



Original article

CTD² Dashboard: a searchable web interface to connect validated results from the Cancer Target Discovery and Development Network

Bülent Arman Aksoy^{1,†}, Vlado Dančik^{2,†}, Kenneth Smith^{3,†},
Jessica N. Mazerik⁴, Zhou Ji³, Benjamin Gross¹, Olga Nikolova⁵,
Nadia Jaber⁴, Andrea Califano³, Stuart L. Schreiber², Daniela S.
Gerhard⁴, Leandro C. Hermida⁴, Subhashini Jagu⁴, Chris Sander^{1,*},
Aris Floratos^{3,6,*} and Paul A. Clemons^{2,*}

¹Computational Biology Center, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA, ²Chemical Biology and Therapeutics Science Program, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA, ³Department of Systems Biology, Columbia University, New York, NY 10032, USA, ⁴Office of Cancer Genomics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA, ⁵Computational Biology Program, School of Medicine, Oregon Health and Science University, Portland, OR 97239, USA and ⁶Department of Biomedical Informatics, Columbia University, New York, NY 10032, USA

*Corresponding author: Tel: +1 617 714 7346; Fax: +1 617 714 8969; Email: pclemons@broadinstitute.org

Correspondence may also be addressed to Chris Sander: Tel: +1 617 5829717; Email: sander.research@gmail.com

Correspondence may also be addressed to Aris Floratos: Tel +1 212 8515139; Email: floratos@c2b2.columbia.edu

[†]These authors contributed equally to this work.

Present addresses: Bülent Arman Aksoy, Icahn School of Medicine, Mount Sinai, New York, NY 10029, USA; Chris Sander, ^scBio Center, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02215, USA.

Citation details: Aksoy, B.A., Dančik, V., Smith, K. *et al.* CTD² dashboard: a searchable web interface to connect validated results from the cancer target discovery and development network. *Database* (2017) Vol. 2017: article ID bax054; doi:10.1093/database/bax054

Received 30 January 2017; Revised 25 May 2017; Accepted 25 June 2017

Abstract

The Cancer Target Discovery and Development (CTD²) Network aims to use functional genomics to accelerate the translation of high-throughput and high-content genomic and small-molecule data towards use in precision oncology. As part of this goal, and to share its conclusions with the research community, the Network developed the 'CTD² Dashboard' [<https://ctd2-dashboard.nci.nih.gov/>], which compiles CTD² Network-generated conclusions, termed 'observations', associated with experimental entities, collected by its member groups ('Centers'). Any researcher interested in learning about a given gene, protein, or compound (a 'subject') studied by the Network can come to the CTD² Dashboard to quickly and easily find, review, and understand Network-generated experimental results. In particular, the Dashboard allows visitors to connect experiments

about the same target, biomarker, etc., carried out by multiple Centers in the Network. The Dashboard's unique knowledge representation allows information to be compiled around a subject, so as to become greater than the sum of the individual contributions. The CTD² Network has broadly defined levels of validation for evidence ('Tiers') pertaining to a particular finding, and the CTD² Dashboard uses these Tiers to indicate the extent to which results have been validated. Researchers can use the Network's insights and tools to develop a new hypothesis or confirm existing hypotheses, in turn advancing the findings towards clinical applications.

Database URL: <https://ctd2-dashboard.nci.nih.gov/>

Introduction

The current Cancer Target Discovery and Development (CTD²) initiative is a collaborative group of 13 different research teams, or Centers, organized under the auspices of the National Cancer Institute's Office of Cancer Genomics. It aims to functionally validate discoveries from large-scale adult and pediatric cancer-genome characterization initiatives, and advance them towards use in precision medicine. CTD² Centers use high-throughput experimental and bioinformatic approaches to mine the data and find alterations that potentially influence tumor biology. The Network characterizes the functional roles of candidate alterations in cancers, and identifies novel approaches that target causative alterations through their associated biological pathways. Through the individual and synergistic efforts of the Centers and broad sharing of data with the research community, CTD² contributes to understanding the mechanisms of cancer initiation, maintenance, and progression, and supports the accelerated development of clinically useful markers, targets, and therapeutics.

CTD² is a 'community resource project' [<https://www.genome.gov/pages/research/wellcomereport0303.pdf>] meaning all data are openly available to the scientific community and can be accessed without restrictions. The Network as a whole applies a wide range of investigational approaches, including computational biology, genome-wide loss-of-function or gain-of-function *in vitro* and *ex vivo* screening, high-throughput small-molecule screening, and protein-protein interaction analysis, among others. The web-based open-access CTD² Data Portal [<https://ocg.cancer.gov/programs/ctd2/data-portal/>] is the primary outlet by which the Network shares these diverse datasets (positive and negative results generated as part of each Center's research) and the associated project descriptions and methodologies. Although the Network has a dedicated process to harmonize the data, a certain level of bioinformatics expertise is needed to use the data in the Data Portal to their fullest potential. The data in the Portal are generated with technology-

relevant controls and undergo quality assessment, but have not been independently validated [https://ocg.cancer.gov/sites/default/files/CTD2CaveatEmptor_final.pdf]. Therefore, researchers who are using the data must further validate findings in order to determine the significance and strength of a particular result. Network results are shared through published manuscripts, which contain data and related information. However, these data are not always provided in an easily digestible format for all types of researcher. Such publications concentrate on the biological relevance of a few findings the authors find interesting, although in large-scale high-throughput experiments, there may be many other positive findings.

