SBCDDB: *Sleeping Beauty* Cancer Driver Database for gene discovery in mouse models of human cancers

Justin Y. Newberg[1,2,*], Karen M. Mann[1,2,*], Michael B. Mann[1,2], Nancy A. Jenkins[1,3,*] and Neal G. Copeland[1,3,*]

[1]Cancer Research Program, Houston Methodist Research Institute, Houston, Texas, USA, [2]Department of Molecular Oncology, H. Lee Moffitt Cancer Center, Tampa, FL, USA and [3]Genetics Department, University of Texas, MD Anderson Cancer Center, Houston, TX, USA

ABSTRACT

Large-scale oncogenomic studies have identified few frequently mutated cancer drivers and hundreds of infrequently mutated drivers. Defining the biological context for rare driving events is fundamentally important to increasing our understanding of the druggable pathways in cancer. *Sleeping Beauty* (SB) insertional mutagenesis is a powerful gene discovery tool used to model human cancers in mice. Our lab and others have published a number of studies that identify cancer drivers from these models using various statistical and computational approaches. Here, we have integrated SB data from primary tumor models into an analysis and reporting framework, the Sleeping Beauty Cancer Driver DataBase (SBCDDB, http://sbcddb.moffitt.org), which identifies drivers in individual tumors or tumor populations. Unique to this effort, the SBCDDB utilizes a single, scalable, statistical analysis method that enables data to be grouped by different biological properties. This allows for SB drivers to be evaluated (and re-evaluated) under different contexts. The SBCDDB provides visual representations highlighting the spatial attributes of transposon mutagenesis and couples this functionality with analysis of gene sets, enabling users to interrogate relationships between drivers. The SBCDDB is a powerful resource for comparative oncogenomic analyses with human cancer genomics datasets for driver prioritization.

INTRODUCTION

Sleeping Beauty (SB) insertional mutagenesis is a powerful system for cancer gene discovery in mouse models of can-cer. The transposon mutagenizes the genome by disrupting gene expression (1–3). The transposon contains an internal promoter and splice donor and, when inserted in-frame and upstream or in the 5′ end of an oncogene, can drive expression of downstream exons, thereby generating an oncogenic transcript. Alternatively, bi-directional splice acceptor sequences and transcriptional termination sites, also encoded in the transposon, can act as a gene trap, essentially disabling gene expression when the transposon is inserted into a tumor suppressor gene in either transcriptional orientation. Existing databases such as the RTCGD (4) or CCGD (http://ccgd-starrlab.oit.umn.edu/about.php) (5) organize gene lists from various transposon studies that have identified thousands of candidate cancer genes. A challenge arising from these large-scale efforts is prioritizing results, as different datasets often yield vastly different gene lists. Efforts to address this challenge include applying statistical learning and meta-analysis methods to evaluate the statistics reported in comparative studies (6). Another approach is to allow investigators to independently assess the validity of cancer genes by providing a digestible context behind the statistics. Efforts such as cBioPortal (http://www.cbioportal.org) (7,8) ICGC Data Portal (https://dcc.icgc.org) (9) and GDC Data Portal (https://portal.gdc.cancer.gov) (10) have provided investigators with tools for this regarding human cancer genomics data, but no such resource is available for mouse transposon data.

To address this need, we have developed a resource, the Sleeping Beauty Cancer Driver DataBase (SBCDDB, http://sbcddb.moffitt.org) that couples statistical tests with spatial information to provide context behind decisions identifying cancer driver genes. The database contains 19 primary mouse models of human tumor types from both published and provisional studies. This resource provides a step beyond identifying drivers to categorize drivers as acting in an oncogenic or tumor suppressive manner, based on the transposon insertion patterns. In addition to providing a

*To whom correspondence should be addressed. Tel: +1 813 745 1030; Fax: +1 813 745 3829; Email: Karen.Mann@moffitt.org
Correspondence may also be addressed to Justin Y. Newberg. Email: justin.newberg@moffitt.org
Correspondence may also be addressed to Nancy A. Jenkins. Email: njenkins1@mdanderson.org
Correspondence may also be addressed to Neal G. Copeland. Email: ncopeland1@mdanderson.org

