SEECancer: a resource for somatic events in evolution of cancer genome

Hongyi Zhang[†], Shangyi Luo[†], Xinxin Zhang[†], Jianlong Liao[†], Fei Quan, Erjie Zhao, Chenfen Zhou, Fulong Yu, Wenkang Yin, Yunpeng Zhang[*], Yun Xiao[*] and Xia Li[*]

College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150081, China

ABSTRACT

Cancer cells progressively evolve from a premalignant to a malignant state, which is driven by accumulating somatic alterations that confer normal cells a fitness advantage. Improvements in high-throughput sequencing techniques have led to an increase in construction of tumor phylogenetics and identification of somatic driver events that specifically occurred in different tumor progression stages. Here, we developed the SEECancer database (http://biocc.hrbmu.edu.cn/SEECancer), which aims to present the comprehensive cancer evolutionary stage-specific somatic events (including early-specific, late-specific, relapse-specific, metastasis-specific, drug-resistant and drug-induced genomic events) and their temporal orders. By manually curating over 10 000 published articles, 1231 evolutionary stage-specific genomic events and 5772 temporal orders involving 82 human cancers and 23 tissue origins were collected and deposited in the SEECancer database. Each entry contains the somatic event, evolutionary stage, cancer type, detection approach and relevant evidence. SEECancer provides a user-friendly interface for browsing, searching and downloading evolutionary stage-specific somatic events and temporal relationships in various cancers. With increasing attention on cancer genome evolution, the necessary information in SEECancer will facilitate understanding of cancer etiology and development of evolutionary therapeutics, and help clinicians to discover biomarkers for monitoring tumor progression.

INTRODUCTION

Cancer is a disease of the genome and the accumulation of genomic alterations leads to aggressive phenotypes such as increased proliferation, angiogenesis and invasion (1–3). Looking back into the long history of tumor research, there are many cases where mutation of important cancer genes occurred in specific stages of cancer evolution. Well-described examples include *DNMT3A* mutation and *TET2* mutation, which in most cases occur in the early evolutionary stage of acute myeloid leukemia (AML) (4–7). While *FLT3* mutation was reported by many studies as a late event in AML (8–10). These events occurring in different evolutionary stages exerted distinct influence on tumor progression. Early events are always considered to be pathogenesis-associated and responsible for tumor initiation, and thus are suitable as biomarkers for early diagnosis and targets for early intervention. Later events often confer additional aggressive hallmarks on cancer cells, which can be used as markers for tumor monitoring (11,12). In addition, the temporal order of the acquisition of cancer driver events was found to influence disease outcomes (13–15). For example, myeloproliferative neoplasm patients with *JAK2* mutation occurring first, as compared to those with *TET2* mutating first, are younger at the onset of the disease and are more likely to have thrombosis (16). Such temporal order allows to build and characterize the history of cancer evolution.

Large-scale cancer genomic sequencing projects, such as The Cancer Genome Atlas (17) (TCGA; http://cancergenome.nih.gov) and International Cancer genome Consortium (18) (ICGC; https://dcc.icgc.org) have identified millions of somatic mutations across the human cancers, with thousands of alleles to be implicated in disease causation. These data supported the development of numerous databases, such as cBioPortal (19) focusing on analyzing multidimensional cancer genomics data, COSMIC (20) providing a comprehensive resource of somatic mutations in cancer, DriverDB (21) aiming at novel driver identification. In recent years, multi-region sequencing, high-

[*]To whom correspondence should be addressed. Tel: +86 451 86615922; Fax: +86 451 86615922; Email: lixia@hrbmu.edu.cn
Correspondence may also be addressed to Yun Xiao. Email: xiaoyun@ems.hrbmu.edu.cn
Correspondence may also be addressed to YunPeng Zhang. Email: zyp19871208@126.com
[†]These authors contributed equally to the paper as first authors.

coverage whole-exome sequencing and single cell sequencing were applied to characterize tumor genome evolution. Such applications generated a large number of genetic alterations that preferentially occurred in distinct evolutionary stages of tumor development, providing a clear picture of evolutionary history in multiple tumor types (22,23). However, such evolutionary stage-specific events and their temporal orders are dispersed in thousands of published papers, without online repository collecting these information. To address this gap, we developed a manually curated database entitled 'SEECancer' (Somatic Events in Evolution of Cancer genome) (http://biocc.hrbmu.edu.cn/SEECancer) with the aim to provide a comprehensive resource of cancer evolutionary stage-specific events (Figure 1) and their temporal orders. As of July 2017, SEECancer documented more than 1200 manually curated evolutionary stage-specific events and more than 5700 temporal order relationships involving 82 human cancers. We hope that this elaborate database can serve as an important resource for future cancer evolution research.

DATA COLLECTION

Our data collection relied on expert manual curation of scientific publications. To obtain the papers associated with cancer genome evolution as many as possible, we searched the PubMed database using more than 100 combinations of different keywords (Supplementary Table S1), such as 'cancer evolution mutation', 'cancer evolution alteration', 'cancer clonal mutation', 'cancer clonal alteration', 'cancer timing mutation', 'cancer timing alteration', 'cancer temporal mutation', 'cancer temporal alteration', 'cancer order mutation', 'cancer order alteration', 'tumor evolution mutation' and 'neoplasm evolution mutation'. We obtained a total of 10 539 articles that include these keywords in the title or abstract, and then downloaded their abstracts through NCBI E-utilities API. These abstracts were manually checked by eight biological researchers, requiring that each abstract was reviewed by two biological researchers. In total, 1696 published papers that did study the evolution of cancer genome were selected. After that, eight biological researchers participated in reviewing the full texts of these 1696 papers, with each paper being scrutinized by two or more researchers. Two types of information were extracted 'by hand', including cancer somatic events tending to occur in specific evolutionary stages and temporal relationships between different events. Cancer evolutionary stage-specific somatic events were mainly collected from the text and tables of the papers, while temporal relationships were almost retrieved from the figures of the papers, which illustrated the 'phylogenetic tree' or 'progression path' of a particular cancer.

