CancerTracer: a curated database for intrapatient tumor heterogeneity

Chen Wang[1,†], Jian Yang[1,†], Hong Luo[1], Kun Wang[1], Yu Wang[1], Zhi-Xiong Xiao[1], Xiang Tao[2,*], Hao Jiang[3,*] and Haoyang Cai [1,*]

[1]Center of Growth, Metabolism and Aging, Key Laboratory of Bio-Resources and Eco-Environment, College of Life Sciences, Sichuan University, Chengdu 610064, China, [2]College of Life Science, Sichuan Normal University, Chengdu 610101, China and [3]Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610041, China

ABSTRACT

Comprehensive genomic analyses of cancers have revealed substantial intrapatient molecular heterogeneities that may explain some instances of drug resistance and treatment failures. Examination of the clonal composition of an individual tumor and its evolution through disease progression and treatment may enable identification of precise therapeutic targets for drug design. Multi-region and single-cell sequencing are powerful tools that can be used to capture intratumor heterogeneity. Here, we present a database we've named CancerTracer (http://cailab.labshare.cn/cancertracer): a manually curated database designed to track and characterize the evolutionary trajectories of tumor growth in individual patients. We collected over 6000 tumor samples from 1548 patients corresponding to 45 different types of cancer. Patient-specific tumor phylogenetic trees were constructed based on somatic mutations or copy number alterations identified in multiple biopsies. Using the structured heterogeneity data, researchers can identify common driver events shared by all tumor regions, and the heterogeneous somatic events present in different regions of a tumor of interest. The database can also be used to investigate the phylogenetic relationships between primary and metastatic tumors. It is our hope that Cancer-Tracer will significantly improve our understanding of the evolutionary histories of tumors, and may facilitate the identification of predictive biomarkers for personalized cancer therapies.

INTRODUCTION

Over the past few years, advances in next-generation sequencing and molecular diagnostics have significantly increased our understanding of the heterogeneities and evolutions of mutations within individual tumors (1–3). The process of conversion from a normal to a malignant cell is known to occur through the sequential accumulation of alterations in oncogenes and tumor suppressor genes (4,5). Over the course of tumor initiation and progression, cancers continue to evolve and may ultimately give rise to heterogeneous mixtures of tumor cells with distinct molecular signatures (6,7). Tumor heterogeneity has been found in virtually all types of cancer, and can be broadly divided into interpatient and intrapatient heterogeneity. Interpatient tumor heterogeneity, which results from patient-specific factors, has long been known about, and has been extensively studied (2,8). Intrapatient tumor heterogeneity refers to heterogeneity among the tumor cells of an individual tumor or patient. This heterogeneity may manifest as spatial heterogeneity, which reflects the uneven distribution of genetically diverse tumor subclones across and within disease sites, or as temporal heterogeneity, which presents as differences that develop within a single tumor over time (9–15). Tumor heterogeneity at an individual patient level has important clinical implications (16,17). Accumulating evidence suggests that the presence of low-frequency subclones harboring resistance mutations may contribute to treatment failure (18,19). Furthermore, small sub clones undetectable within the primary tumor at the time of diagnosis may later become responsible for local or distant metastases in the same patient. Therefore, understanding intrapatient tumor heterogeneity is essential for predicting therapeutic responses, and for developing effective therapies.

Multi-region sequencing and single-cell sequencing are emerging technologies that can effectively capture intratumor heterogeneity. Multi-region sequencing can be used to sample different spatial regions within a single tumor, in order to elucidate the complex clonal architectures of cancers (20–22). The clonal and subclonal compositions of individual tumors can be deduced by studying multiple biopsies from the same tumor. This information can then be used to construct distance-based phylogenetic trees that reveal the evolutionary trajectories of different subclones in each patient. Clonal mutations representing early or initiating genetic events are depicted on the trunk of a tumor's phylogenetic tree, whereas subclonal mutations reflecting later or progression events, are shown on the branches. The value of multi-region sequencing for phylogenetic reconstruction was first demonstrated in renal cancer, and, more recently, in a number of other cancer types (9,12,23–26). Single-cell sequencing is another innovative and informative technology that facilitates the discovery of clonal versus subclonal alterations in tumors, at even higher levels of resolution (27–30).

Rapid advances in DNA sequencing technologies have resulted in the accumulation of tumor genomic data at an unprecedented speed. A number of excellent resources on cancer genomics and mutations are available. The Cancer Genome Atlas (TCGA) (31) and the International Cancer Genome Consortium (ICGC) (32), for example, provide comprehensive sets of genomic data and complete clinical information of patients. Several web resources focus on curating and annotating mutations and variants, such as the Catalogue Of Somatic Mutations In Cancer (COSMIC) (33), ClinVar (34), OncoBase (35), BioMuta and BioXpress (36), IntOGen (37), and PreMedKB (38). SEECancer is a database that aims to provide information on cancer evolutionary stage-specific somatic events and their temporal orders (39). These resources have greatly facilitated the discoveries of novel cancer-associated genes and have greatly advanced the field of cancer genomics. However, none of the existing tumor genome repositories provide well-documented phylogenies and 'road-maps' of each tumor, since these programs were not designed to address the heterogeneity component of cancer. Generally, these databases use platforms that characterize tumors in bulk, giving results that average across all tumor clones. Therefore, a resource for systematically tracking and annotating the evolution of individual cancer genomes over space and time is necessary in order to decipher the extent and pattern of tumor heterogeneity at the single-patient level.

To address this gap, we developed CancerTracer, a manually curated database to track and characterize the evolutionary trajectories of tumor growths in patients. Patient-specific phylogenetic trees were constructed based on available alteration data. Somatic mutations and Copy Number Alterations (CNAs) were included in the database, as two principal factors driving heterogeneity in cancer. Currently, the database contains over 6000 tumor samples from 1548 patients across 45 different cancer types. We hope that this elaborate database will serve as a unique resource for researchers in a broad range of cancer-related fields to explore and understand intrapatient tumor heterogeneity.

DATA COLLECTION AND PROCESSING

Data collection

CancerTracer aims to provide a comprehensive resource for documenting tumor heterogeneities in individual patients. In order to retrieve articles that are suitable for inclusion in our database, we first searched the PubMed database using a set of topically relevant keywords, including: 'intratumor heterogeneity', 'spatial heterogeneity', 'temporal heterogeneity', 'genomic heterogeneity', 'primary heterogeneity', 'metastasis heterogeneity', 'multi-region sequencing' and 'single cell sequencing'. Each retrieved article was assessed independently by two investigators for inclusion or exclusion. The citation-tracking capability of Google Scholar was applied to highly relevant publications in order to broaden the search. Of the >1000 articles retrieved as potentially relevant, 145 were identified as eligible for our subsequent data curation process. These studies were classified into two categories according to sampling strategies: (i) intratumor heterogeneity studies—sampling of multiple regions within a single primary lesion (which can itself be further divided into spatial and temporal types) and (ii) intertumor heterogeneity studies—sampling of primary tumors as well as either single or multiple metastatic sites. Somatic mutations and CNAs were manually extracted from tables, figures, texts and supplementary materials of the qualified articles. Clinical data of the patients, sampling sites and information on phylogenetic trees was also collected if available. All data were integrated and organized at the patient level. The final data were confirmed by at least two independent experts.