Table 1. Published and unpublished mouse models of human cancer in SBCDDB

Datasets by Organ	Description	Ref.	Diagnoses within datasets	Tumors(n)	Mice(n)	Progression driver genes	Trunk driver genes
Bone	Osteosarcomas with *Trp53* R172H or *Trp53* wild-type alleles.	*Unpub.*	Osteosarcoma	45	36	131	41
Brain	Medulloblastomas with *Ptch1*-LacZ oncogenic allele.	(11)	Medulloblastoma	93	93	37	18
Breast	Breast tumors with *Pten* KO or *Pten* wild-type alleles.	(12)	Adenocarcinoma, adenoma, adenomyoepithelioma, adenosquamous carcinoma	39	33	137	36
Intestine	Intestine tumors with *Apc* KO, *Apc* Min, *Kras* G12D, *Smad4* KO, or *Trp53* R172H mutant alleles.	(13,14)	Intestine cancer (inferred)	839	190	673	30
Liver	Hepatocellular adenomas and carcinomas with *Trp53* KO, *Trp53* R172H, HBV, or wild-type background.	(15)	Hepatocellular adenoma, hepatocellular carcinoma (inferred)	597	110	2073	60
Lymph node and Thymus	Lymphomas with *Trp53* KO, *Trp53* R172H, or *Trp53* wild-type alleles.	*Unpub.*	B-cell lymphoma, lymphoblastic lymphoma (non-T non-B), T-cell lymphoblastic lymphoma, T-cell lymphoma	39	39	34	7
Muscle	Rhabdomyosarcomas with *Ptch1*-LacZ oncogenic allele.	*Unpub.*	Rhabdomyosarcoma	43	37	106	21
Pancreas	Pancreatic ductal adenocarcinomas with *Pdx1*-cre and *Kras* G12D or *Kras* wild-type alleles.	(16,17)	Pancreatic ductal early carcinoma, Pancreatic ductal adenocarcinoma	172	122	813	50
Prostate	Adenocaricnomas of the prostate with *Kras* G12D or *Kras* wild-type alleles.	*Unpub.*	Adenocarcinoma	54	53	182	32
Skin	Cutaneous keratoacanthomas, melanomas, and squamous cell carcinomas with mutant *Braf* V600E, *Trp53* KO, *Trp53* R172H, *Pten* KO alleles or wild-type background.	(18)	Keratoacanthoma, melanoma, squamous cell carcinoma	199	137	273	118
Spleen	Myeloid leukemias with *Trp53* KO, *Trp53* R172H or *Trp53* wild-type alleles.	(19)	Undifferentiated myeloid leukemia	168	168	34	29
Stomach	Gastric tumors with *Smad4* KO allele.	(20)	Gastric cancer (inferred)	66	24	770	27
Total*	-		–	**2354**	**956**	**2629**	**380**

* The non-redundant sum of progression and trunk driver genes across the various datasets. *Unpub.* are unpublished studies.
The datasets are grouped by the organ in which cancer develops, with an accompanying description of the respective tumor type(s). Published datasets include references to original reports. All other datasets are provisional. For most datasets a veterinary pathologist diagnosed tumors; in some cases diagnoses were inferred.

valuable aid for discovery of cancer drivers, the SBCDDB can be used to prioritize the long tail of low-frequency mutations in genes detected in human cancer genomes for hypothesis driven validation studies.

DATABASE CONTENTS AND DESIGN

Source data

The SBCDDB contains *Sleeping Beauty* transposon data derived from 2354 tumors representing 19 distinct mouse models of human cancer. These data were obtained from published and unpublished studies (Table 1). Allele information for these mouse models is included in Supplemental Table S1. Transposon insertions represented in this database were ascertained from tumors using the splinkerette-PCR method (2) and sequenced using the Roche 454 pyrosequencing platform. Here, we have represented these data in a comprehensive format to enable queries of statistically defined cancer drivers across SB models of human cancers.