PROCESSING AND ANNOTATION

For each evolutionary stage-specific alteration, we carefully curated essential information, including gene/event symbol, cancer name, variant type (e.g. mutation, structural variation, LOH), evolutionary stage (i.e. early, late, relapse, metastasis, drug resistance, drug induced), study model (i.e. human, mouse, cell line), detection approach (low or high

throughout experiment or bioinformatics), cancer subtype, PubMed ID, publication year and journal. Moreover, we directly extracted the description of evidence from scientific publications. For somatic mutation events, the protein changes as a result of these mutations were also recorded. For drug associated events (including drug-resistant and drug-induced events), we further obtained the associated drug names. In this work, somatic events with annotations of 'early', 'trunk', 'founder', 'clonal' or 'initiation' in scientific publications usually occur in the early evolutionary stage of cancer and they were regarded as early-specific events. Somatic events with annotations of 'late' or 'subclonal' tend to occur in the late evolutionary stage and they were regarded as late-specific events. Notably, the separation of early- and late-specific events also depended on the context of scientific papers. In addition, care was taken to avoid duplication of entries within the database. Specifically, entries with identical gene/event names, cancer types, variant types and evolutionary stages were merged into one entry. One example is early mutation of *IDH1* in glioma, which was reported by twelve different studies. All entries from these studies were integrated into one entry. Another example is that *TP53* mutation was demonstrated as an early event of hepatocellular cancer by different researches in human tissue and mouse models (24–26). The entries derived from these researches were merged into one entry.

For temporal order, every entry contains seven annotation items, including antecedent event, subsequent event and their associated variant types, cancer name, method used to determine the temporal order and PubMed ID. Similarly, we merged duplicated entries with identical antecedent events, subsequent events, cancer types and variant types into one entry. Considering that the temporal orders were detected using different sample sizes, we set a weighted score for each temporal order relationship. Details about how to set the weighted score are available on the 'Help' page of the website.

Finally, we organized these evolutionary stage-specific events and temporal relationships into a standardized classification scheme based on tissue origin of cancers by reference to cBioPortal (27) and tumor classification vocabulary (https://www.oncolink.org/). Meanwhile, all human gene/protein names and synonyms were mapped to the official symbols of human genes provide by HUGO Gene Nomenclature Committee (www.genenames.org/).

DATA ACCESS

The SEECancer website is available online at http://biocc.hrbmu.edu.cn/SEECancer and requires no registration. It provides a user-friendly interface to the scientific community to browse, search and download. We have tested it in Mozilla Firefox, Google Chrome and Apple Safari browsers. All datasets including evolutionary stage-specific events and temporal order relationships are available as tab-delimited files on the 'Download' page. More recently, multi-region, multi-time-point, ultra-depth and single cell tumor sequencing have produced extensive data on cancer genome evolution, which will provide the opportunity to further extend SEECancer. We will continue to manually curate new evolutionary stage-specific alterations and tem-
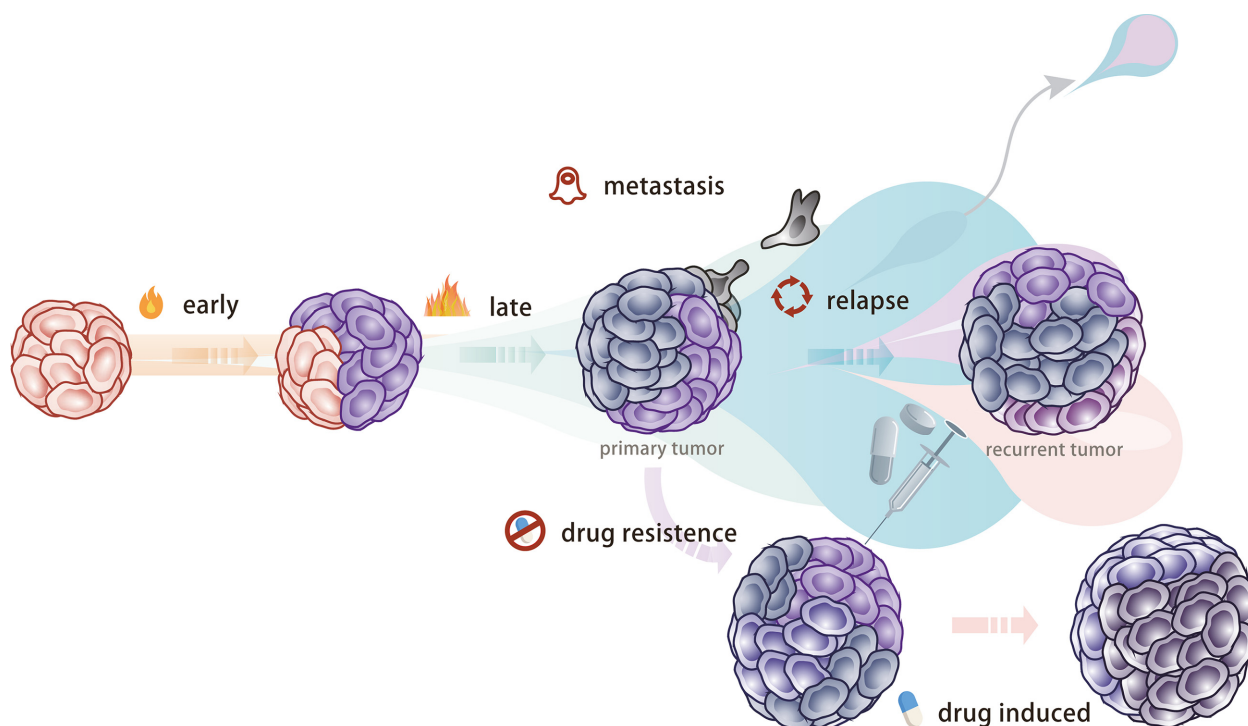
Figure 1. The cancer evolutionary stages studied in SEECancer.

poral order relationships. Our database will be updated four to six times per year depending on the number of newly released data.