Intratumor heterogeneity data annotation

Intratumor heterogeneity can manifest as spatial and temporal patterns of genetic diversity. Spatial separation of subclones within an individual tumor is usually investigated by multi-region sequencing. We collected data on somatic mutations and CNAs from publications, to add to the CancerTracer database. According to the original descriptions in the literature, these genomic alterations were classified into three categories: (i) trunk mutations, present in all intratumoral regions; (ii) branch mutations, shared by some regions but not all and (iii) private mutations, observed in only one region. This type of data is frequently used to construct phylogenetic trees, which can intuitively present phylogenetic relationships of the tumor regions (20–22) (Figure 1). These trees are usually rooted at the germline DNA sequence, determined by sequencing of DNA from normal tissues or patient blood samples. Trunk and branch lengths are proportional to the number of nonsynonymous mutations acquired. We adapted and redrew the original phylogenetic trees, if available, for the database. In about 75% of publications, the authors provided scale bars on the dendrograms or indicated the total numbers of trunk, branch and private mutations in the main text. We present this information in the figure of phylogenetic tree. In some cases, driver genes or CNAs that contributed to the evolution were indicated in the branches or trunk, and we also kept this information in the redrawn dendrograms. Some studies pro-
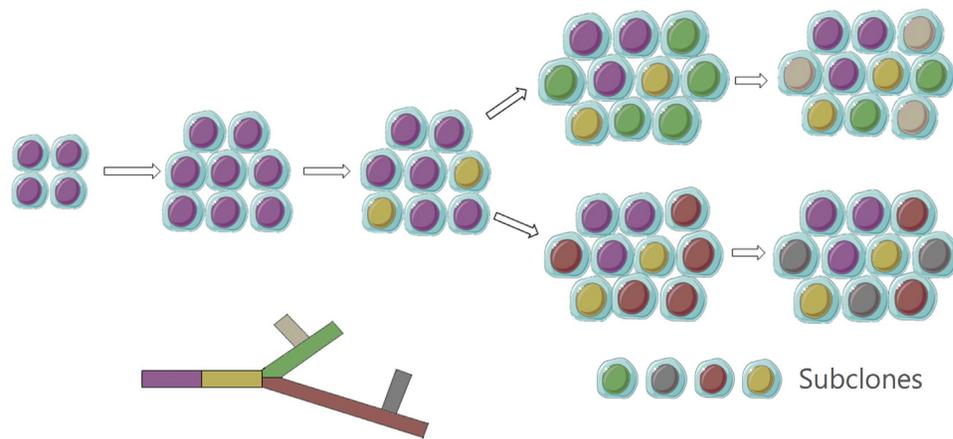
Figure 1. Schematic representation of tumor clonal evolution and the construction of phylogenetic tree. Distinct subclones are designated with distinct colors. In the phylogenetic tree, the trunk and branch lengths are proportional to the number of alterations acquired.

vided detailed information on gene variations such as substitution, insertion, deletion and duplication. We converted such data into a uniform format to facilitate the comparison of data from different studies. In order to allow users to access the potential medical values of genes of interest, we downloaded the drug-gene annotation and therapeutic targets annotation from the DGIdb (40) and the project Score database (41), respectively. CCLE (42), GDSC (43), CellMiner (44) and NCI-60 (45) contain drug resistance data. We provide links to these websites where appropriate. To systematically annotate the functional impact of somatic mutations in proteins, we employed ANNOVAR (46) and PolyPhen2 (47) to perform the analysis. About 88.78% mutations were successfully annotated by ANNOVAR or PolyPhen2, and 84.62% of non-synonymous point mutations were annotated by SIFT and/or PolyPhen2. A small number of mutations were not able to be annotated due to the following possible reasons: (i) some publications did not provide genome assembly of mutations; (ii) irregular mutation descriptions; (iii) some studies did not provide details of gene mutations. We also provided predictions of disease-causing potential of variants from several in silico tools, such as MutationTaster (48) and FATHMM (49). A recent study identified 299 common driver genes in cancer (50), and we marked these genes in the gene information page. CaPSSA (51) and KM Plotter (52) were used to assess the effect of genes on survival. CaPSSA enables evaluation of cancer biomarker genes for patient stratification and survival analysis based on mutation and expression data. KM Plotter is a web application, developed for meta-analysis-based biomarker assessment. Longitudinal tumor sampling before and after therapy was also often carried out to assess treatment-induced molecular changes. In such cases, the phylogenetic trees illustrated the temporal acquisitions of driver events that represented valuable information relevant to treatment outcome.

Intertumor heterogeneity data annotation

Substantial intertumor heterogeneity between primary tumors and metastases has been reported in many studies uti-

lizing high-throughput sequencing data (53–57). In these cases, metastatic tumors tend to evolve under new microenvironmental pressures, and present different genomic alterations than those found in the primary tumor. For studies that investigated genetic differences between primary tumors and multiple metastatic sites in the same patient, data curation was similar to that of intratumor heterogeneity data processing. We collected somatic genomic alterations and classified them into trunk (present in all metastases and primary lesions), branch (shared by some samples but not all) and private (only detected in one sample) alterations. The primary and metastatic tumor-specific mutations or copy number changes were used to construct phylogenetic trees. The trunks were usually rooted by germline DNA sequences that did not have somatic mutations. Trunk and branch lengths were proportional to the numbers of non-synonymous mutations. The organ-specific branches within each dendrogram represent the degree of clonal diversity between primary and metastasis tumors. This type of phylogenetic tree may help facilitate a deeper understanding of the phylogenetic relationships between primary and metastatic neoplasms, as well as identify founder mutations and processes that are relevant to metastatic progression.

Database architecture and data visualization

CancerTracer runs on a linux-based Apache web server using MySQL as the back-end database. The web interface was developed using HTML5, PHP, CSS and JavaScript. The website has been tested thoroughly to ensure functionality across various operating systems and web browsers such as Google Chrome, Safari, Firefox, Opera, and Microsoft Edge. The data visualization was performed using the R programming language, Metascape (58), and the ggplot2 R package (59). Metascape is an online tool that can be used to perform gene annotation and enrichment analysis (58). The analysis results returned by Metascape were reconstructed and integrated into CancerTracer. In most enrichment analyses, *P*-values were used for statistical significance evaluation. This is to avoid missing interesting enriched terms and valuable information, since a false discov-
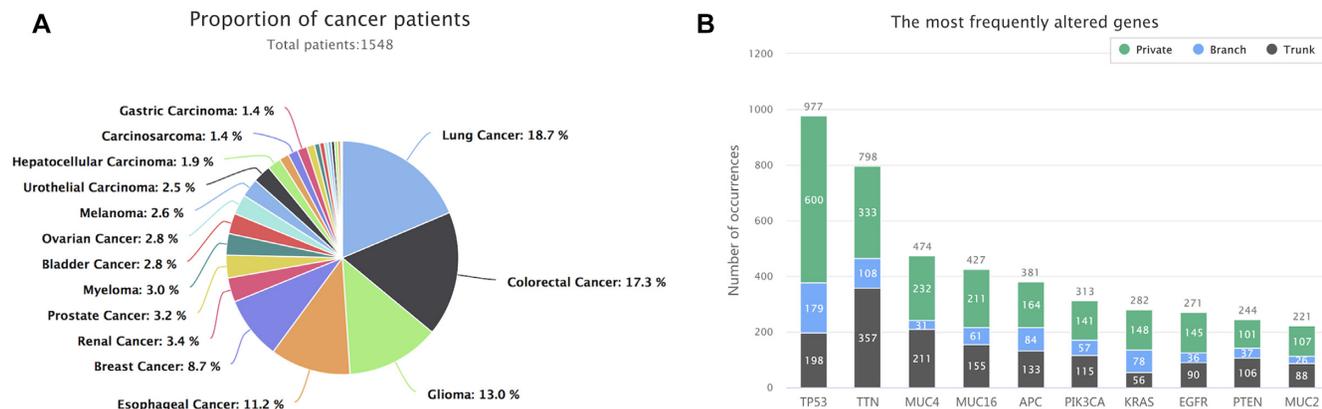
A
Proportion of cancer patients
Total patients:1548