Cancer drivers

Genes that drive transformation of normal cells and subsequent cancer progression are identified for each dataset using a gene-guided SB driver analysis pipeline. Within each

gene, SB mutations are categorized as exhibiting activating, inactivating, or indeterminate insertion patterns (Supplemental Figure S1). The pipeline is also applied to a filtered subset of high read-depth insertion sites for each dataset to identify candidate trunk (early initiating) drivers. Details regarding the determination of cancer drivers can be found on the FAQs page of the database and will be described in a companion manuscript. A secondary, locus-centric driver analysis pipeline was also applied (13) to provide additional genomic context for SB-driven tumors. Results from driver analysis have been compared with published studies and show significant overlap with previously reported gene lists, as expected, since the driver analysis incorporates design considerations from previous statistical analysis pipelines, such as gCIS (21). However, differences in the drivers reported in the SBCDDB compared to published studies do occur. Our driver pipeline applies standard criteria across all datasets to filter insertions included in statistical enrichment analysis to define drivers, applies a conservative FWER (family-wise error rate) multiple testing metric, corrects for known SB hotspots, and defines criteria for insertion pattern analysis. The gene sets in SBCDDB can be exported directly to Enrichr (http://amp.pharm.mssm.edu/Enrichr/) pathway analysis (22,23); this enables enrichment analysis of cancer drivers in known signaling pathways and
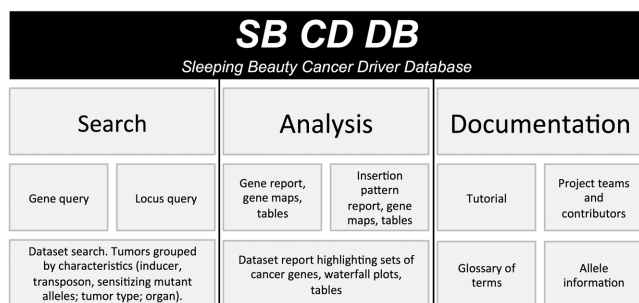
Figure 1. SBCDDB functionality. SBCDDB is comprised of three main sections. The 'Search' section on the main page enables users to query the database for a gene of interest using gene symbols or locus information. The output shows the frequency of insertions across tumor types modeled by SB and in which biological context(s) the gene is defined as a driver. The 'Analysis' section is accessed by clicking on the hyperlinked 'Search' output. Here, users can interrogate individual tumor models or biological contexts for a gene of interest. Clicking on the 'Dataset' hyperlink for a given tumor type provides details of all the defined drivers and the individual tumors within the dataset. Drivers defined by the biological context (organ, transposon, sensitizing allele, etc) can be found by clicking on the frequency charts. This analysis output can also be accessed from the main page by clicking on the pie chart under 'Datasets.' Finally, the 'Documentation' section of the SBCDDB provides a tutorial for users, defines the terms and allele information for the models, and provides additional information. This section can be accessed by a series of links at the top of each page.

other biological processes, including protein-protein interactions and transcription factor binding sites.

Architecture

The database contains *Sleeping Beauty* insertions in BED format (https://genome.ucsc.edu/FAQ/FAQformat.html), mouse and tumor annotations, and gene annotations. Tumors are grouped by biological properties to produce insertion datasets. Insertions are mapped to UCSC RefGene gene annotations, and cancer driver genes are statistically defined from these insertion datasets. Results of these analyses are cached in database tables (Supplemental Figure S2). A web-based querying and reporting framework sits on top of the database.

Functionality

Queries of the SBCDDB can be performed at the gene, locus and tumor-type level (Figure 1). Gene searches utilize official mouse (MGI, http://www.informatics.jax.org) and human (HGNC, http://www.genenames.org) nomenclature, and relate these to chromosomal locations in the mm9 genome reference. The Tutorial page details how to input search terms. Queries produce visual and numerical reports of SB insertions in a gene(s) of interest with mutational frequencies, and report enriched driver genes defined by experimental contexts. Graphics and tables presented in these reports can be downloaded as .svg and .txt files, respectively, for informational purposes.

WEBSITE FEATURES

Portal and navigation

The SBCDDB has an entry page that allows the user to navigate to different types of reports and documentation. Navigation between reports is also possible through blue hyperlinks and interactive SVG objects, which are recognizable by the mouse cursor changing from an arrow to a pointer. Orange hypertext denotes links to external resources.