RESULTS

Evaluating the reproducibility of manual curation

Reproducibility is always the main problem of manual curation. Here, we conducted analyses for evaluating the reproducibility of our curation processes. We first evaluated the reproducibility for the selection of the full papers from the abstracts. Two researchers were requested to review the same 500 abstracts randomly sampled from the original 10 539 articles for selecting the evolution-related full papers. After 2 weeks, the same two researchers re-selected the full papers from these 500 abstracts. By comparison, 49 and 51 full papers were selected by the two researchers at the first time point, respectively, and 47 (95.9%) and 49 (96.1%) papers were selected again two weeks later (Supplementary Figure S1A). The full papers selected by these two researchers showed high consistency at both of the two time points (90% at the first time point and 86.1% at the second time point) (Supplementary Figure S1A).

Likewise, we also evaluated the reproducibility of the manual curation of evolutionary stage-specific events from full papers. We randomly sampled 100 full papers from the 1696 selected full texts. This evaluation was conducted by another two researchers in the same way as described above. As a result, 89 and 84 somatic events were retrieved by the two researchers at the first time point, respectively, and 86 (96.6%) and 80 (95.2%) events were selected again two weeks later (Supplementary Figure S1B). The manual cura-

tion between the two researchers showed high reproducibility at both of the two time points (89.1% at the first time point and 88.9% at the second time point) (Supplementary Figure S1B). We noted that the reproducibility of manual curation for individual researchers between different time points is higher than that between different researchers at the same time point, suggesting that manual curation of every entry by two or more researchers is necessary to maintain the quality of the database.

SEECancer statistics

The current version of SEECancer consists of 1231 cancer evolutionary stage-specific somatic events (including 663 mutations, 385 copy number alterations, 87 structural variations, 69 methylation changes and 27 other alterations) involving 82 human cancers and 23 tissue origins (Figure 2A).

The majority of evolutionary stage-specific events came from early-specific events ($n = 727$, Figure 2B). SEECancer also contains evolutionary late-specific ($n = 201$), relapse-specific ($n = 73$) and metastasis-specific ($n = 70$) somatic alterations. In addition, during tumor evolution, some somatic events emerged under the selective pressure of drug treatment. We therefore collected drug-associated somatic events, including 146 drug-resistant and 14 drug-induced alterations (Figure 2B).

Among the 23 tissue origins, blood-derived cancers (e.g. AML and chronic lymphocytic leukemia) contain the most evolutionary stage-specific events ($n = 341$) (Figure 2C). One possible reason is that the samples of blood cancer are easier to collect, leading to substantial increases in studies dedicated to analyzing the genome evolution of blood cancer. Except for blood-derived cancers, SEECancer also
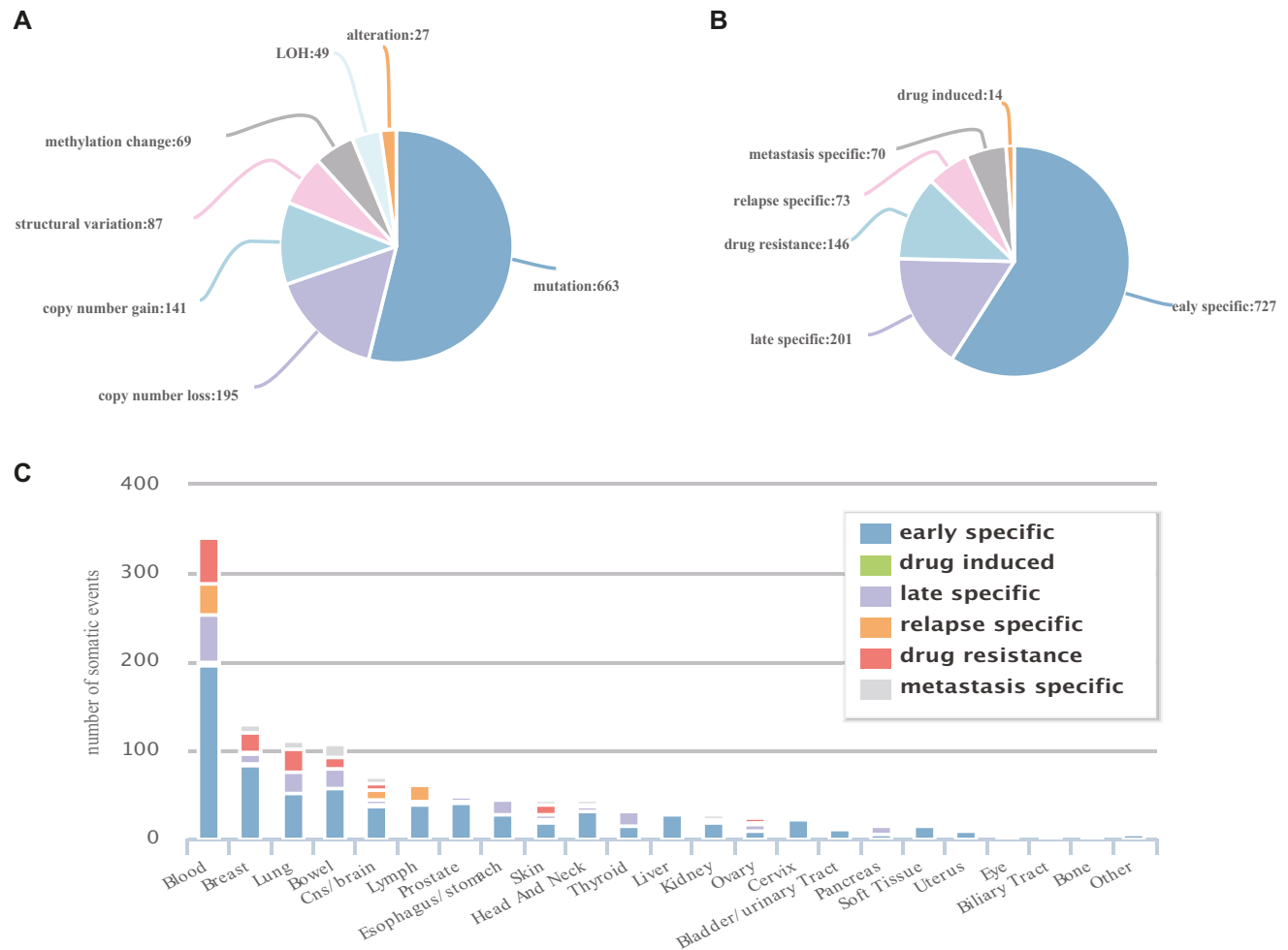
Figure 2. Statistics of evolutionary stage-specific events in SEECancer. (**A**) Distribution of evolutionary stage-specific events for different variant types. (**B**) Distribution of somatic events across different evolutionary stages. (**C**) The numbers of different evolutionary stage-specific somatic events related to different tissue origins.