The CTD² Dashboard [<https://ctd2-dashboard.nci.nih.gov/>] addresses these data-access and usability challenges and fills the need for a searchable and browsable web interface that assembles concisely summarized results, connects them with subsequent evidence that reinforces or builds upon the original finding, and conveys the extent to which the results have been functionally validated. The Dashboard was developed to address the need of the community to find data generated by CTD², and to adhere to the FAIR (findable, accessible, interoperable and reusable) principles (1). The Dashboard was designed to allow easy navigation and use by a range of scientists, including both computationalists and non-computational cancer experts.

Database structure and organization

The Dashboard concept

The CTD² 'Dashboard' (which was developed by Centers at the Broad Institute, Cold Spring Harbor Laboratories/Memorial Sloan Kettering Cancer Center, and Columbia University, with input from the entire CTD² Network) compiles CTD² Network-generated conclusions, termed 'observations', associated with particular experimental subjects. Any researcher interested in learning about a particular gene, protein, or compound that has been the

subject of experimentation by Network members can come to the CTD² Dashboard to quickly and easily find, review, and download Network-generated results. In particular, the CTD² Dashboard allows users rapidly to connect different cancer biology experiments about the same biological entity, including its function in the experiment, carried out by multiple Centers in the Network.

Because of the diversity of data types generated by the contributing Centers, the concept of the CTD² Dashboard needed to be very flexible. Indeed, one of the early conceptual challenges was to provide for a situation where the types of experiments to be performed were not known in advance. As described in detail in later sections, we addressed this challenge by defining a small number of subjects that represent actual biological entities (genes, proteins, etc.), but did not pre-define the types of data that could be connected. Instead, we articulated a system of evidence that is defined in terms of *electronic* data types (e.g. numbers, text labels, images, URLs, etc.). This strategy enables the CTD² Dashboard to connect observations from diverse types of experiments and analyses to represented subjects (Figure 1).

To distinguish the *strength* of different types of evidence, the CTD² Network has developed broadly defined Tiers of validation for evidence pertaining to a particular finding (2). Specifically, Tier 1 evidence comprises high-throughput biology, such as screening or profiling experiments, unbiased or genome-wide computational analyses, and other similar hypothesis-generating activities. Evidence is classified as Tier 2 when an initial hypothesis is

validated *in vitro*, such as with protein-binding assays or cell-culture models. Tier 3 evidence is supported by validation of a hypothesis *in vivo*, such as in mice, organoids, or other models. Through the CTD² Dashboard and application of the Tier ranking system, any investigator can utilize the Network-generated experimental results as a starting point to extend or confirm hypotheses.

Observations—atomic units of the CTD² Dashboard

The main challenge in developing the CTD² Dashboard was devising a data model capable of representing datasets generated by the large and diverse collection of computational and experimental assays performed by the CTD² Network Centers (Table 1). Our solution was to organize data around the concept of an *observation*. Observations are statements about biological or chemical entities (e.g., genes, compounds, cell lines, etc.) that concisely summarize experimental results. As described above, CTD² Dashboard observations are annotated with Tiers based on the strength of the supporting evidence. The highest, Tier 3, is assigned to observations representing hypotheses validated *in vivo*. Typically, these observations correspond to research findings reported in the literature and are the most comprehensively validated results generated from analysis of Network data. For example, a high-throughput screen designed to assess how different genetic backgrounds influence response to therapeutics could give rise to the following observation: ‘Across many cancer types, cell lines *sensitive* to *BIX-01294* are enriched in cells