B
The most frequently altered genes

Figure 2. Statistics on the data contents of CancerTracer. (**A**) Distribution of patients across tumor types. (**B**) The most frequently altered genes and their regional distributions.

ery rate adjustment procedure may produce conservative *P*-values and declare very few terms as significant. For enrichment network visualization, Metascape employed a heuristic algorithm to select the most informative terms from the obtained GO clusters. The top 20 clusters were sampled and up to 10 GO terms with lowest *P*-values within each cluster were selected. Pairwise similarities between any two enriched terms were calculated based on a Kappa-test score. All terms pairs with Kappa similarity >0.3 were connected and visualized by Cytoscape (60).

DATABASE CONTENT AND USAGE

Data summary

The current version of CancerTracer contains genomic alteration data profiled from over 6000 tumor samples representing 1548 individual patients. These datasets were curated from 145 publicly available studies across 45 cancer types. Additional information of the 145 articles can be found in Supplementary Table S1. Lung cancer accounts for the largest proportion of patients (18.7%) in our dataset, followed by colorectal cancer (17.3%), glioma (13.0%), esophageal cancer (11.2%) and breast cancer (8.7%) (Figure 2A). The most commonly mutated gene in all cancers is TP53, which has nearly 1000 alteration entries in the database. The top 10 most frequently altered genes are shown in Figure 2B. For each gene, we counted the number of mutations identified as trunk, branch, or private. Most genes appeared as both trunk and private mutations in different samples. The data thus indicated that tumors represent a complex dynamic ecological system, and the roles of trunk or private mutations are influenced by many microenvironmental factors surrounding each tumor.

Web design and interface

The CancerTracer website features a user-friendly interface for exploring and visualizing tumor heterogeneity data within patients. The main functional pages include 'Browse', 'Intratumor', 'Intertumor', 'Download' and 'Tu-

torial'. The 'Browse' page is the main entry point to explore the datasets, while 'Intratumor' and 'Intertumor' are two query options to retrieve specific types of data from the database. The 'Download' page provides links to download the complete datasets of CancerTracer, and instructions on how to use the online interface can be found on the 'Tutorial' page.

The 'Browse' page provides an overview of all data included in the database, and data can be browsed by cancer types or gene/event. When clicking a specific cancer type, all related entries are shown in a table (Figure 3A). The most frequently mutated genes in this tumor type are presented in a histogram at the bottom of the page. Clicking on the patient ID will open a new page that displays sample information and data analysis results. In the new page, the 'Patient Details' field shows basic clinical information of the patient (Figure 3B). The 'Supported Literature' field describes the source from which the data was extracted, such as PubMed ID, table/figure number, etc. Users can thus easily trace the data back to the original literature. If available, the technical details of raw data processing were also provided, such as sequencing platforms and bioinformatics tools applied. We also curated the information of tumor purity, patient prognosis, and treatment strategies from publications. Some studies provide information on the exact points of sampling. Since sampling strategies may influence the sensitivity and accuracy of clonality analysis, we adapted and illustrated this information in a schematic diagram (Figure 3C). Each patient-specific phylogenetic tree reveals the spatial composition or the evolutionary trajectory of all identified subclones (Figure 3D). Driver genes with possible functional mutations are usually mapped along the phylogenetic trees. To facilitate the assessment of tumor heterogeneity in all patients in a study, a heatmap is generated to show the regional distribution of all cancer gene mutations (Figure 3E). Altered genes are listed at the bottom of the heatmap, and patient IDs are shown on the right. The status of each gene in each patient is color-coded. 'Gene Lists Analysis' field provides the results of a functional enrichment analysis based on trunk gene lists of a specific study. Since trunk mutations are shared by every region of an individual tumor, the anal-
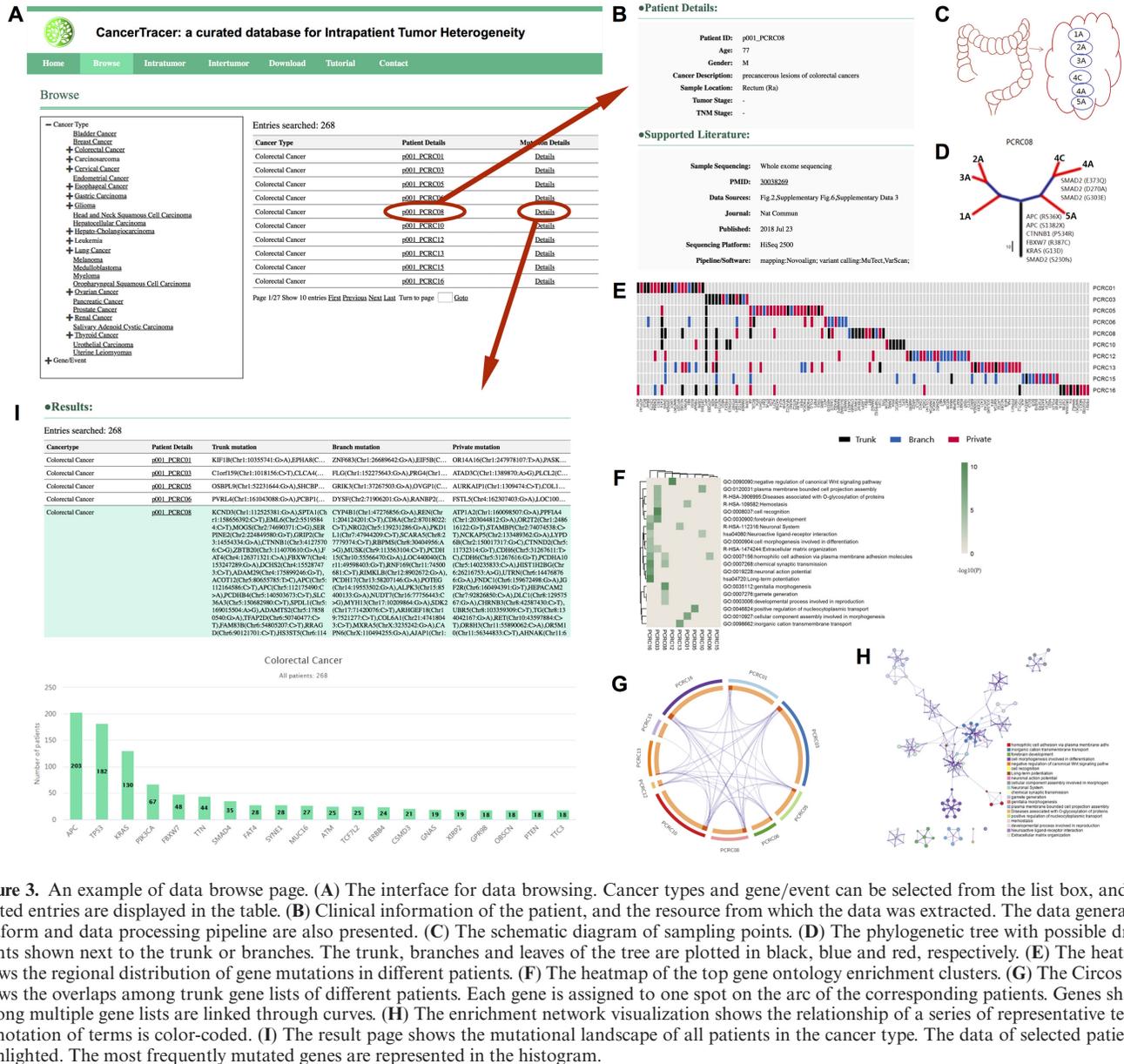
Figure 3. An example of data browse page. (**A**) The interface for data browsing. Cancer types and gene/event can be selected from the list box, and the related entries are displayed in the table. (**B**) Clinical information of the patient, and the resource from which the data was extracted. The data generation platform and data processing pipeline are also presented. (**C**) The schematic diagram of sampling points. (**D**) The phylogenetic tree with possible driver events shown next to the trunk or branches. The trunk, branches and leaves of the tree are plotted in black, blue and red, respectively. (**E**) The heatmap shows the regional distribution of gene mutations in different patients. (**F**) The heatmap of the top gene ontology enrichment clusters. (**G**) The Circos plot shows the overlaps among trunk gene lists of different patients. Each gene is assigned to one spot on the arc of the corresponding patients. Genes shared among multiple gene lists are linked through curves. (**H**) The enrichment network visualization shows the relationship of a series of representative terms. Annotation of terms is color-coded. (**I**) The result page shows the mutational landscape of all patients in the cancer type. The data of selected patient is highlighted. The most frequently mutated genes are represented in the histogram.