contains a number of evolutionary stage-specific somatic events from other tissue origins, including breast ($n = 129$), bowel ($n = 111$), lung ($n = 107$), brain ($n = 71$), lymph ($n = 62$), prostate ($n = 56$), head and neck ($n = 45$), esophagus/stomach ($n = 45$), skin ($n = 44$), liver ($n = 34$), thyroid ($n = 32$), kidney ($n = 27$), ovary ($n = 25$), cervix ($n = 22$), urinary tract ($n = 19$), pancreas ($n = 17$), soft tissue ($n = 15$), uterus ($n = 14$), eye ($n = 4$), biliary tract ($n = 3$), bone ($n = 1$) and other ($n = 7$).

Another important feature of SEECancer is that it collected the temporal orders between somatic alterations in multiple cancers. Cancer evolution is dependent on antecedent events that influence the subsequent evolutionary trajectories, which is thus constrained by mutation order (15,28–30). It will be interesting for researchers to understand the order of genetic events in tumor development, which can influence the future clinical behavior of tumors. As of July 2017, there were 5772 temporal relationships involving 4250 various somatic events in 45 different cancer types deposited in SEECancer.

Website

We provide a user-friendly web interface that facilitates searching, browsing and retrieving cancer evolutionary stage-specific events and their temporal order relationships in the SEECancer. In the 'Home' page, users can quickly explore genetic events contributing to tumor development in specific evolutionary stages through clicking hyperlinks embedded in two web images—'Dynamic evolution of tumor' and 'Anatomical location map of common cancers' (Figure 3A). In the 'Browse' page, a particular hierarchical system was built, avoiding the need to type complex pathology terms and event symbols. Users can browse SEECancer by clicking a cancer name or a specific event, and the complete list of matched entries can be returned (Figure 3B). Two query sections 'EvolutionaryStage' and 'TemporalOrder' are provided for searching evolutionary stage-specific events and temporal order relationships, respectively. In addition, all data in the database can be downloaded in the 'Download' page, and a detailed tutorial showing how to use SEECancer is available on the 'Help' page.
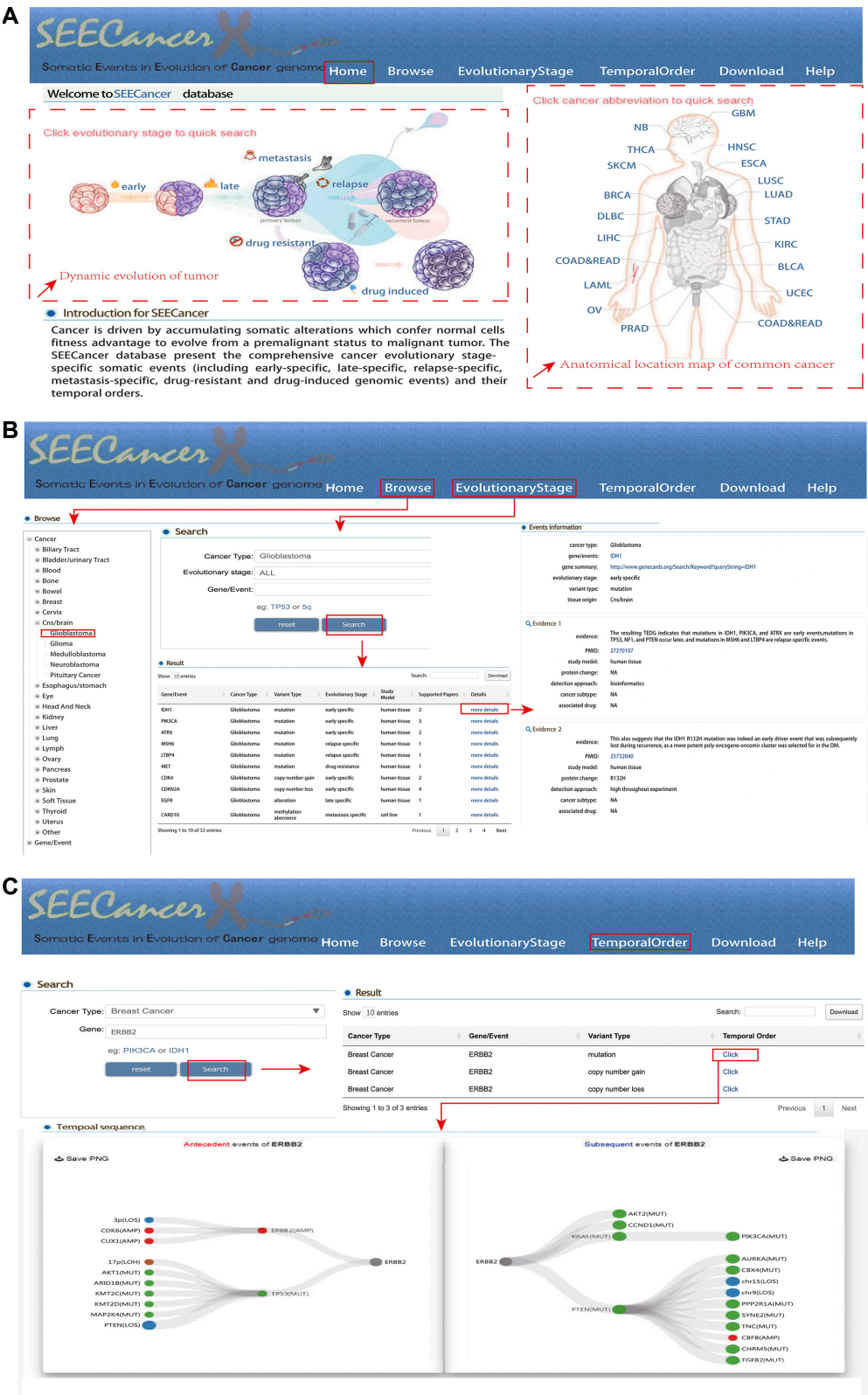
Figure 3. A schematic workflow of SEECancer. (**A**) The web images in the home page allow to quick research for evolutionary stage-specific events. (**B**) The 'Browse' and 'EvolutionaryStage' pages allow the users to browse and search evolutionary stage-specific events. (**C**) The 'TemporalOrder' page allows to search the antecedent events and subsequent events of queried alteration.