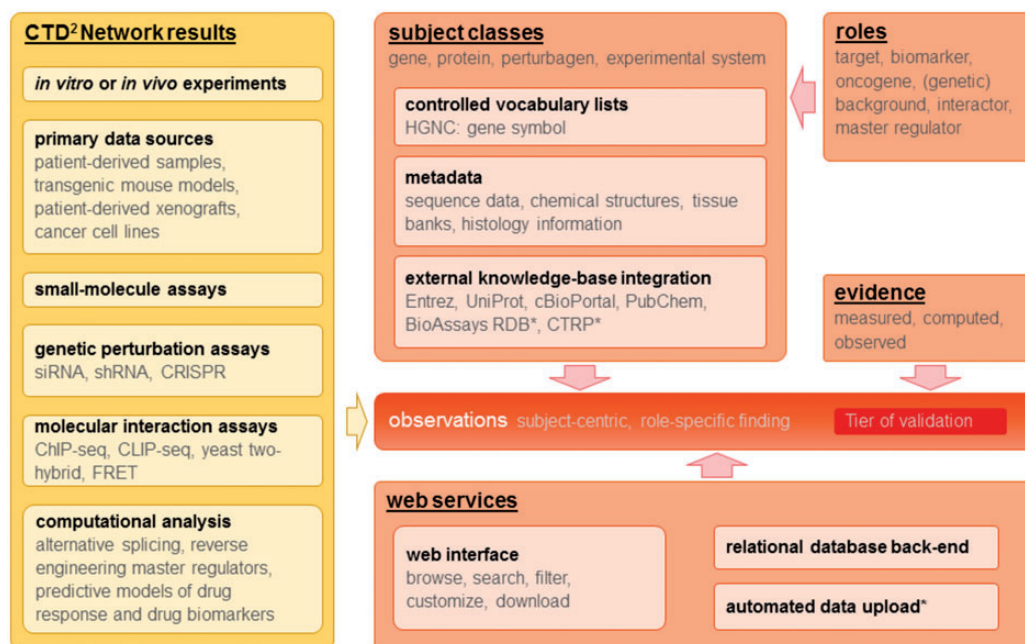


Figure 1. Overview of CTD² Dashboard data types, key concepts, and web services. Items marked with (*) represent work in progress or future plans.

Table 1. Overview of CTD² Dashboard datasets

High-throughput or bioinformatic method	# Tier 1 observations (May 2017)
Cancer cell-line sensitivity profiling (9)	10 828
Oncogenomic screening – average tumor volumes (18)	18
Master-regulator analyses (17,28)	38
shRNA screens and ATARIS (19,23)	6961
siRNA functional kinomics screen (21)	58
MethylMix analysis (25)	58
Elastic-net analysis of TCGA (27)	7176
Evaluation of dependency differentiality (EDDY) analysis (20)	164
Chemical-genetic interaction mapping (24)	156
Quantitative-trait locus variant identification (26)	19
Drug-sensitivity screening (22)	66
Reverse phase protein arrays (22)	24

Tier 1 observations associated with unique CTD² Dashboard ‘projects’ (9,17–28) are listed for each high-throughput assay or bioinformatics approach, as of May 2017.

harboring *LYN* mutations’. The italicized words in this example highlight the building blocks of an observation, namely the subjects, and are hyperlinked in the Dashboard interface. The bold text ‘sensitive’ summarizes a piece of evidence that defines the result (enrichment among sensitive cell lines, in this case). Additionally, observations are linked to background information and other evidence, i.e., assay-specific data which corroborate the observation statement. For instance, for the example observation shown above, relevant evidence could include the false discovery rate, an enrichment plot and the number of cell lines used. These CTD² Dashboard principles are illustrated in more detail in subsequent sections.

Subjects–focus of the experiments

As one mission of the CTD² Network is to discover and develop new therapeutic hypotheses, we expected that cancer-relevant genes and the proteins they encode (e.g., oncoproteins) would need to be explicit in our model. In addition, perturbation reagents such as small molecules or shRNAs, and certain common experimental systems, such as named cell lines and mouse models, were also included. ‘Subjects’ are the focus of experiments and are categorized into classes, such as genes, proteins, perturbagens, and model systems. A critical first decision in conceptualizing the CTD² Dashboard was to decide which entities would be considered well-defined classes for the purposes of cross-linking different types of evidence. Together, these subject classes form the core knowledge skeleton of the

CTD² Dashboard, allowing researchers from multiple Centers to connect observations and other evidence to the same abstract entity, even when studied in separate laboratories. These subject names are made consistent through vocabulary control and synonym reconciliation, ideally *via* an external authority (e.g., gene names and synonyms from HUGO Gene Nomenclature Committee database). In some cases, such as for cell lines and mouse models, no suitable external authority was already available, so the CTD² Dashboard development team curated and reconciled synonyms for these subjects. This construction allows the knowledge built around a subject to be greater than the sum of the individual contributions.

Subject choices permit extensive connection to structured resources outside of the CTD² Dashboard for additional external data and metadata. Importantly, each subject class warrants connection to different types of metadata and resources. For example, genes and proteins are described with controlled names and identifiers, and the CTD² Dashboard links out to other resources, including Entrez and UniProt, which provide additional summary information, genomic context, functional annotations, macromolecule sequences, and references. Genes are also linked to their monograph pages in the cBioPortal (3,4), which provides frequencies of alteration of the gene in different cancer types, as well as mutation and expression-level information in collections of samples.

Similarly, small-molecule probes and drugs can connect to sources of outside information, as many public databases of such entities are available (5–7). At present, small-molecule monographs in the Dashboard link to their corresponding entries in PubChem (7), which provide summaries of activity and mechanism of action, where available, along with chemical information and descriptors, structure representations, and links to bioactivity assays using the compound. New versions of the Dashboard will add links to other resources, such as the BioAssay Research Database (8) and the Cancer Therapeutics Response Portal (9–11), through which additional information about compound activity, including in cancer cell models, can be accessed. By linking to external data about subjects whenever possible, we enrich the CTD² Dashboard with content beyond that generated explicitly by the CTD² Network.

Roles–functions assigned to subjects based on supporting data

Due to the diversity of experiments done by CTD² Network Centers and the inherent characteristic of subjects to have multiple possible biological functions, subject classes alone are not sufficient to provide all the contextual information needed to understand the relationships between subjects.

