	16
Evolution and heterogeneity		
Timing of cancer evolution	● Each tumour type has a distinct pattern of early and late-occurring driver events. ● Earliest somatic mutations may occur decades prior to diagnosis, providing opportunities for early diagnosis. ● Intra-tumour heterogeneity is widespread and tumour subclones contain drivers that are under positive selection.	7
Structural variants		
Patterns of structural variation	● Replication-based mechanisms of genome rearrangement frequent in many cancers, often causing driver structural variants. ● 16 signatures of SV, including break-and-ligate patterns and copy-and-insert patterns, varying by size range, replication timing, tumour type and patient.	6
Functional consequence of structural variation	● 52 regions with recurrent structural breakpoints and 90 recurrently fused pairs of loci show evidence of positive selection. ● Oncogenic fusions are shaped by juxtaposition of proto-oncogenes with tissue-specific regulatory elements.	4
Patterns of retrotransposition	● Many flavours of somatic retrotransposition in many cancers: LINE element mobilisation; transductions, pseudogenes, Alu elements. ● Retrotranspositions can induce genomic instability, including large deletions and breakage-fusion-bridge cycles amplifying cancer genes.	10
Chromothripsis	● Chromothripsis pervasive across cancers, with frequency >50% in several tumour types. ● Replicative processes and templated insertions contribute to rearrangement.	18
Mutational signatures		
Signatures of point mutations	● >70 distinct mutational signatures, encompassing SNVs, doublet subs and indels. ● Multiple signatures from unknown processes of DNA damage, repair and replication.	5
Mutation distribution across genome	● Uneven distribution of somatic mutations and structural variants across the genome explained by epigenetic state of tissue, cell of origin and topological associated domains. ● Can be used to identify a tumour's type and presumed tissue/cell of origin.	11,12,15
Transcriptional consequences of somatic mutation		
RNA effects of somatic mutation	● Genomic basis for RNA alterations across ~1200 tumours, including quantitative trait loci, allele specific expression and alternative splicing. ● Link between mutational signatures and expression; classification of gene fusions; identification of genes recurrently altered at RNA level.	8,9
Others		
Tumour subtypes from genome sequencing	● Genomic distribution of somatic mutations, mutational signatures and driver mutations accurately distinguish major tumour types of primaries and metastases.	12
Mitochondrial DNA mutations	● Somatic mitochondrial truncating mutations frequent in certain cancer types, associated with activation of critical signaling pathways.	14
Telomere biology and sequences	● Activating *TERT* promoter mutations are the single most frequent non-coding driver. ● In *ATRX*/*DAXX*-mutant tumours, aberrant telomere variant repeat distribution is common.	4,13

Key findings are described further in associated papers[4-18].

nature research

Corresponding author(s): Peter Campbell, Joshua Stuart, Gad Getz, Jan Korbel, Lincoln Stein

Last updated by author(s): Oct 23, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data and metadata were collected from International Cancer Genome Consortium (ICGC) consortium members using custom software packages designed by the ICGC Data Coordinating Centre. The general-purpose core libraries and utilities underlying this software have been released under the GPLv3 open source license as the "Overture" package and are available at https://www.overture.bio. Other data collection software used in this effort, such as ICGC-specific portal user interfaces, are available upon request to contact@overture.bio.
Data analysis	The workflows executing core WGS alignment, QC and variant-calling software are packaged as executable Dockstore images and available at: https://dockstore.org/search?labels.value.keyword=pcawg&searchMode=files. Individual software components are as follows: BWA-MEM v0.78.8-r455; DELLY v0.6.6; ACEseq v1.0.189; DKFZ somatic SNV workflow v1.0.132-1; Platypus v0.7.4; ascatNgs v1.5.2; BRASS v4.012; grass v1.1.6; CaVEMan v1.50; Pindel v1.5.7; ABSOLUTE/JaBbA v1.5; SvABA 2015-05-20; dRanger 2016-03-13; BreakPointer 2015-12-22; MuTect v1.1.4; MuSE v1.0rc; SMuFIN 2014-10-26; OxoG 2016-4-28; VAGrENT v2.1.2; ANNOVAR v2014Nov12; VariantBAM v2017Dec12; SNV-Merge v2017May26; SV-MERGE v2017Dec12; DKFZ v2016Dec15; Chromothripsis Explorer v1.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