Query on the *'EvolutionaryStage'* page. To query evolutionary stage-specific events, SEECancer allows users to search by selecting cancer name or evolutionary stage from pull-down menu, or by entering a gene(event) name in the 'Gene/Event' search box (Figure 3B). The gene(event) symbols from SEECancer are auto-completed after typing some letters in the 'Gene/Event' search box. After clicking the 'Search' button, the query results will be displayed in a table showing the event name, variant type, cancer name, evolutionary stage, study model as well as the supported literature number. By clicking the hyperlink 'more details' for an individual event, users can obtain more detailed information for this event, including evidence descriptions and related PubMed IDs, associated drugs, protein change and the cross references to external databases. In addition, SEECancer provides several features for browsing the query results. First, users can filter entries by typing any term in the 'Search' field on the top right side above the table. Only entries matching the search term will be kept. Second, you can sort columns in an ascending (or a descending) fashion by clicking the arrows in the column headers. Third, the table size per page can be selected by using the 'Show entries' pull-down menu on the top left side above the table. The full query result can be saved as a tab-delimited file through the 'Download' button on the top right side.

Query on the *'TemporalOrder'* page. In the 'TemporalOrder' page, users can select a cancer type from pull-down menu in the 'cancer Type' search box, and/or enter a gene name in the 'gene' search box to search temporal orders for interested genes in a specific cancer type (Figure 3C). After clicking the 'Search' button, the search engine will return all matched somatic events. By choosing one somatic event of interest and then clicking the hyperlink in the column of 'Temporal order', its antecedent events and subsequent events can be obtained and visualized in a directed weighted tree.

Use of SEECancer

Besides general database search, browse and download, users can perform their specific researches based on the datasets in SEECancer. One way to analyze data from SEECancer is to look at the prevalence of events in specific evolutionary stages. We assessed those events involved in the early or late evolutionary stage and revealed that *TP53*, *KRAS* and *CDKN2A* alterations are the most common early events, which are associated with 28, 15 and 11 cancer types, respectively, providing further evidence for the importance of these genes in the onset of cancer (Figure 4A). Analysis of late events showed that the top 10 most common events were *TP53*, *PIK3CA*, *KRAS*, *PTEN*, *NRAS*, *SMAD4*, *FLT3*, *MYC*, *BRAF* and *CEBPA* alterations. Notably, among these highly prevalent late events, *TP53*, *KRAS* and *BRAF* alterations were also reported as early events in many cancer types (Figure 4A and C). In addition, the evolutionary information documented in the SEECancer allows us to have an overview of evolutionary stage-specific events and their related biological processes. We performed functional enrichment analysis for the early-specific and late-specific events, respectively (Figure 4B and

D). We noted that many critical functions were consistently enriched for early and late-specific events. For example, the immune system development was enriched in both classes, implying that immune dysfunction contributes to the whole trajectory of cancer evolution and suggesting the potential clinical benefits of immunotherapy in any stages of cancer evolution. Interestingly, the negative regulation of cell cycle pathway showed the most significant enrichment for early events, suggesting that the cell cycle abnormalities play important roles in the initiation of cancer. The aging biological process that refers to cellular senescence and apoptosis was highly enriched for late-specific events. This is probably due to the fact that resistance to cell senescence and escape from apoptosis can promote malignant transformation and resistance to therapy (31,32). These are just two ways to explore SEECancer and we look forward to seeing more researchers to confirm and consolidate the usefulness of this database.