We introduced additional functional classifications for subjects, termed ‘roles’, which are assigned to each subject in a CTD² Dashboard submission to delineate how it is acting or being used in an experiment, and to help interpret the associated evidence. In general, roles are standardized and new terms are added to the list of potential roles only after careful consideration. This strategy prevents submitters from assigning very specific and unique roles to each use of a subject, which would render the use of roles as a grouping mechanism less effective. In the Dashboard, roles for genes and proteins include target, biomarker, (genetic) background, interactor, (candidate) master regulator, and oncogene. In the Dashboard interface, a focus on targets and biomarkers is selected by default, but users can add or remove roles in the Browse interface using a check-box list under the button ‘Select Roles’. For compounds and other perturbagens, the roles of perturbagen and (candidate) drug are default, with an additional role of control compound available. Finally, cell and tissue samples associated with a particular disease context have roles of disease, metastasis, and tissue (of origin). Roles thus allow for easier browsing and searching.

Evidence–data supporting an observation

While subjects must conform to their type-specific controlled vocabularies, submitters are allowed complete flexibility when providing evidence pertinent to an observation. This information can include numeric values or text displayed directly on observation pages, files (including images) that are served by the CTD² Dashboard’s web-server, or URL hyperlinks to other websites. Thus, CTD² Dashboard evidence is constrained and categorized not by its biological meaning or context, but rather by its electronic data type. Furthermore, to indicate *how* a particular piece of evidence was obtained, we categorize them as ‘observed’, ‘measured’, or ‘computed’. Again, these classifications are more general than any particular biological context or interpretation, and together these broad evidence types and categories encourage submitters to define evidence in new and creative ways. Each of the existing Dashboard evidence types and roles are mapped to their corresponding Evidence Ontology (<http://www.evidenceontology.org/>) codes and appropriate hyperlinks are provided to the corresponding Evidence Ontology pages. Furthermore, we encourage submitters to use Evidence Ontology terms in their submissions to provide further resolution of classes of evidence.

Data submissions and curation

Data submitted to the Dashboard by CTD² Network Centers can be accessed through the web interface [[\[ctd2-dashboard.nci.nih.gov/\]\(https://ctd2-dashboard.nci.nih.gov/\)\]. The primary goal of the interface is to enable inter-connecting, browsing, searching, and presentation of the CTD² Dashboard data through a front-end that is intuitive to use and accessible to scientists at all levels of bioinformatics expertise. A key element that enables this functionality is the consistent use of controlled subject classes across all Center submissions. These serve as unambiguous identifiers that facilitate cross-referencing of data from different Centers to identify observations referring to the same abstract biological entities.](https://</p></div><div data-bbox=)

The CTD² Dashboard is geared towards sharing observations with scientists and others not specializing in computational biomedicine (Figure 1). New content is contributed to the CTD² Dashboard through individual ‘submissions’ by the CTD² Network Centers. A submission is a collection of observations that were obtained from one investigation, possibly integrating more than one type of experiment, and whose results share the same structure and format. For functional biology experiments (Tiers 2 and 3), submissions often consist of only a single observation. Currently the process of preparing, curating, and uploading submissions is manual. Submitters, with the help of the CTD² Dashboard team, identify fields (subjects, background information, and evidence) pertinent to observations, assign classes, roles, and types to the fields, and provide, in a tabular file format, descriptive text for each field displayed on the corresponding observation page. A submission must contain an ‘observation summary’ that describes each observation and is displayed on browse and search pages. These summaries use variables (i.e., subject classes that are ‘filled in’ by the submitting Center) and additional text to provide context for inter-linking subjects appropriately. Multiple observations of the same type (e.g., describing multiple results from a certain experiment or analysis), but about different subjects of the same class, can be handled by a single observation summary. Prior to loading to the CTD² Dashboard, submissions are automatically checked for internal consistency and for consistency with current CTD² Dashboard content. Recently, the CTD² Dashboard team created a web-based submission-generation system to guide contributors through the process. To manage quality control, at this time only CTD² Network Centers can contribute new submissions to the Dashboard. However, we envision that this system may open the door for future Dashboard submissions from outside the CTD² Network. To keep users up-to-date on the latest additions to the Dashboard, an RSS feed feature is provided, which allows subscription to receive a notification of submissions or observations involving a subject of interest (e.g., gene, protein, or compound).

Contextual architecture is independent of subject matter

The fundamental structure of the CTD² Dashboard architecture is not coupled to cancer targets, cancer biology, or even biology. Indeed, any system of controlled subject names could serve as a knowledge skeleton for attaching evidence. We have not fully explored alternative knowledge domains beyond the conceptual stage, but certainly encourage others to do so. Controlled-name subjects with any entity-relational framework could be coupled with evidence classes used in the current CTD² Dashboard: labels, numeric values, URLs, images, and attached files. In this sense, the CTD² Dashboard serves as a way to link an ‘operating system’ of file types with each member subject of a given class. In other words, the Dashboard does not need to understand the contents of the evidence being linked, only the format in which that evidence is being presented.