Gene report

Gene queries are performed when a gene of interest is entered into the search field in the SBCDDB website interface (Figure 2). A query produces a report that cross-references the gene with external genomics and cancer data portals. The report highlights tumor models in which the gene is defined as a cancer driver. Furthermore, queries produce readouts that highlight insertion patterns within genes of interest and surrounding genomic areas; these visual maps allow users to independently assess driver gene status. Alternate gene identifiers (gene symbol synonyms) can be entered into the search field and up to five genes can be queried at one time. Users can view insertion patterns in individual datasets, which is useful for identifying context-dependent putative oncogenic or tumor suppressive behavior (Supplemental Figure S3).

Dataset report

Datasets in the SBCDDB are queried via a dropdown menu of pre-determined groupings based on biological parameters including tumor type, affected organ, transposon allele, sensitizing mutation and other phenotypic characteristics. A report of driver genes is generated for a given dataset (Figure 3). This dataset report outputs a co-occurrence matrix for the top eighty most frequently mutated drivers, providing a visual representation of relationships between genes across tumors in a dataset. To interrogate co-occurrence relationships between drivers, users can query individual genes of interest and click on the 'insertion pattern' hyperlink for a dataset of interest on the Gene report page to navigate to a list of tumors with mapped insertions in the driver. These data can be downloaded as .txt files and further annotated by the end user to statistically define co-occurrence relationships between genes using external methods.
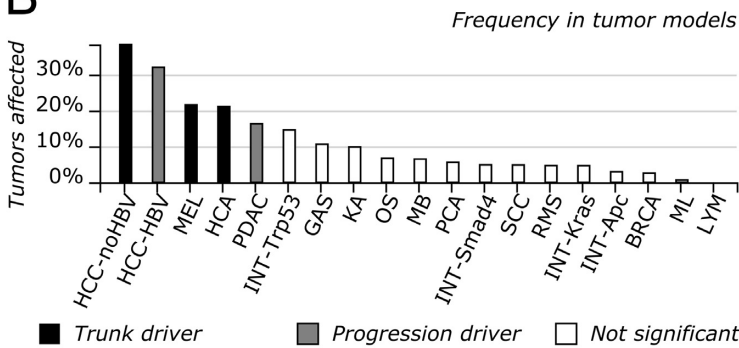
Pathway analysis

Driver gene sets in a Dataset report are presented as lists that can be automatically analyzed in Enrichr (22,23), identifying sets of enriched pathway that may be useful in future integrative studies. Enrichr was chosen because of its flexible API, the simple way in which it treats gene sets as inputs, its intuitive reporting interface and its annotations of a broad scope of biological pathways and processes. Notably, pathway analyses of SBCDDB drivers in primary tumor models recapitulate findings from published reports. For example, Wnt signaling is identified in intestinal tumors with an *Apc^Min* or *Apc* conditional knock-out allele (13), and various metabolic pathways are identified in liver tumors induced by HBV (15).
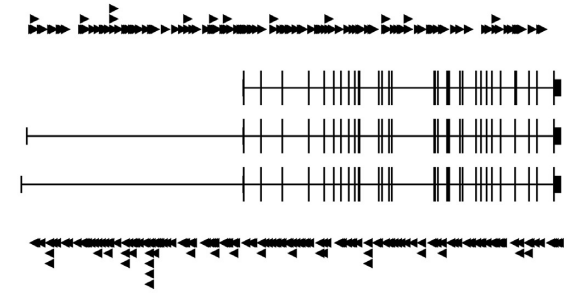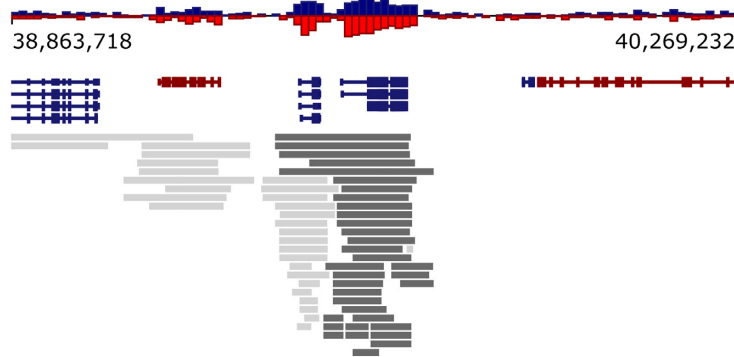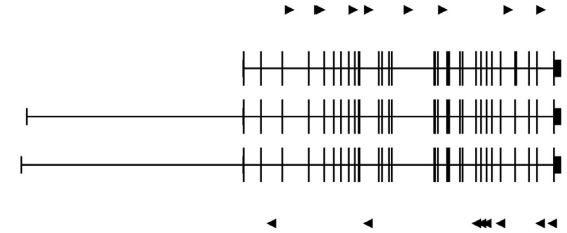
Figure 2. Gene report for *Stag2*, a tumor suppressor and driver. A gene report details conditions in which a gene may be a cancer driver. (**A**) Genes are matched to official mouse and human gene symbols, and links to external annotation and cancer databases are provided for the given gene. (**B**) The frequency of tumors with SB insertions in *Stag2* is shown across all tumor types, with an indication of whether the gene is a trunk driver (black bars), a progression driver (gray bars), or is not significantly altered (white bars). (**C**) An insertion map shows the various mapped insertions (triangles) in *Stag2* transcripts across all of the primary tumor models. Right facing arrows (above transcripts) show forward insertions, while left facing arrows (below transcripts) correspond to reverse insertions. The presence of insertions across the *Stag2* locus, in both the forward and reverse orientation, indicates that this locus is selectively inactivated. (**D**) A genomic context map roughly depicts insertion densities and genes around *Stag2*. Insertion densities are only shown in datasets in which *Stag2* is a driver. Blue bars on the top track represent forward insertion density (with respect to the chromosome), while red bars represent reverse insertion density. Transcripts are drawn below this, and their strand is denoted by their color (blue = +, red = −). Genomic regions containing significant insertion densities are shown in the gray bars, with the darker bars associated with the gene of interest. E) The insertion map for *Stag2* is shown for melanoma. BRCA, breast cancer; GAS, gastric cancer; HCA, hepatocellular adenoma; HCC, hepatocellular carcinoma; INT, Intestinal cancers; KA, keratoacanthoma; LYM, lymphoma; MB, medulloblastoma; MEL, melanoma; ML, myeloid leukemia; OS, osteosarcoma; PCA, prostate cancer; PDAC, pancreatic ductal adenocarcinoma; RMS, rhabdomyosarcoma; SCC, cutaneous squamous cell carcinoma.