DISCUSSION

In this work, we developed SEECancer as a resource that provides a comprehensive collection and assessment of somatic events in evolution of cancer genome. SEECancer is the first public database, to our best knowledge, focusing on cancer evolutionary stage-specific events and their temporal relationships. We know of several well-known databases (e.g. COSMIC, cBioPortal and DriverDB) that provide somatic mutations or driver genes in many human cancers, but without the evolutionary information, thus lacking a view of somatic events from the perspective of tumor evolution. Understanding the evolutionary dynamics of somatic events is very important for biologists to dissect cancer etiology and develop evolutionary therapeutics, and for clinicians to discover biomarkers for monitoring tumor progression. In recent years, numerous researches used high-coverage whole-exome/genome sequencing, multi-region/time-point sequencing or single cell sequencing to construct the evolutionary histories of tumors. Thus, the successful development of SEECancer will be expected to facilitate cancer evolution research.

Notably, in the database, certain alterations were reported to be early events in some studies, but in other studies, these alterations were found be late events in the same cancer type. We found ∼5% of events documented in both early and late evolutionary stages. As an example, *TP53* mutation was revealed as an early event driving the initiation of cancer in four studies (33–36). However, in the other four researches, *TP53* mutation was deemed as a late event favoring the expansion of *TP53* mutated clones in CLL (23,37–39). This phenomenon may be due to the high heterogeneity of cancer and the extraordinary diversity of clonal evolution.

In the SEECancer, the collected somatic events that preferentially occur in specific cancer evolutionary stages (termed 'evolutionary stage-specific events' here) can be considered as potential molecular targets and monitoring markers in future cancer treatment (11,40). For example, targeting early clonal events that present in every tumor cell may be an attractive model for drug development. Indeed, many successful targeted therapies take advantage of early clonal events present at all sites of disease (41). Considering
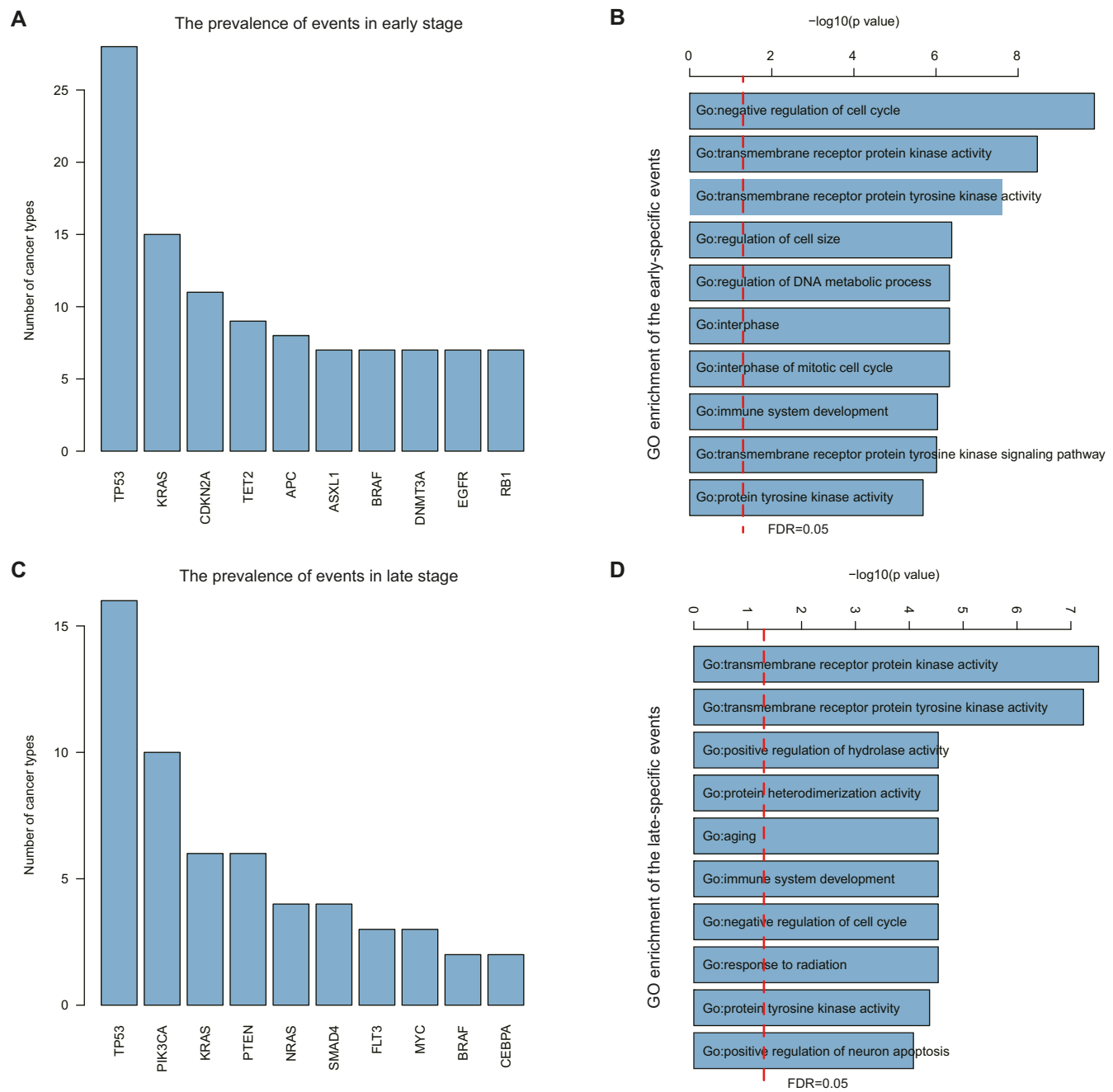
A The prevalence of events in early stage

B

C The prevalence of events in late stage

D

Figure 4. Analyses of evolutionary early and late-specific events and their associated biological processes. Histogram of the top 10 most prevalent events in the evolutionary early-specific stage (**A**) and evolutionary late-specific stage (**C**). The top 10 significantly enriched biological processes for early-specific events (**B**) and late-specific events (**D**).

the dynamic characteristics of tumor, monitoring its progression to guide therapeutic interventions has become a vital research area. In recent years, circulating tumor DNA (ctDNA) has the ability to capture somatic mutations acquired at different evolutionary stages, which will greatly facilitate the use of evolutionary stage-specific events to trace the tumor evolution (42,43). Therefore, the SEECancer database will help researchers to design biomarker panels of ctDNA sequencing for monitoring tumor progression. In addition, the temporal relationships of somatic

events can help researchers build links between genes underlying tumorigenesis, which will further enhance our understanding of cancer development. For instance, the early event *DNMT3A* mutation and its subsequent event *FLT3* mutation could define a high-risk AML subtype, which has a higher percentage of bone marrow blasts and extremely poor prognosis (44).