User interface

The end-user-facing web application is constructed in two layers. The back-end is implemented as a series of RESTful web services based on the Spring framework, a widely used application framework and inversion-of-control container for the Java platform [https://spring.io/guides/gs/rest-service/]. The front-end is developed using the lightweight backbone.js JavaScript library [http://backbonejs.org/docs/backbone.html]. The design provides a loose and flexible coupling between the two layers, allowing for each part to be developed and maintained independently from the other. The database supporting the CTD² Dashboard web interface is built using the Hibernate implementation of the Java Persistence API, which is seamlessly supported by the Spring framework. The back-end RDBMS used is MySQL. The CTD² Dashboard source code is open and publicly available on GitHub [https://github.com/CBIIT/nci-ctd2-dashboard/].

Researchers who visit the CTD² Dashboard to investigate a specific gene or compound will find the Search function to be the most expedient method for locating relevant data. CTD² Dashboard data are indexed against subject terms and can be queried using the search box located in the top menu bar of the CTD² Dashboard. For instance, searching for a full or partial gene symbol will generate a list of all matching genes along with a count of the observations involving each gene; each count is hyperlinked to a detailed list of the relevant observations.

In addition to search, data can be explored *via* several browsing options.

- *Subject-based browsing*: Exploration of CTD² Dashboard entries can be initiated from a summary

view. The CTD² Dashboard offers three “Browse” pages focusing on genes, compounds, and cancer types. Each page provides a list of biological entities prioritized according to the number, breadth, and level of validation of the observations involving those entities. For example, the ‘Biomarkers, Targets, Genes & Proteins’ browse page focuses on genes and lists roles and observations associated with each subject (Figure 2). Users can download this table as an Excel file using the function called “Export as Spreadsheet”.

- *Submission-based browsing*: The ‘Centers’ link at the site’s top navigation bar leads to a page that lists all the CTD² Network Centers. From that list, users can access all data submissions produced by a Center and drill down to the observation data in each submission. There is also a link which allows the user to easily and directly download the entire submission source data package.
- *Story-based browsing*: Research findings that have been explored in detail are featured prominently on the homepage of the CTD² Dashboard in the form of stories. The aim of stories is to distill the essential elements of reported results, including integration of many experiments at different levels of evidence, at a level of detail that is appropriate for a general scientific audience.

Search and browse operations ultimately result in a list of entry points to detailed information about observations and subjects. For instance, clicking on any gene symbol under the ‘Name’ column in the gene-browser list leads to a gene profile page (Figure 3) where the top portion contains key gene descriptors and links to relevant annotations and resources, and the bottom portion lists the observations generated by the Centers involving that gene. As with the browse feature, users can download this entire table as an Excel file using the button ‘Export to Spreadsheet’. Similarly, clicking on the ‘details’ link of any observation brings up a page (Figure 4) describing the observation attributes, including:

- a short summary of the related investigational finding (i.e., the *observation summary*);
- the biological entities (genes, compounds, cell lines, etc.) involved in the study and their specific roles;
- description of the submission, submission date, and a link to download all the data (evidence) associated with that submission;
- experimentally and computationally derived evidence supporting the reported finding.

Gene Cart

Many CTD² Dashboard pages contain green ‘+’ icons adjacent to gene symbols (see Figures 3 and 4). During

CTD² Dashboard Centers Resources Browse Gene Cart e.g. CTNNB1 or ABT-737 Search

Explore: Biomarkers, Targets

Entries listed below are ordered by relevance, based on the number of Centers providing observations and Tier evidence level. For each Center, only its highest Tier of evidence counts toward the score. Entries must have Tier 2 evidence or higher to be in the list. Currently displaying observations involving Biomarkers, Targets, Genes & Proteins (genes) that have been assigned one of the following roles: Biomarker, Target (see [background information](#) for the meaning of observations, roles, and Tiers).

Reset initial ordering Select Roles

Showing 1 to 10 of 519 entries Export as Spreadsheet Filter Table:

Observations					
Class	Name	Role	Tier 3	Tier 2	Tier 1
gene	SMARCA4	biomarker	1	33 (1 center)	13 (2 centers)
gene	TP53	biomarker	2 (2 centers)		23 (2 centers)
gene	CTNNB1	biomarker	1	51 (2 centers)	29 (2 centers)
gene	KRAS	biomarker	1	6 (1 center)	10 (2 centers)
gene	AURKA	target	2 (2 centers)		4 (1 center)
gene	PIK3CA	biomarker	1		36 (2 centers)
gene	WEE1	target	1	1	9 (1 center)
gene	AKT1	target	4 (1 center)	5 (2 centers)	1
gene	EGFR	target	1	1	
gene	TFRC	biomarker	1		3 (1 center)

Show 10 entries Previous 1 2 3 4 5 Next

Figure 2. ‘Biomarkers, Targets, Genes & Proteins’ browse page. Every table row corresponds to a gene–role pair, and displays for each evidence Tier the number of observations where the gene has been reported to play the specific role. The number of Centers who have contributed towards the observations is also shown. The ‘Select Roles’ button can be used to customize this view by adding or removing roles beyond those used in the default listing (‘biomarker’ and ‘target’). The initial ordering of the table rows is determined by a relevance score calculated as the sum across Centers of the highest Tier evidence provided by each Center contributing evidence.