ysis of truncal mutations can provide common results. This analysis was performed by an online tool called MetaScape (58), and its results were used to reconstruct parts of each patient's diagram using the R language. The heatmap of the top Gene Ontology (GO) enrichment clusters is included to help facilitate the identification of key processes or pathways that are common to all patients (Figure 3F). The Circos plot (61) intuitively shows the overlaps among multiple trunk gene lists, and the functional overlaps of genes that share the same ontology term (Figure 3G). Furthermore, a series of representative terms from the full cluster were selected and converted into a network layout (Figure 3H). When clicking on the graph, a new page containing an interactive network will appear. User can click on nodes to explore GO terms of interest, and get related gene list.

The complete list of mutated genes identified in the patient, and their distribution levels, can be viewed by clicking

on the 'Mutation Details' link in the table of retrieved entries. The result page displays the mutational landscape of all patients with the specific cancer type, and the record of the selected patient will be highlighted (Figure 3I).

Evolutionary events query

The data in CancerTracer can be split into two conceptual categories: intratumor and intertumor heterogeneity. We developed query interfaces for these two data types, respectively. Intratumor heterogeneity data can be queried by cancer type or gene/event (Figure 4A). In the interface of query by cancer type, we provide another option to select data heterogeneity type, including spatial and temporal heterogeneity, which represent different sampling strategies. For researchers only focused on known cancer genes, the 'Cancer Gene Census' (cancer-related genes annotated
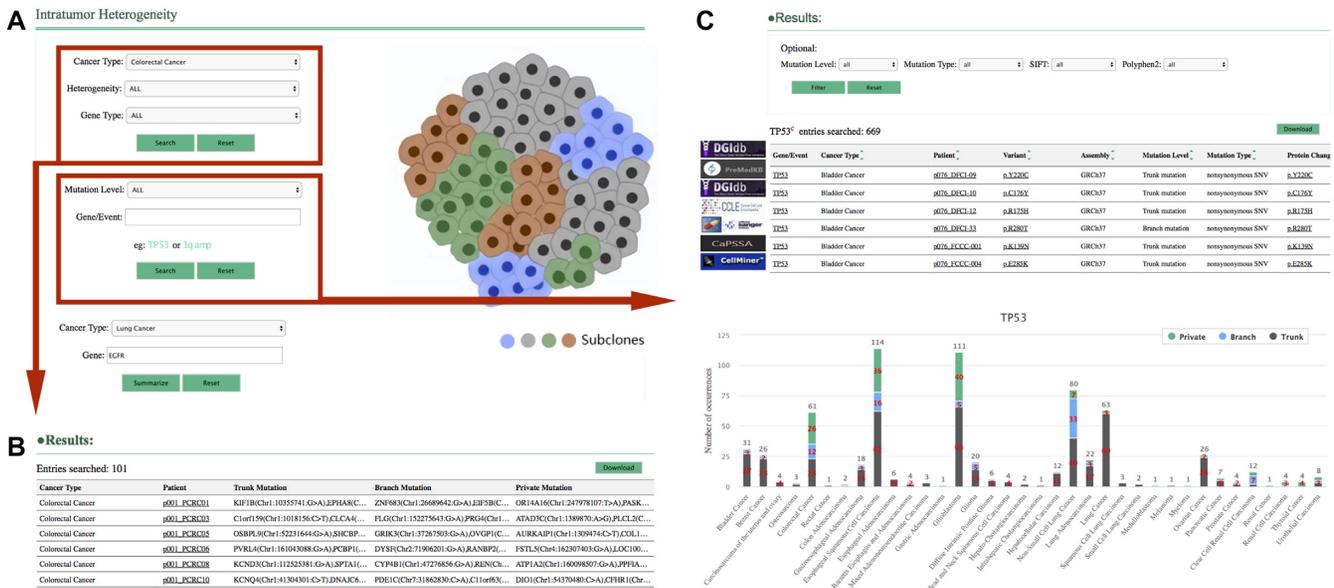
Figure 4. Examples of intratumor heterogeneity data query. (**A**) The interface and related parameters for intratumor heterogeneity data query. (**B**) The results of query by cancer type. The table presents patient-level result. (**C**) The results of query by Gene/Event. The table provides detailed information about the queried gene/event and links to other sources. Several filter options are provided above the table. The histogram represents the regional distribution of all mutations in the queried gene/event across different cancer types.

in the COSMIC database) (33) option in the 'Gene Type' selection box can be selected, instead of 'All'. The result page presents data at the patient level, and provides lists of affected genes and links to individual patient data (Figure 4B). The retrieved data can be downloaded as an Excel file by clicking on the 'Download' button.