In summary, the SEECancer database will continue to grow and be an informative and valuable resource on cancer research. Further extensions will be conducted in the

following aspects. First, we are going to develop a unified framework to build the evolution atlas for major human cancers by combing their evolutionary trajectories. Second, we will collect evolutionary stage-specific somatic events at the individual level. Third, we will develop new visualized tools for displaying the complex temporal network of each cancer in SEECancer. Finally, as the cost of whole-genome sequencing is dropping rapidly, more somatic events in non-coding regions, such as events in regulatory elements, will be reported. Therefore, non-coding somatic events are expected to be collected in the future versions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Nowell,P.C. (1976) The clonal evolution of tumor cell populations. *Science*, **194**, 23–28.
2. Martincorena,I., Roshan,A., Gerstung,M., Ellis,P., Van Loo,P., McLaren,S., Wedge,D.C., Fullam,A., Alexandrov,L.B., Tubio,J.M. *et al.* (2015) Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*, **348**, 880–886.
3. Yates,L.R. and Campbell,P.J. (2012) Evolution of the cancer genome. *Nat. Rev. Genet.*, **13**, 795–806.
4. Papaemmanuil,E., Gerstung,M., Bullinger,L., Gaidzik,V.I., Paschka,P., Roberts,N.D., Potter,N.E., Heuser,M., Thol,F., Bolli,N. *et al.* (2016) Genomic classification and prognosis in acute myeloid leukemia. *N. Engl. J. Med.*, **374**, 2209–2221.
5. Kim,T., Yoshida,K., Kim,Y.K., Tyndel,M.S., Park,H.J., Choi,S.H., Ahn,J.S., Jung,S.H., Yang,D.H., Lee,J.J. *et al.* (2016) Clonal dynamics in a single AML case tracked for 9 years reveals the complexity of leukemia progression. *Leukemia*, **30**, 295–302.
6. Itzykson,R. and Solary,E. (2013) An evolutionary perspective on chronic myelomonocytic leukemia. *Leukemia*, **27**, 1441–1450.
7. Bochtler,T., Frohling,S. and Kramer,A. (2015) Role of chromosomal aberrations in clonal diversity and progression of acute myeloid leukemia. *Leukemia*, **29**, 1243–1252.
8. Jan,M., Snyder,T.M., Corces-Zimmerman,M.R., Vyas,P., Weissman,I.L., Quake,S.R. and Majeti,R. (2012) Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci. Transl. Med.*, **4**, 149ra118.
9. Corces-Zimmerman,M.R. and Majeti,R. (2014) Pre-leukemic evolution of hematopoietic stem cells: the importance of early mutations in leukemogenesis. *Leukemia*, **28**, 2276–2282.
10. Welch,J.S. (2014) Mutation position within evolutionary subclonal architecture in AML. *Semin. Hematol.*, **51**, 273–281.
11. Aparicio,S. and Caldas,C. (2013) The implications of clonal genome evolution for cancer medicine. *N. Engl. J. Med.*, **368**, 842–851.
12. Snuderl,M., Fazlollahi,L., Le,L.P., Nitta,M., Zhelyazkova,B.H., Davidson,C.J., Akhavanfard,S., Cahill,D.P., Aldape,K.D., Betensky,R.A. *et al.* (2011) Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell*, **20**, 810–817.
13. 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.
14. Fisher,R., Horswell,S., Rowan,A., Salm,M.P., de Bruin,E.C., Gulati,S., McGranahan,N., Stares,M., Gerlinger,M., Varela,I. *et al.* (2014) Development of synchronous VHL syndrome tumors reveals contingencies and constraints to tumor evolution. *Genome Biol.*, **15**, 433.
15. Swanton,C. (2015) Cancer evolution constrained by mutation order. *N. Engl. J. Med.*, **372**, 661–663.
16. Kent,D.G., Ortmann,C.A. and Green,A.R. (2015) Effect of mutation order on myeloproliferative neoplasms. *N. Engl. J. Med.*, **372**, 1865–1866.
17. Collins,F.S. and Barker,A.D. (2007) Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci. Am.*, **296**, 50–57.
18. International Cancer Genome,C., Hudson,T.J., Anderson,W., Artez,A., Barker,A.D., Bell,C., Bernabe,R.R., Bhan,M.K., Calvo,F., Eerola,I. *et al.* (2010) International network of cancer genome projects. *Nature*, **464**, 993–998.
19. Gao,J., Aksoy,B.A., Dogrusoz,U., Dresdner,G., Gross,B., Sumer,S.O., Sun,Y., Jacobsen,A., Sinha,R., Larsson,E. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.*, **6**, pl1.
20. Forbes,S.A., Beare,D., Boutselakis,H., Bamford,S., Bindal,N., Tate,J., Cole,C.G., Ward,S., Dawson,E., Ponting,L. *et al.* (2017) COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.*, **45**, D777–D783.
21. Chung,I.F., Chen,C.Y., Su,S.C., Li,C.Y., Wu,K.J., Wang,H.W. and Cheng,W.C. (2016) DriverDBv2: a database for human cancer driver gene research. *Nucleic Acids Res.*, **44**, D975–D979.
22. Johnson,B.E., Mazor,T., Hong,C., Barnes,M., Aihara,K., McLean,C.Y., Fouse,S.D., Yamamoto,S., Ueda,H., Tatsuno,K. *et al.* (2014) Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science*, **343**, 189–193.
23. Landau,D.A., Carter,S.L., Stojanov,P., McKenna,A., Stevenson,K., Lawrence,M.S., Sougnez,C., Stewart,C., Sivachenko,A., Wang,L. *et al.* (2013) Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*, **152**, 714–726.
24. Zhai,W., Lim,T.K., 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.
25. Pasquinelli,C., Bhavani,K. and Chisari,F.V. (1992) Multiple oncogenes and tumor suppressor genes are structurally and functionally intact during hepatocarcinogenesis in hepatitis B virus transgenic mice. *Cancer Res.*, **52**, 2823–2829.
26. Xue,R., Li,R., Guo,H., Guo,L., Su,Z., Ni,X., Qi,L., Zhang,T., Li,Q., Zhang,Z. *et al.* (2016) Variable intra-tumor genomic heterogeneity of multiple lesions in patients with hepatocellular carcinoma. *Gastroenterology*, **150**, 998–1008.
27. Cerami,E., Gao,J., Dogrusoz,U., Gross,B.E., Sumer,S.O., Aksoy,B.A., Jacobsen,A., Byrne,C.J., Heuer,M.L., Larsson,E. *et al.* (2012) The cBio cancer genomics portal: an open platform for