CTD² Dashboard Centers Resources Browse Gene Cart PIK3CB Search

PIK3CB

Gene symbol: PIK3CB +

Synonyms/Related terms: PI3K, P110BETA, PIK3C1, PI3KBETA

Organism: Homo sapiens

References: Entrez: 5291, UniProt: P42338

Genomic alterations: view in cBioPortal

Related observations for the role of target

Showing 1 to 2 of 2 entries Export as Spreadsheet Filter Table:

Date	Observation Summary	Tier	Center
Aug 3, 2016	A pooled in vivo tumorigenic screen identifies rare KRAS, NFE2L2, PIK3CB, and POT1 transforming alleles, and a parallel gene-expression correlation analysis indicates the functional roles of these rare alleles. (details »)	Tier 3	DANA-FARBER CANCER INSTITUTE
Sep 25, 2014	RNAi-mediated knockdown of PIK3CB kinase with siRNAs SIO2622214, SIO2622221, and SIO0085862 significantly reduced cell viability in head and neck squamous cell carcinoma cell line UMSCC-38, which harbors TP53 mutation p.K132N. (details »)	Tier 1	FRED HUTCH CURE START HERE

Show 10 entries Previous 1 Next

Subscribe to the 'PIK3CB' RSS feed to receive alerts when new observations are posted to the site.

Figure 3. Example gene-profile page. Gene-profile page for *PIK3CB* shows synonyms (as defined by Entrez), provides links to external annotation web sites, and lists observations involving *PIK3CB* along with their Tiers and the submitting Center names. Similar detail pages are available for all subject types supported by the CTD² Dashboard.

CTD² Dashboard Centers Resources Browse Gene Cart **PIK3CB** Search

Observation (Tier 3)

A pooled *in vivo* tumorigenic screen identifies rare **KRAS**, **NFE2L2**, **PIK3CB**, and **POT1** transforming alleles, and a parallel gene-expression correlation analysis indicates the functional roles of these rare alleles.

Name	Class	Role	Description
KRAS +	Gene	target	gene with transforming rare allelic variant
NFE2L2 +	Gene	target	gene with transforming rare allelic variant
PIK3CB +	Gene	target	gene with transforming rare allelic variant
POT1 +	Gene	target	gene with transforming rare allelic variant

Submission (hide details)

Project	Mapping the function of rare oncogenic variants
Description	Discovery of rare oncogenic variants through high-throughput screening and gene-expression analysis (details →)
Summary	Story about the identification and validation of rare tumorigenic alleles.
Date	Aug 3, 2016
Source Data	download

Evidence

Showing 1 to 3 of 3 entries Filter Table:

Type	Description	Details
reference	Kim, E. et al. Systematic Functional Interrogation of Rare Cancer Variants Identifies Oncogenic Alleles. <i>Cancer Discovery</i> . 2016 July 1; 6(7):714-26.	(open link)
resource	link to source data in CTD ² Data Portal	(open link)
written	Story about the identification and validation of rare tumorigenic alleles	(view)

Show 25 entries ← Previous 1 Next →

Figure 4. Example observation page. This page comprises an observation summary, a list of implicated subjects, a download link to information about the data submission containing the observation, and the evidence supporting the observation findings.

Dashboard data exploration, users can click on these icons to incrementally compile a list of genes of interest. This list can be accessed through the ‘Gene Cart’ link available in CTD² Dashboard’s navigation menu (Figure 5a). From there, users are able to edit the contents of the list and to query the Cellular Networks Knowledge Base (CNKB) (12) repository of molecular interaction networks to retrieve interactions involving the genes in the Gene Cart (Figure 5b and c). CNKB (12) is a database of protein and gene interactions maintained at Columbia University. It comprises curated pathway collections and cancer tissue-specific regulatory and signaling networks derived computationally through the analysis of large gene-expression datasets (13) such as those generated by The Cancer Genome Atlas, and also predicted or measured protein-protein interactions (14,15). The ability to interrogate CNKB and to explore the interactions involving key tumor regulators can help users better contextualize and interpret Dashboard observations. For instance, if an shRNA screen

reduces the viability of a tumor cell line, looking at the interaction network involving the shRNA targets may reveal the underlying mechanism of action by identifying genes regulated by these targets.

Discussion

As content continues to grow and submitters continue to find creative uses for the Dashboard as a data-sharing platform, the CTD² Dashboard team will continue to improve the database and interface (Figure 1). Our current major undertaking to improve CTD² Dashboard functionality is creating a web-based submission-generation system that uses standardized formats, thereby reducing the time and complexity of the submission process by automating data import and validation. Plans to improve the user experience and further enrich the content include linking to additional outside resources that the Network selects as useful. For example, on the gene-entry pages, hyperlinks will point users