In the query by 'Gene/Event' interface, users can select different mutation levels, including trunk, branch and private, which may indicate the time of occurrence of the mutation during tumor evolution (Figure 4A). The 'Gene/Event' input box suggests plausible gene or event names, and supports auto-completion. In the result page, a table is displayed that lists related information on the queried gene, including cancer type, links to patient-level data, gene variations, and version of reference-genome assembly, etc. (Figure 4C). The variations associated with amino acid changes were normalized to the standard format and linked to the COSMIC database for details. We also provide systematic annotation on the curated somatic mutations by ANNOVAR. The annotations from SIFT, PolyPhen2, etc., as well as protein change, were displayed in the table. The results can be sorted by columns and be downloaded for offline analysis. To allow users to access the potential medical value of genes of interest, we further provide links for each mutation or gene to PreMedKB (38), DGIdb (40) and KM Plotter (52), if available. The regional distribution of all gene mutations in different cancer types is illustrated in the histogram below the table.

Intertumor heterogeneity data can also be queried by cancer type or gene/event. The difference is that the options of heterogeneity type are classified as primary-metastasis or primary-multimetastasis. In this case, the data can be used to investigate phylogenetic relationships between primary and metastatic tumors. In the 'Intertumor Hetero-

geneity' and 'Intratumor Heterogeneity' pages, we provide interfaces to display summary of gene mutations. This function allows users to investigate metastasis specific mutations and mutations that appear both in primary and metastasis samples.

Case study

To exhibit the utility and potential application of Cancer-Tracer, we used EGFR as an example to query its mutation states in lung cancer (Supplementary Figure S1). In 'Intratumor Heterogeneity' page, we selected lung cancer and EGFR to perform the query, and found that EGFR: p.L858R mutations, which are major selection markers for EGFR tyrosine kinase inhibitors (TKIs) therapy, were usually presented as trunk or branch mutations. However, EGFR: p.T790M mutations, which cause resistance to first and second generations EGFR TKIs, were found as trunk or branch mutations in two patients (p107_P012 and p109_p4–6992). The patient p107_P012 had both p.L858R (as branch) and p.T790M (as trunk) mutations, which may suggest higher EGFR activity and weak response to first and second generations TKIs. These results may also suggest that p.T790M mutation occurred early in disease progression of this patient. Thus, CancerTracer allows users to investigate the distribution of specific mutations of interest, and to deduce the time of occurrence of the mutation during tumor evolution.

Data access

To enable users to perform further metadata analysis, the entire contents of CancerTracer can be downloaded on the 'Download' page without login or registration. The full

dataset was split into sample level data and publication information, and can be downloaded as tab-delimited files. The dataset was also organized by intra- or intertumor heterogeneity and gene type to provide users with flexible data choices. The emerging studies utilizing spatial and longitudinal sampling strategies provide in-depth understanding of tumor evolution during progression and treatment. The continuously growing data is a valuable resource and we will update our database several times per year.

DISCUSSION

To date, an increasing number of studies have highlighted the fact that single biopsies may not be able to provide adequate reflections of the clonal compositions of whole tumors (9,29,62). Molecular and genetic heterogeneity within a single tumor, as well as between different sites of neoplasia in a single patient, confound our ability to design and select effective therapies to curtail treatment resistance. We developed CancerTracer to help researchers gain a deeper insight into the extent of tumor heterogeneity and how it evolves over the course of the disease. This manually curated platform brings together clinical and genomic alteration data concerning tumor heterogeneity, to allow users to quickly explore the spatial compositions and the evolutionary trajectories of tumor subclones.

The study of tumor evolution over space and time can facilitate the identification of truncal and branched driver mutations. An example of such a study is the TRACERx (TRAcking non-small cell lung Cancer Evolution through therapy (Rx)), which attempts to map cancer subclones over time and to understand the impact of intratumor heterogeneity on therapeutic outcomes (63). Clonally dominant or truncal driver mutations could be evaluated in order to find the most effective drug targets for certain tumors. However, the roles of driver and passenger mutations are constrained by spatial and temporal contexts - somatic events can act as drivers at one stage of tumorigenesis, and as passengers at another stage, and vice-versa, in some cases. Moreover, a driver mutation might give way to a passenger mutation when the tumor is under selection pressure from a course of treatment (14,64–66). The matter is complicated even further by the fact that the extent and patterns of heterogeneity vary within histological tumor subtypes and between tumors of different tissue types. Therefore, further exploration of large-scale tumor heterogeneity data is necessary in order to reveal the molecular mechanisms driving intrapatient heterogeneity. We will therefore commit to continuously updating and expanding our data content to meet the growing requirements of the research community.

It is important to note that besides genetic alterations, epigenetic changes also play significant roles in shaping the clonal architectures and heterogeneities of many tumors. Epigenetic clonality can reflect the potential of the tumor to respond to changing environments during progression and therapy. Furthermore, an increasing number of studies have demonstrated that epigenetics provides a complementary paradigm to the analysis of genetic mutations in tumorigenesis (67–69). The degree and patterns of heterogeneity may be influenced by the interplay of the genome and epigenome in each tumor cell, and thus the integrative analysis of both data types will facilitate a clearer understanding of the mechanisms driving tumor heterogeneity. However, despite the importance of intratumor epigenetic heterogeneity, it has rarely been examined and has only been described in several tumor types, including brain tumors (70), prostate cancer (71) and hepatocellular cancer (72). Although the current version of CancerTracer focuses primarily on genetic heterogeneity, data on epigenetic heterogeneity will be added to the database once enough data is collected.

Recent progress in single-cell genome sequencing has enabled characterization of both somatic mutations and copy number alterations in a single tumor cell. It is a powerful tool to explore genetic and functional heterogeneity, and to detect rare subpopulations. Compared to single-cell sequencing, multi-region sequencing may be insufficient for precise detection of subclonal CNAs and low-frequency mutations. One recent single-cell sequencing study detected CNAs that were not detected in the multi-region sequencing (73). It indicated that a detailed genetic variation of the tumor may be better uncovered by single-cell sequencing. However, for single-cell sequencing, a large number of cells are usually required to be sequenced in order to obtain meaningful results, which is costly and has its limitations in clinical applications. Furthermore, single-cell RNA sequencing (scRNA-seq) can also be used to assess tumor heterogeneity (74). It allows researchers to investigate the diversity of transcriptional profiles present in an individual tumor, and to explore multiple cell states with distinct transcriptional profiles. Many scRNA-seq studies have revealed new insights into tumor-related mechanism in detail. For example, tumor microenvironment was found to be consisted of fibroblasts, T cells, macrophages, etc., besides malignant cells (75). Puram *et al.* revealed tumor subpopulation with partial epithelial-to-mesenchymal transition in head and neck carcinoma, which may be used to predict tumor invasion and metastasis (76). In the future, we plan to integrate scRNA-seq data into CancerTracer. Given that scRNA-seq provides another type of data, we will develop new pages and visualization tools to present this type of tumor heterogeneity data.