Tumor page

When individual tumors are selected from dataset pages, various properties are highlighted that integrate insertion data across the tumors. For example, one set of metrics is the number of forward and reverse reads, which may provide insight into whether a tumor is driven by activation of proto-oncogenes. These insertion properties provide additional information that supplements the gene set and pathway annotations that can be used to categorize the tumors.
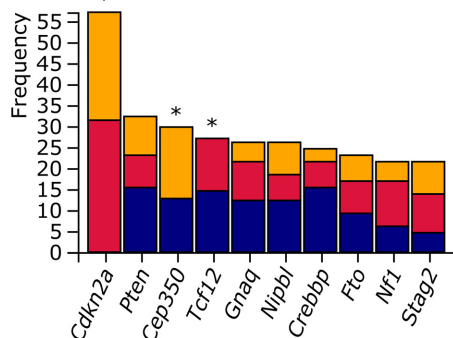
A

Description	Pubmed Ids	Investigators	Diagnosis	Organs	Number of mice	Number of tumors	Progression drivers
A melanoma model consisting of tumors with a *Braf* V600E allele.	25848750	Michael Mann	Melanoma	Skin	52	65	151 E

B
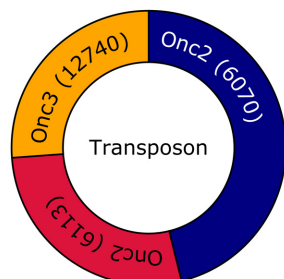
C

D
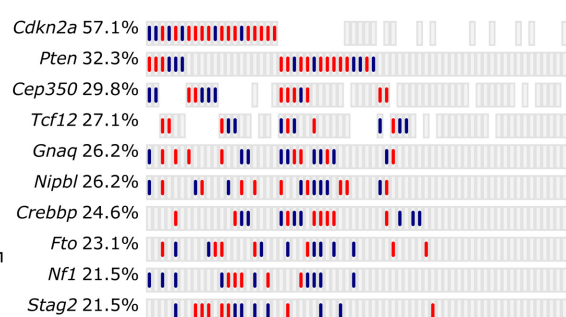
E

Gene ▲	Pattern	Tumors with insertions ▼	P-value	Reads per tumor
Pten	Inactivating	21	< 1e-12	162.2
Cdkn2a	Inactivating	20	< 1e-12	103.7
Crebbp	Inactivating	16	< 1e-12	21.3
Cep350	Inactivating	14	< 1e-12	127.7
Sae1	Indeterminate	10	< 1e-12	1.7