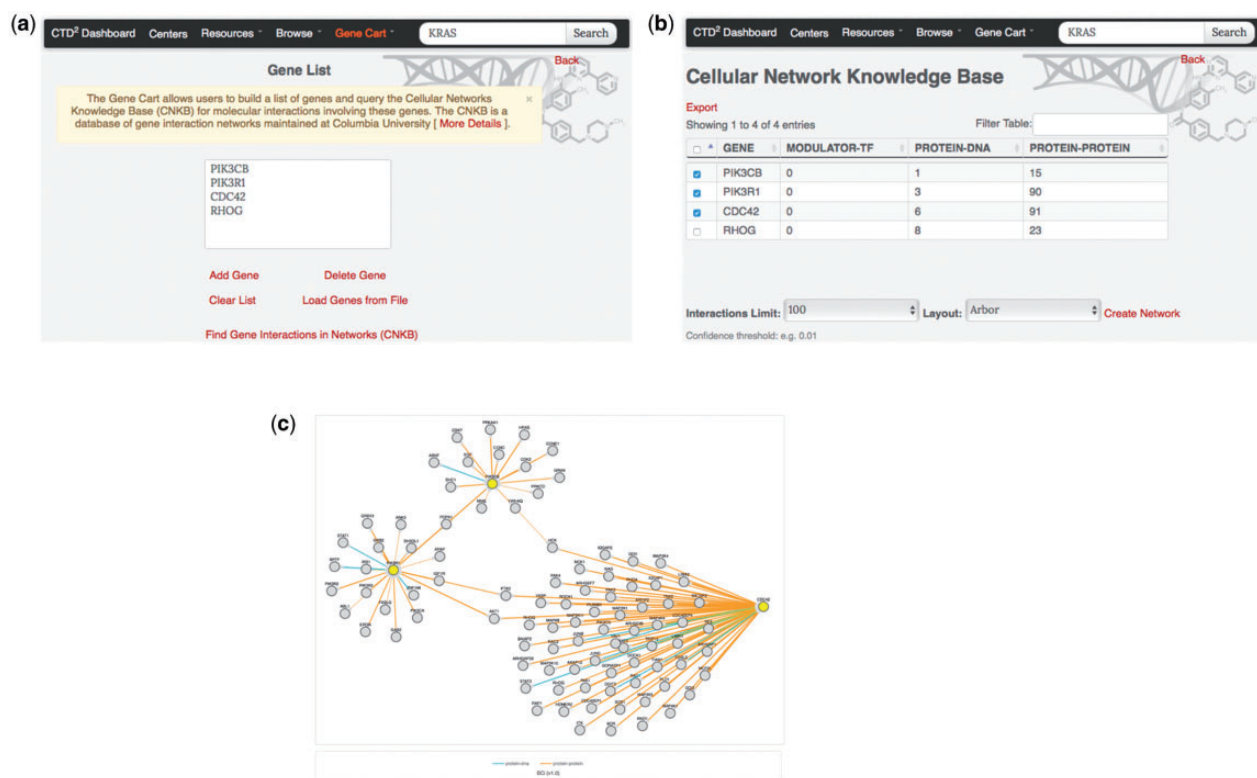


Figure 5. The CTD² Dashboard Gene Cart feature. (a) The ‘Gene Cart’ link available on the top navigation bar leads to a page that lists all genes that have been added by the user. From this page, users can add or remove genes from the list and initiate a query of the Cellular Networks Knowledge Base. (b) After selecting interactome type and version (in this example, B-Cell Interactome v1.0 was chosen) query results are summarized in a tabular view, with one row per query gene. Each row shows the number of interactions involving the corresponding gene, categorized according to interaction type (protein–DNA or protein–protein). Results can be exported as a SIF format file for display in third-party software such as Cytoscape (29,30). (c) Users can also select one or more of the query genes and click on the ‘Create Network’ link to generate a graph view of the interactions. Nodes in yellow represent query genes. Edges represent interactions and are colored according to the type of interaction.

to GeneCards and HGNC. The ‘Compounds & Perturbagens’ browse pages will include links to ChemBank (6,16), BioAssay Research Database (8), and ZINC (5). Links to relevant cell lines at American Type Culture Collection (ATCC) will be added to the tissue-type entries. In keeping with the FAIR principles (1), further evolution of the Dashboard will continue to expand the use of cross-links with other community resources.

As CTD² Dashboard content grows, evidence to support higher-Tier rankings increases, and additional external resources are integrated, the opportunities for its use will expand. Already, many Dashboard submissions about the same subjects are interconnected across multiple Tiers, reflecting the evolving validation within CTD² Network projects. The wealth of information embedded in the CTD² Dashboard can be taken advantage of to confirm hypotheses and initiate new collaborations. We plan to take advantage of this information to explore new bioinformatic analyses that may yield new cancer dependencies. We encourage the research community to use this database to study the experimental approaches the Network has taken, explore the data using an easily

minable format, develop new hypotheses, and build on CTD² Network findings.

Acknowledgement

The authors are indebted to the Investigators and members of the CTD² Network for helpful feedback during initiation, development, and beta-testing of the Dashboard, as well as their continued participation in providing CTD² Dashboard content. The authors would also like to thank NCI CBIIT Application Hosting, DBA, and SCM teams for their work in deploying and hosting the CTD² Dashboard, in particular Tracy Truong and Phil Hartman.

Funding

The CTD² Dashboard was supported by the National Cancer Institute’s Cancer Target Discovery and Development Network (Grant nos. U01CA168409, U01CA168426 and U01CA176152]. Additional work preparing the manuscript and figures was supported by U01CA176303 (O.N.). S.L.S. is an Investigator at the Howard Hughes Medical Institute. Funding for open access charge: National Cancer Institute.

Conflict of interest. None declared.