Several limitations due to tumor purity need to be noticed. Solid tumor samples typically contain normal cell contamination. A recent study reported that the proportion of private mutations in a given biopsy was positively correlated with its purity (77). Private mutations are less likely to be identified in biopsies of relatively low purity. Thus, large variability in tumor purity between biopsies from the same patient may cause the overestimation of intratumor heterogeneity. Furthermore, in most publications, the authors did not provide information on tumor purity. The included studies used different methods to assess tumor purity. Some studies evaluated tumor purity based on histopathological slides, while others applied computational methods for purity speculation. These limitations pose considerable challenges for downstream data integration, and increase the risk of inaccurate inferences.

In summary, the extent of intrapatient tumor heterogeneity is a complex issue that researchers are only beginning to understand. To the best of our knowledge, CancerTracer is the first database dedicated to exploring, integrating and

mining tumor heterogeneity data at an individual patient level. To overcome the challenges of tumor heterogeneity, collaborations between computational biologists, cancer biologists, technology developers, and clinicians are required. We hope that our database will substantially contribute to this research field.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Dr Yang Cao for his valuable suggestions and help in website construction.

FUNDING

REFERENCES

1. Alexandrov,L.B., Nik-Zainal,S., Wedge,D.C., Aparicio,S.A.J.R., Behjati,S., Biankin,A.V., Bignell,G.R., Bolli,N., Borg,A., Børresen-Dale,A.-L. *et al.* (2013) Signatures of mutational processes in human cancer. *Nature*, **500**, 415–421.
2. Lawrence,M.S., Stojanov,P., Mermel,C.H., Robinson,J.T., Garraway,L.A., Golub,T.R., Meyerson,M., Gabriel,S.B., Lander,E.S. and Getz,G. (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*, **505**, 495–501.
3. Anderson,K., Lutz,C., van Delft,F.W., Bateman,C.M., Guo,Y., Colman,S.M., Kempski,H., Moorman,A.V., Titley,I., Swansbury,J. *et al.* (2011) Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature*, **469**, 356–361.
4. Vogelstein,B., Papadopoulos,N., Velculescu,V.E., Zhou,S., Diaz,L.A. and Kinzler,K.W. (2013) Cancer genome landscapes. *Science*, **339**, 1546–1558.
5. Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
6. Swanton,C. (2014) Cancer evolution: the final frontier of precision medicine? *Ann. Oncol.*, **25**, 549–551.
7. Yates,L.R. and Campbell,P.J. (2012) Evolution of the cancer genome. *Nat. Rev. Genet.*, **13**, 795–806.
8. Wood,L.D., Parsons,D.W., Jones,S., Lin,J., Sjöblom,T., Leary,R.J., Shen,D., Boca,S.M., Barber,T., Ptak,J. *et al.* (2007) The genomic landscapes of human breast and colorectal cancers. *Science*, **318**, 1108–1113.
9. Gerlinger,M., Rowan,A.J., Horswell,S., Math,M., Larkin,J., Endesfelder,D., Gronroos,E., Martinez,P., Matthews,N., Stewart,A. *et al.* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.*, **366**, 883–892.
10. Jamal-Hanjani,M., Wilson,G.A., McGranahan,N., Birkbak,N.J., Watkins,T.B.K., Veeriah,S., Shafi,S., Johnson,D.H., Mitter,R., Rosenthal,R. *et al.* (2017) Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.*, **376**, 2109–2121.
11. Gerlinger,M., Horswell,S., Larkin,J., Rowan,A.J., Salm,M.P., Varela,I., Fisher,R., McGranahan,N., Matthews,N., Santos,C.R. *et al.* (2014) Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.*, **46**, 225–233.
12. Yates,L.R., Gerstung,M., Knappskog,S., Desmedt,C., Gundem,G., Van Loo,P., Aas,T., Alexandrov,L.B., Larsimont,D., Davies,H. *et al.* (2015) Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat. Med.*, **21**, 751–759.
13. Hao,J.-J., Lin,D.-C., Dinh,H.Q., Mayakonda,A., Jiang,Y.-Y., Chang,C., Jiang,Y., Lu,C.-C., Shi,Z.-Z., Xu,X. *et al.* (2016) Spatial intratumoral heterogeneity and temporal clonal evolution in esophageal squamous cell carcinoma. *Nat. Genet.*, **48**, 1500–1507.
14. Murugaesu,N., Wilson,G.A., Birkbak,N.J., Watkins,T., McGranahan,N., Kumar,S., Abbassi-Ghadi,N., Salm,M., Mitter,R., Horswell,S. *et al.* (2015) Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discov.*, **5**, 821–831.
15. Chen,X.-X., Zhong,Q., Liu,Y., Yan,S.-M., Chen,Z.-H., Jin,S.-Z., Xia,T.-L., Li,R.-Y., Zhou,A.-J., Su,Z. *et al.* (2017) Genomic comparison of esophageal squamous cell carcinoma and its precursor lesions by multi-region whole-exome sequencing. *Nat. Commun.*, **8**, 524.
16. Juric,D., Castel,P., Griffith,M., Griffith,O.L., Won,H.H., Ellis,H., Ebbesen,S.H., Ainscough,B.J., Ramu,A., Iyer,G. *et al.* (2015) Convergent loss of PTEN leads to clinical resistance to a PI(3)Kα inhibitor. *Nature*, **518**, 240–244.
17. Kwak,E.L., Ahronian,L.G., Siravegna,G., Mussolin,B., Borger,D.R., Godfrey,J.T., Jessop,N.A., Clark,J.W., Blaszkowsky,L.S., Ryan,D.P. *et al.* (2015) Molecular heterogeneity and receptor coamplification drive resistance to targeted therapy in met-amplified esophagogastric cancer. *Cancer Discov.*, **5**, 1271–1281.
18. Russo,M., Siravegna,G., Blaszkowsky,L.S., Corti,G., Crisafulli,G., Ahronian,L.G., Mussolin,B., Kwak,E.L., Buscarino,M., Lazzari,L. *et al.* (2016) Tumor heterogeneity and lesion-specific response to targeted therapy in colorectal cancer. *Cancer Discov.*, **6**, 147–153.
19. Dagogo-Jack,I. and Shaw,A.T. (2018) Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.*, **15**, 81–94.
20. Jamal-Hanjani,M., Quezada,S.A., Larkin,J. and Swanton,C. (2015) Translational implications of tumor heterogeneity. *Clin. Cancer Res.*, **21**, 1258–1266.
21. Yap,T.A., Gerlinger,M., Futreal,P.A., Pusztai,L. and Swanton,C. (2012) Intratumor heterogeneity: seeing the wood for the trees. *Sci. Transl. Med.*, **4**, 127ps10.
22. Gerlinger,M., McGranahan,N., Dewhurst,S.M., Burrell,R.A., Tomlinson,I. and Swanton,C. (2014) Cancer: evolution within a lifetime. *Annu. Rev. Genet.*, **48**, 215–236.
23. Zhang,J., Fujimoto,J., Zhang,J., Wedge,D.C., Song,X., Zhang,J., Seth,S., Chow,C.-W., Cao,Y., Gumbs,C. *et al.* (2014) Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science*, **346**, 256–259.
24. Rasche,L., Chavan,S.S., Stephens,O.W., Patel,P.H., Tytarenko,R., Ashby,C., Bauer,M., Stein,C., Deshpande,S., Wardell,C. *et al.* (2017) Spatial genomic heterogeneity in multiple myeloma revealed by multi-region sequencing. *Nat. Commun.*, **8**, 268.
25. Yan,T., Cui,H., Zhou,Y., Yang,B., Kong,P., Zhang,Y., Liu,Y., Wang,B., Cheng,Y., Li,J. *et al.* (2019) Multi-region sequencing unveils novel actionable targets and spatial heterogeneity in esophageal squamous cell carcinoma. *Nat. Commun.*, **10**, 1670.
26. Zhai,W., Lim,T.K.-H., Zhang,T., Phang,S.-T., Tiang,Z., Guan,P., Ng,M.-H., Lim,J.Q., Yao,F., Li,Z. *et al.* (2017) The spatial organization of intra-tumour heterogeneity and evolutionary trajectories of metastases in hepatocellular carcinoma. *Nat. Commun.*, **8**, 4565.
27. Francis,J.M., Zhang,C.-Z., Maire,C.L., Jung,J., Manzo,V.E., Adalsteinsson,V.A., Homer,H., Haidar,S., Blumenstiel,B., Pedamallu,C.S. *et al.* (2014) EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discov.*, **4**, 956–971.
28. Xu,X., Hou,Y., Yin,X., Bao,L., Tang,A., Song,L., Li,F., Tsang,S., Wu,K., Wu,H. *et al.* (2012) Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell*, **148**, 886–895.
29. Navin,N., Kendall,J., Troge,J., Andrews,P., Rodgers,L., McIndoo,J., Cook,K., Stepansky,A., Levy,D., Esposito,D. *et al.* (2011) Tumour evolution inferred by single-cell sequencing. *Nature*, **472**, 90–94.
30. Kim,C., Gao,R., Sei,E., Brandt,R., Hartman,J., Hatschek,T., Crosetto,N., Foukakis,T. and Navin,N.E. (2018) Chemoresistance evolution in triple-negative breast cancer delineated by single-cell sequencing. *Cell*, **173**, 879–893.
31. Cancer Genome Atlas Research Network, Weinstein,J.N., Collisson,E.A., Mills,G.B., Shaw,K.R.M., Ozenberger,B.A., Ellrott,K., Shmulevich,I., Sander,C. and Stuart,J.M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.