F

Tumor	Number drivers	Inducer	Transposon	Sensitizing mutation
0555-001	53 E	*Tyr*-creERT2	Onc2 (6113)	*Braf* CA
0544-001	50 E	*Tyr*-creERT2	Onc2 (6113)	*Braf* CA
0546-003	47 E	*Tyr*-creERT2	Onc3 (12740)	*Braf* CA
0575-001	43 E	*Tyr*-creERT2	Onc2 (6070)	*Braf* CA
0563-001	41 E	*Tyr*-creERT2	Onc2 (6113)	*Braf* CA

Figure 3. Dataset report for melanoma model highlighting sets of mutated cancer genes. Cancer models can be used to identify sets of significantly altered genes across tumors. (**A**) A summary table provides general information about the melanoma dataset with a list of significantly altered genes. Pathway analysis via Enrichr can be run on this set of genes. (**B**) The melanoma dataset consists of tumors derived from various transposons. (**C**) The top ten most frequently mutated statistically significant genes are shown along with the proportion of tumors for each transposon. Datasets can contain tumors with different transposons; the chromosome containing the original SB transposon array is called the donor. Genes that map to a donor chromosome in some tumors can still be identified as significant in other tumors with a different donor (asterisks). (**D**) A waterfall plot shows occurrences of frequently mutated melanoma genes. Rows are genes and columns are tumors. Blue and red hashmarks denote forward and reverse insertions, respectively. White gaps appear when a gene is on the donor chromosome. (**E**) A table of statistically significant melanoma genes with metrics that can be used in subsequent analysis to rank genes of interest. (**F**) Significantly altered melanoma genes can be viewed in each melanoma tumor, and these gene sets can be run through Enrichr for pathway analysis. *Cdkn2a*, Cyclin-dependent kinase inhibitor 2A; *Lpp*, LIM domain containing preferred translocation partner in lipoma; *Pten*, Phosphatase and tensin homolog; *Cep350*, Centrosomal protein 350; *Tcf12*, Transcription factor 12; *Gnaq*, Guanine nucleotide binding protein, alpha q polypeptide; *Nipbl*, Nipped-B homolog (Drosophila); *Crebbp*, CREB binding protein; *Fto*, Fat mass and obesity associated; *Setd2*, SET domain containing 2; *Nf1*, Neurofibromin 1; *Sae1*, SUMO1 activating enzyme subunit 1.

Documentation pages

The Tutorial page of the SBCDDB contains visual tutorials that guide the user through the steps to generate reports. Within each report are toggles that provide explanations of the different graphics and tables. A Frequently Asked Questions (FAQs) page provides answers to questions regarding specific data analysis decisions, underlying datasets in the SBCDDB, and how users may contribute additional datasets. *Sleeping Beauty* insertional mutagenesis terms are defined in a glossary. Finally, contact information for the database is provided in the About Us page.

SOFTWARE

Data is stored server-side in an SQLite database. Web pages are served by Apache, which uses mod_wsgi (https://github.com/GrahamDumpleton/mod_wsgi) to invoke Python scripts to retrieve and process data. Data is served in response to AJAX requests from the client. Client-side data analysis and reporting is performed in JavaScript using the D3 (https://d3js.org) (24), d3-tip (https://github.com/Caged/d3-tip), jQuery (https://jquery.com), and dataTables (https://datatables.net) libraries. Users should access the SBCDDB using an HTML5 and

ECMAScript 6 supported browser, such as Chrome, Firefox, or Safari. Software details and a disclaimer for use of the SBCDDB can be found in the Terms of Use page.