References

1. Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J. *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, 3, 160018.
2. Gerhard, D.S., Clemons, P.A., Shamji, A.F. *et al.* (2016) Transforming Big Data into cancer-relevant insight: an initial, multi-tier approach to assess reproducibility and relevance. *Mol Cancer Res.*, 14, 675–682.
3. Cerami, E., Gao, J., Dogrusoz, U. *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, 2, 401–404.
4. Gao, J., Aksoy, B.A., Dogrusoz, U. *et al.* (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.*, 6, p11.
5. Irwin, J.J. and Shoichet, B.K. (2005) ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.*, 45, 177–182.
6. Seiler, K.P., George, G.A., Happ, M.P. *et al.* (2008) ChemBank: a small-molecule screening and cheminformatics resource database. *Nucleic Acids Res.*, 36, D351–D359.
7. Wang, Y., Xiao, J., Suzek, T.O. *et al.* (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.*, 37, W623–W633.
8. Howe, E.A., de Souza, A., Lahr, D.L. *et al.* (2015) BioAssay Research Database (BARD): chemical biology and probe-development enabled by structured metadata and result types. *Nucleic Acids Res.*, 43, D1163–D1170.
9. Basu, A., Bodycombe, N.E., Cheah, J.H. *et al.* (2013) An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell*, 154, 1151–1161.
10. Seashore-Ludlow, B., Rees, M.G., Cheah, J.H. *et al.* (2015) Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discov.*, 5, 1210–1223.
11. de Waal, L., Lewis, T.A., Rees, M.G. *et al.* (2016) Identification of cancer-cytotoxic modulators of PDE3A by predictive chemogenomics. *Nat. Chem. Biol.*, 12, 102–108.
12. Mani, K.M., Lefebvre, C., Wang, K. *et al.* (2008) A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol. Syst. Biol.*, 4, 169.
13. Basso, K., Margolin, A.A., Stolovitzky, G. *et al.* (2005) Reverse engineering of regulatory networks in human B cells. *Nat. Genet.*, 37, 382–390.
14. Zhang, Q.C., Petrey, D., Deng, L. *et al.* (2012) Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature*, 490, 556–560.
15. Zhang, Q.C., Petrey, D., Garzon, J.I. *et al.* (2013) PrePPI: a structure-informed database of protein-protein interactions. *Nucleic Acids Res.*, 41, D828–D833.
16. Petri Seiler, K., Kuehn, H., Pat Happ, M. *et al.* (2008) Using ChemBank to probe chemical biology. *Curr. Protoc. Bioinformatics*, Chapter 14, Unit 14.15.
17. Carro, M.S., Lim, W.K., Alvarez, M.J. *et al.* (2010) The transcriptional network for mesenchymal transformation of brain tumours. *Nature*, 463, 318–325.
18. Sawey, E.T., Chanrion, M., Cai, C. *et al.* (2011) Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by Oncogenomic screening. *Cancer Cell*, 19, 347–358.
19. Shao, D.D., Tsherniak, A., Gopal, S. *et al.* (2013) ATARis: computational quantification of gene suppression phenotypes from multisample RNAi screens. *Genome Res.*, 23, 665–678.
20. Jung, S. and Kim, S. (2014) EDDY: a novel statistical gene set test method to detect differential genetic dependencies. *Nucleic Acids Res.*, 42, e60.
21. Moser, R., Xu, C., Kao, M. *et al.* (2014) Functional kinomics identifies candidate therapeutic targets in head and neck cancer. *Clin. Cancer Res.*, 20, 4274–4288.
22. Cheung, L.W., Yu, S., Zhang, D. *et al.* (2014) Naturally occurring neomorphic PIK3R1 mutations activate the MAPK pathway, dictating therapeutic response to MAPK pathway inhibitors. *Cancer Cell*, 26, 479–494.
23. Cowley, G.S., Weir, B.A., Vazquez, F. *et al.* (2014) Parallel genome-scale loss of function screens in 216 cancer cell lines for the identification of context-specific genetic dependencies. *Sci. Data*, 1, 140035.
24. Martins, M.M., Zhou, A.Y., Corella, A. *et al.* (2015) Linking tumor mutations to drug responses via a quantitative chemical-genetic interaction map. *Cancer Discov.*, 5, 154–167.
25. Gevaert, O. (2015) MethylMix: an R package for identifying DNA methylation-driven genes. *Bioinformatics*, 31, 1839–1841.
26. Chen, J., Hackett, C.S., Zhang, S. *et al.* (2015) The genetics of splicing in neuroblastoma. *Cancer Discov.*, 5, 380–395.
27. Lee, H., Palm, J., Grimes, S.M. and Ji, H.P. (2015) The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical-genomic driver associations. *Genome Med.*, 7, 112.
28. Bisikirska, B., Bansal, M., Shen, Y. *et al.* (2016) Elucidation and pharmacological targeting of novel molecular drivers of follicular lymphoma progression. *Cancer Res.*, 76, 664–674.
29. Shannon, P., Markiel, A., Ozier, O. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13, 2498–2504.
30. Smoot, M.E., Ono, K., Ruscheinski, J. *et al.* (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, 27, 431–432.