32. International Cancer Genome Consortium, Hudson,T.J., Anderson,W., Artez,A., Barker,A.D., Bell,C., Bernabé,R.R., Bhan,M.K., Calvo,F., Eerola,I. *et al.* (2010) International network of cancer genome projects. *Nature*, **464**, 993–998.

33. Tate,J.G., Bamford,S., Jubb,H.C., Sondka,Z., Beare,D.M., Bindal,N., Boutselakis,H., Cole,C.G., Creatore,C., Dawson,E. *et al.* (2019) COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.*, **47**, D941–D947.

34. Landrum,M.J., Lee,J.M., Benson,M., Brown,G.R., Chao,C., Chitipiralla,S., Gu,B., Hart,J., Hoffman,D., Jang,W. *et al.* (2018) ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.*, **46**, D1062–D1067.

35. Li,X., Shi,L., Wang,Y., Zhong,J., Zhao,X., Teng,H., Shi,X., Yang,H., Ruan,S., Li,M. *et al.* (2019) OncoBase: a platform for decoding regulatory somatic mutations in human cancers. *Nucleic Acids Res.*, **47**, D1044–D1055.

36. Dingerdissen,H.M., Torcivia-Rodriguez,J., Hu,Y., Chang,T.C., Mazumder,R. and Kahsay,R. (2018) BioMuta and BioXpress: mutation and expression knowledgebases for cancer biomarker discovery. *Nucleic Acids Res.*, **46**, D1128–D1136.

37. Gonzalez-Perez,A., Perez-Llamas,C., Deu-Pons,J., Tamborero,D., Schroeder,M.P., Jene-Sanz,A., Santos,A. and Lopez-Bigas,N. (2013) IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods*, **10**, 1081–1082.

38. Yu,Y., Wang,Y., Xia,Z., Zhang,X., Jin,K., Yang,J., Ren,L., Zhou,Z., Yu,D., Qing,T. *et al.* (2019) PreMedKB: an integrated precision medicine knowledgebase for interpreting relationships between diseases, genes, variants and drugs. *Nucleic Acids Res.*, **47**, D1090–D1101.

39. Zhang,H., Luo,S., Zhang,X., Liao,J., Quan,F., Zhao,E., Zhou,C., Yu,F., Yin,W., Zhang,Y. *et al.* (2018) SEECancer: a resource for somatic events in evolution of cancer genome. *Nucleic Acids Res.*, **46**, D1018–D1026.

40. Cotto,K.C., Wagner,A.H., Feng,Y.Y., Kiwala,S., Coffman,A.C., Spies,G., Wollam,A., Spies,N.C., Griffith,O.L. and Griffith,M. (2018) DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.*, **46**, D1068–D1073.

41. Behan,F.M., Iorio,F., Picco,G., Gonçalves,E., Beaver,C.M., Migliardi,G., Santos,R., Rao,Y., Sassi,F., Pinnelli,M. *et al.* (2019) Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature*, **568**, 511–516.

42. Ghandi,M., Huang,F.W., Jané-Valbuena,J., Kryukov,G.V., Lo,C.C., McDonald,E.R. 3rd, Barretina,J., Gelfand,E.T., Bielski,C.M., Li,H. *et al.* (2019) Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*, **569**, 503–508.

43. Yang,W., Soares,J., Greninger,P., Edelman,E.J., Lightfoot,H., Forbes,S., Bindal,N., Beare,D., Smith,J.A., Thompson,I.R. *et al.* (2013) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.*, **41**, D955–D961.

44. Reinhold,W.C., Sunshine,M., Liu,H., Varma,S., Kohn,K.W., Morris,J., Doroshow,J. and Pommier,Y. (2012) CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res.*, **72**, 3499–3511.

45. Su,G., Burant,C.F., Beecher,C.W., Athey,B.D. and Meng,F. (2011) Integrated metabolome and transcriptome analysis of the NCI60 dataset. *BMC Bioinformatics*, **12**, S36.

46. Wang,K., Li,M. and Hakonarson,H. (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.*, **38**, e164.

47. Adzhubei,I.A., Schmidt,S., Peshkin,L., Ramensky,V.E., Gerasimova,A., Bork,P., Kondrashov,A.S. and Sunyaev,S.R. (2010) A method and server for predicting damaging missense mutations. *Nat Methods*, **7**, 248–249.

48. Schwarz,J.M., Rödelsperger,C., Schuelke,M. and Seelow,D. (2010) MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*, **7**, 575–576.

49. Rogers,M.F., Shihab,H.A., Mort,M., Cooper,D.N., Gaunt,T.R. and Campbell,C. (2018) FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics*, **34**, 511–513.

50. Bailey,M.H., Tokheim,C., Porta-Pardo,E., Sengupta,S., Bertrand,D., Weerasinghe,A., Colaprico,A., Wendl,M.C., Kim,J., Reardon,B. *et al.*

51. Jang,Y., Seo,J., Kim,S. and Lee,S. (2019) CaPSSA: Visual evaluation of cancer biomarker genes for patient stratification and survival analysis using mutation and expression data. *Bioinformatics*, btz516.