DISCUSSION

The SBCDDB is an interactive, integrative resource of transposon insertion data from mouse models of human cancers driven by *Sleeping Beauty* mutagenesis. The SBCDDB addresses a need to integrate data from SB transposon datasets to define high-confidence cancer drivers across various tumor types. The principles guiding the design and implementation of SBCDDB include a desire to provide rich visualization and data analysis, features that enable (i) a systematic, comprehensive and scalable analysis approach to define cancer drivers; (ii) the identification and classification of oncogenic and tumor suppressive drivers from a variety of tumor histologies and (iii) the ability to define cooperating and mutually exclusive relationships between cancer drivers in mouse models of cancer. Our efforts to present SB data in an accessible format empowers the cancer research community to interrogate cancer drivers identified in SB mouse models and integrate these findings with data from their own labs and with publicly available human genomic data from complementary cancer driver databases, such as cBioPortal (7,8), the ICGC Data Portal (9) and the GDC Data Portal (10). The cancer drivers identified by SB mutagenesis provide a powerful comparative dataset for identifying the biologically important cancer drivers present in the long tail of rare mutations found in human tumors, thereby increasing our understanding of the druggable pathways for cancer therapeutics. SB mutagenesis may also help identify human cancer genes that are deregulated by mechanisms other than somatic mutations, such as copy number alterations, chromatin modifications, DNA methylation changes and mutations in distal enhancers. We believe that the SBCDDB will provide a valuable tool for defining and prioritizing new drivers of cancer initiation and progression in support of developing promising novel ways to prevent and treat cancers.

AVAILABILITY

The SBCDDB can be accessed at http://sbcddb.moffitt.org.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Dr David Steffen, Dr Mik Black, and the members of the Copeland-Jenkins lab for their helpful suggestions and comments during the development of the SBCDDB. The authors wish to thank David Adams and Alistair Rust for 454 sequencing and informatics support and Dr Kenneth Hon Kim Ban and Dr Ching Ging Ng for providing datasets prior to publication for prostate and rhabdomyosarcoma, respectively. N.G.C. and N.A.J. are CPRIT Scholars in Cancer Research.

REFERENCES

1. Copeland,N.G. and Jenkins,N.A. (2010) Harnessing transposons for cancer gene discovery. *Nat. Rev. Cancer*, **10**, 696–706.
2. Mann,K.M., Jenkins,N.A., Copeland,N.G. and Mann,M.B. (2014) Transposon insertional mutagenesis models of cancer. *Cold Spring Harbor Protoc.*, **2014**, pdb.top069849.
3. Mann,M.B., Jenkins,N.A., Copeland,N.G. and Mann,K.M. (2014) Sleeping Beauty mutagenesis: exploiting forward genetic screens for cancer gene discovery. *Curr. Opin. Genet. Dev.*, **24**, 16–22.
4. Akagi,K., Suzuki,T., Stephens,R.M., Jenkins,N.A. and Copeland,N.G. (2004) RTCGD: retroviral tagged cancer gene database. *Nucleic Acids Res.*, **32**, D523–D527.
5. Abbott,K.L., Nyre,E.T., Abrahante,J., Ho,Y.-Y., Isaksson Vogel,R. and Starr,T.K. (2015) The Candidate Cancer Gene Database: a database of cancer driver genes from forward genetic screens in mice. *Nucleic Acids Research*, **43**, D844–D848.
6. Tokheim,C.J., Papadopoulos,N., Kinzler,K.W., Vogelstein,B. and Karchin,R. (2016) Evaluating the evaluation of cancer driver genes. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 14330–14335.
7. 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.
8. 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.
9. Zhang,J., Baran,J., Cros,A., Guberman,J.M., Haider,S., Hsu,J., Liang,Y., Rivkin,E., Wang,J., Whitty,B. *et al.* (2011) International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. *Database*, **2011**, bar026.
10. Grossman,R.L., Heath,A.P., Ferretti,V., Varmus,H.E., Lowy,D.R., Kibbe,W.A. and Staudt,L.M. (2016) Toward a shared vision for cancer genomic data. *N. Engl. J. Med.*, **375**, 1109–1112.
11. Genovesi,L.A., Ng,C.G., Davis,M.J., Remke,M., Taylor,M.D., Adams,D.J., Rust,A.G., Ward,J.M., Ban,K.H., Jenkins,N.A. *et al.* (2013) Sleeping Beauty mutagenesis in a mouse medulloblastoma model defines networks that discriminate between human molecular subgroups. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E4325–E4334.
12. Rangel,R., Lee,S.-C., Hon-Kim Ban,K., Guzman-Rojas,L., Mann,M.B., Newberg,J.Y., Kodama,T., McNoe,L.A., Selvanesan,L., Ward,J.M. *et al.* (2016) Transposon mutagenesis identifies genes that cooperate with mutant Pten in breast cancer progression. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E7749–E7758.
13. March,H.N., Rust,A.G., Wright,N.A., ten Hoeve,J., de Ridder,J., Eldridge,M., van der Weyden,L., Berns,A., Gadiot,J., Uren,A. *et al.* (2011) Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. *Nat. Genet.*, **43**, 1202–1209.
14. Takeda,H., Wei,Z., Koso,H., Rust,A.G., Yew,C.C.K., Mann,M.B., Ward,J.M., Adams,D.J., Copeland,N.G. and Jenkins,N.A. (2015) Transposon mutagenesis identifies genes and evolutionary forces driving gastrointestinal tract tumor progression. *Nat. Genet.*, **47**, 142–150.
15. Bard-Chapeau,E.A., Nguyen,A.-T., Rust,A.G., Sayadi,A., Lee,P., Chua,B.Q., New,L.-S., de Jong,J., Ward,J.M., Chin,C.K.Y. *et al.* (2014) Transposon mutagenesis identifies genes driving hepatocellular carcinoma in a chronic hepatitis B mouse model. *Nat. Genet.*, **46**, 24–32.