exploring multidimensional cancer genomics data. *Cancer Discov.*, **2**, 401–404.

28. 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.

29. Zhang,H., Deng,Y., Zhang,Y., Ping,Y., Zhao,H., Pang,L., Zhang,X., Wang,L., Xu,C., Xiao,Y. *et al.* (2017) Cooperative genomic alteration network reveals molecular classification across 12 major cancer types. *Nucleic Acids Res.*, **45**, 567–582.

30. Greaves,M. and Maley,C.C. (2012) Clonal evolution in cancer. *Nature*, **481**, 306–313.

31. Gimenez-Bonafe,P., Tortosa,A. and Perez-Tomas,R. (2009) Overcoming drug resistance by enhancing apoptosis of tumor cells. *Curr. Cancer Drug Targets*, **9**, 320–340.

32. Campisi,J. (2013) Aging, cellular senescence, and cancer. *Annu. Rev. Physiol.*, **75**, 685–705.

33. Lazarian,G., Tausch,E., Eclache,V., Sebaa,A., Bianchi,V., Letestu,R., Collon,J.F., Lefebvre,V., Gardano,L., Varin-Blank,N. *et al.* (2016) TP53 mutations are early events in chronic lymphocytic leukemia disease progression and precede evolution to complex karyotypes. *Int. J. Cancer*, **139**, 1759–1763.

34. Damm,F., Mylonas,E., Cosson,A., Yoshida,K., Della Valle,V., Mouly,E., Diop,M., Scourzic,L., Shiraishi,Y., Chiba,K. *et al.* (2014) Acquired initiating mutations in early hematopoietic cells of CLL patients. *Cancer Discov.*, **4**, 1088–1101.

35. Parker,H. and Strefford,J.C. (2016) The mutational signature of chronic lymphocytic leukemia. *Biochem. J.*, **473**, 3725–3740.

36. Wang,J., Khiabanian,H., Rossi,D., Fabbri,G., Gattei,V., Forconi,F., Laurenti,L., Marasca,R., Del Poeta,G., Foa,R. *et al.* (2014) Tumor

37. Rossi,D., Khiabanian,H., Spina,V., Ciardullo,C., Bruscaggin,A., Fama,R., Rasi,S., Monti,S., Deambrogi,C., De Paoli,L. *et al.* (2014) Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood*, **123**, 2139–2147.

38. Izzo,F. and Landau,D.A. (2016) Genetic and epigenetic determinants of B-cell lymphoma evolution. *Curr. Opin. Hematol.*, **23**, 392–401.

39. Rodriguez,D., Bretones,G., Arango,J.R., Valdespino,V., Campo,E., Quesada,V. and Lopez-Otin,C. (2015) Molecular pathogenesis of CLL and its evolution. *Int. J. Hematol.*, **101**, 219–228.

40. Turajlic,S. and Swanton,C. (2017) Implications of cancer evolution for drug development. *Nat. Rev. Drug Discov.*, **16**, 441–442.

41. McGranahan,N. and Swanton,C. (2017) Clonal heterogeneity and tumor evolution: past, present, and the future. *Cell*, **168**, 613–628.

42. Murtaza,M., Dawson,S.J., Tsui,D.W., Gale,D., Forshew,T., Piskorz,A.M., Parkinson,C., Chin,S.F., Kingsbury,Z., Wong,A.S. *et al.* (2013) Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*, **497**, 108–112.

43. Dawson,S.J., Tsui,D.W., Murtaza,M., Biggs,H., Rueda,O.M., Chin,S.F., Dunning,M.J., Gale,D., Forshew,T., Mahler-Araujo,B. *et al.* (2013) Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.*, **368**, 1199–1209.

44. Ahn,J.S., Kim,H.J., Kim,Y.K., Lee,S.S., Jung,S.H., Yang,D.H., Lee,J.J., Kim,N.Y., Choi,S.H., Jung,C.W. *et al.* (2016) DNMT3A R882 mutation with FLT3-ITD positivity is an extremely poor prognostic factor in patients with normal-karyotype acute myeloid leukemia after allogeneic hematopoietic cell transplantation. *Biol. Blood Marrow Transplant.*, **22**, 61–70.

evolutionary directed graphs and the history of chronic lymphocytic leukemia. *Elife*, **3**, doi:10.7554/eLife.02869.