52. Nagy,Á., Lánczky,A., Menyhárt,O. and Győrffy,B. (2018) Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci. Rep.*, **8**, 9227.

53. Wang,D., Niu,X., Wang,Z., Song,C.-L., Huang,Z., Chen,K.-N., Duan,J., Bai,H., Xu,J., Zhao,J. *et al.* (2019) Multiregion sequencing reveals the genetic heterogeneity and evolutionary history of osteosarcoma and matched pulmonary metastases. *Cancer Res.*, **79**, 7–20.

54. Wei,Q., Ye,Z., Zhong,X., Li,L., Wang,C., Myers,R.E., Palazzo,J.P., Fortuna,D., Yan,A., Waldman,S.A. *et al.* (2017) Multiregion whole-exome sequencing of matched primary and metastatic tumors revealed genomic heterogeneity and suggested polyclonal seeding in colorectal cancer metastasis. *Ann. Oncol.*, **28**, 2135–2141.

55. Gibson,W.J., Hoivik,E.A., Halle,M.K., Taylor-Weiner,A., Cherniack,A.D., Berg,A., Holst,F., Zack,T.I., Werner,H.M.J., Staby,K.M. *et al.* (2016) The genomic landscape and evolution of endometrial carcinoma progression and abdominopelvic metastasis. *Nat. Genet.*, **48**, 848–855.

56. Yachida,S., Jones,S., Bozic,I., Antal,T., Leary,R., Fu,B., Kamiyama,M., Hruban,R.H., Eshleman,J.R., Nowak,M.A. *et al.* (2010) Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature*, **467**, 1114–1117.

57. Sanborn,J.Z., Chung,J., Purdom,E., Wang,N.J., Kakavand,H., Wilmott,J.S., Butler,T., Thompson,J.F., Mann,G.J., Haydu,L.E. *et al.* (2015) Phylogenetic analyses of melanoma reveal complex patterns of metastatic dissemination. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 10995–11000.

58. Zhou,Y., Zhou,B., Pache,L., Chang,M., Khodabakhshi,A.H., Tanaseichuk,O., Benner,C. and Chanda,S.K. (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.*, **10**, 1523.

59. Wickham,H. (2016) *ggplot2 - Elegant Graphics for Data Analysis*. Springer-Verlag, NY, p. 260.

60. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

61. Krzywinski,M., Schein,J., Birol,I., Connors,J., Gascoyne,R., Horsman,D., Jones,S.J. and Marra,M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639–1645.

62. Navin,N., Krasnitz,A., Rodgers,L., Cook,K., Meth,J., Kendall,J., Riggs,M., Eberling,Y., Troge,J., Grubor,V. *et al.* (2010) Inferring tumor progression from genomic heterogeneity. *Genome Res.*, **20**, 68–80.

63. Jamal-Hanjani,M., Hackshaw,A., Ngai,Y., Shaw,J., Dive,C., Quezada,S., Middleton,G., de Bruin,E., Le Quesne,J., Shafi,S. *et al.* (2014) Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol.*, **12**, e1001906.

64. Findlay,J.M., Castro-Giner,F., Makino,S., Rayner,E., Kartsonaki,C., Cross,W., Kovac,M., Ulahannan,D., Palles,C., Gillies,R.S. *et al.* (2016) Differential clonal evolution in oesophageal cancers in response to neo-adjuvant chemotherapy. *Nat. Commun.*, **7**, 11111.

65. Harbst,K., Lauss,M., Cirenajwis,H., Isaksson,K., Rosengren,F., Törngren,T., Kvist,A., Johansson,M.C., Vallon-Christersson,J., Baldetorp,B. *et al.* (2016) Multiregion whole-exome sequencing uncovers the genetic evolution and mutational heterogeneity of early-stage metastatic melanoma. *Cancer Res.*, **76**, 4765–4774.

66. Marusyk,A., Almendro,V. and Polyak,K. (2012) Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer*, **12**, 323–334.

67. Assenov,Y., Brocks,D. and Gerhäuser,C. (2018) Intratumor heterogeneity in epigenetic patterns. *Semin. Cancer Biol.*, **51**, 12–21.

68. Landau,D.A., Clement,K., Ziller,M.J., Boyle,P., Fan,J., Gu,H., Stevenson,K., Sougnez,C., Wang,L., Li,S. *et al.* (2014) Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell*, **26**, 813–825.

69. Torres,C.M., Biran,A., Burney,M.J., Patel,H., Henser-Brownhill,T., Cohen,A.-H.S., Li,Y., Ben-Hamo,R., Nye,E., Spencer-Dene,B. *et al.*

(2016) The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity. *Science*, **353**, aaf1644.

70. Mazor,T., Pankov,A., Johnson,B.E., Hong,C., Hamilton,E.G., Bell,R.J.A., Smirnov,I.V., Reis,G.F., Phillips,J.J., Barnes,M.J. *et al.* (2015) DNA methylation and somatic mutations converge on the cell cycle and define similar evolutionary histories in brain tumors. *Cancer Cell*, **28**, 307–317.

71. Brocks,D., Assenov,Y., Minner,S., Bogatyrova,O., Simon,R., Koop,C., Oakes,C., Zucknick,M., Lipka,D.B., Weischenfeldt,J. *et al.* (2014) Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. *Cell Rep.*, **8**, 798–806.

72. Hlady,R.A., Zhou,D., Puszyk,W., Roberts,L.R., Liu,C. and Robertson,K.D. (2017) Initiation of aberrant DNA methylation patterns and heterogeneity in precancerous lesions of human hepatocellular cancer. *Epigenetics*, **12**, 215–225.

73. Liu,M., Liu,Y., Di,J., Su,Z., Yang,H., Jiang,B., Wang,Z., Zhuang,M., Bai,F. and Su,X. (2017) Multi-region and single-cell sequencing reveal variable genomic heterogeneity in rectal cancer. *BMC Cancer.*, **17**, 787.

74. Parra,R.G., Papadopoulos,N., Ahumada-Arranz,L., Kholtei,J.E., Mottelson,N., Horokhovsky,Y., Treutlein,B. and Soeding,J. (2019) Reconstructing complex lineage trees from scRNA-seq data using MERLoT. *Nucleic Acids Res.*, **47**, 8961–8974.

75. Lavin,Y., Kobayashi,S., Leader,A., Amir,E.D., Elefant,N., Bigenwald,C., Remark,R., Sweeney,R., Becker,C.D., Levine,J.H. *et al.* (2017) Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell*, **169**, 750–765.

76. Puram,S.V., Tirosh,I., Parikh,A.S., Patel,A.P., Yizhak,K., Gillespie,S., Rodman,C., Luo,C.L., Mroz,E.A., Emerick,K.S. *et al.* (2017) Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell.*, **171**, 1611–1624.

77. Shi,W., Ng,C.K.Y., Lim,R.S., Jiang,T., Kumar,S., Li,X., Wali,V.B., Piscuoglio,S., Gerstein,M.B., Chagpar,A.B. *et al.* (2018) Reliability of whole-exome sequencing for assessing intratumor genetic heterogeneity. *Cell Rep.*, **25**, 1446–1457.