16. Mann,K.M., Ward,J.M., Yew,C.C.K., Kovochich,A., Dawson,D.W., Black,M.A., Brett,B.T., Sheetz,T.E., Dupuy,A.J., Initiative,A.P.C.G. *et al.* (2012) Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. *Proc. Natl. cad. Sci. U.S.A.*, **109**, 5934–5941.

17. Perez-Mancera,P.A., Rust,A.G., van der Weyden,L., Kristiansen,G., Li,A., Sarver,A.L., Silverstein,K.A.T., Grutzmann,R., Aust,D., Rummele,P. *et al.* (2012) The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. *Nature*, **486**, 266–270.

18. Mann,M.B., Black,M.A., Jones,D.J., Ward,J.M., Yew,C.C.K., Newberg,J.Y., Dupuy,A.J., Rust,A.G., Bosenberg,M.W., McMahon,M. *et al.* (2015) Transposon mutagenesis identifies genetic drivers of BrafV600E melanoma. *Nat Genet*, **47**, 486–495.

19. Mann,K.M., Newberg,J.Y., Black,M.A., Jones,D.J., Amaya-Manzanares,F., Guzman-Rojas,L., Kodama,T., Ward,J.M., Rust,A.G., van der Weyden,L. *et al.* (2016) Analyzing tumor heterogeneity and driver genes in single myeloid leukemia cells with SBCapSeq. *Nat. Biotech.*, **34**, 962–972.

20. Takeda,H., Rust,A.G., Ward,J.M., Yew,C.C.K., Jenkins,N.A. and Copeland,N.G. (2016) Sleeping Beauty transposon mutagenesis identifies genes that cooperate with mutant Smad4 in gastric cancer development. *Proc. Natl. Acad. Sci. U.S.A.s*, **113**, E2057–E2065.

21. Brett,B.T., Berquam-Vrieze,K.E., Nannapaneni,K., Huang,J., Scheetz,T.E. and Dupuy,A.J. (2011) Novel molecular and computational methods improve the accuracy of insertion site analysis in sleeping beauty-induced tumors. *PLOS ONE*, **6**, e24668.

22. Chen,E.Y., Tan,C.M., Kou,Y., Duan,Q., Wang,Z., Meirelles,G.V., Clark,N.R. and Ma'ayan,A. (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, **14**, 128–128.

23. Kuleshov,M.V., Jones,M.R., Rouillard,A.D., Fernandez,N.F., Duan,Q., Wang,Z., Koplev,S., Jenkins,S.L., Jagodnik,K.M., Lachmann,A. *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.

24. Bostock,M., Ogievetsky,V. and Heer,J. (2011) D^3 Data-Driven Documents. *IEEE Trans. Vis. Comput. Graph.*, **17**, 2301–2309.