WGS somatic and germline variant calls, mutational signatures, subclonal reconstructions, transcript abundance, splice calls and other core data generated by the ICGC/TCGA Pan-cancer Analysis of Whole Genomes Consortium are available for download at https://dcc.icgc.org/releases/PCAWG. Additional information on accessing the data, including raw read files, can be found at https://docs.icgc.org/pcawg/data/. In accordance with the data access policies of the ICGC and TCGA

projects, most molecular, clinical and specimen data are in an open tier which does not require access approval. To access potentially identification information, such as germline alleles and underlying sequencing data, researchers will need to apply to the TCGA Data Access Committee (DAC) via dbGaP (https://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login) for access to the TCGA portion of the dataset, and to the ICGC Data Access Compliance Office (DACO; http://icgc.org/daco) for the ICGC portion. In addition, to access somatic single nucleotide variants derived from TCGA donors, researchers will also need to obtain dbGaP authorization

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We compiled an inventory of matched tumour/normal whole cancer genomes in the ICGC Data Coordinating Centre. Most samples came from treatment-naïve, primary cancers, but there were a small number of donors with multiple samples of primary, metastatic and/or recurrent tumours. Our inclusion criteria were: (i) matched tumour and normal specimen pair; (ii) a minimal set of clinical fields; and (iii) characterisation of tumour and normal whole genomes using Illumina HiSeq paired-end sequencing reads. We collected genome data from 2,834 donors, representing all ICGC and TCGA donors that met these criteria at the time of the final data freeze in autumn 2014. No statistical methods were used to predetermine sample size.
Data exclusions	After quality assurance, data from 176 donors were excluded as unusable. Reasons for data exclusions included inadequate coverage, extreme bias in coverage across the genome, evidence for contamination in samples and excessive sequencing errors (for example, through 8-oxoguanine). Exclusion criteria were pre-determined.
Replication	In order to evaluate the performance of each of the mutation-calling pipelines and determine an integration strategy, we performed a large-scale deep sequencing validation experiment. We selected a pilot set of 63 representative tumour/normal pairs, on which we ran the three core pipelines, together with a set of 10 additional somatic variant-calling pipelines contributed by members of the SNV Calling Working Group. Overall, the sensitivity and precision of the consensus somatic variant calls were 95% (CI90%: 88-98%) and 95% (CI90%: 71-99%) respectively for SNVs. For somatic indels, sensitivity and precision were 60% (34-72%) and 91% (73-96%) respectively. Regarding SVs, we estimate the sensitivity of the merging algorithm to be 90% for true calls generated by any one caller; precision was estimated as 97.5% - that is, 97.5% of SVs in the merged SV call-set have an associated copy number change or balanced partner rearrangement.
Randomization	No randomisation was performed - this was a descriptive study, not an experimental study.
Blinding	No blinding was undertaken - this was a descriptive study, not an experimental study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Patient-by-patient clinical data are provided in the marker paper for the PCAWG consortium (Extended Data Table 1 of that manuscript). Demographically, the cohort included 1,469 males (55%) and 1,189 females (45%), with a mean age of 56 years (range, 1-90 years). Using population ancestry-differentiated single nucleotide polymorphisms (SNPs), the ancestry distribution was heavily weighted towards donors of European descent (77% of total) followed by East Asians (16%), as expected for large contributions from European, North American and Australian projects. We consolidated histopathology descriptions of the tumour samples, using the ICD-0-3 tumour site controlled vocabulary. Overall, the PCAWG data set comprises 38 distinct tumour

	types. While the most common tumour types are included in the dataset, their distribution does not match the relative population incidences, largely due to differences among contributing ICGC/TCGA groups in numbers sequenced.
Recruitment	Patients were recruited by the participating centres following local protocols. Samples obtained had to meet criteria on amount of tumour DNA available, meaning that the cohort is potentially somewhat biased towards larger tumours. Otherwise, we anticipate no major recruitment biases.
Ethics oversight	The Ethics oversight for the PCAWG protocol was undertaken by the TCGA Program Office and the Ethics and Governance Committee of the ICGC. Each individual ICGC and TCGA project that contributed data to PCAWG had their own local arrangements for ethics oversight and regulatory alignment.

Note that full information on the approval of the study protocol must also be provided in the manuscript.