
CTD² Data Sharing

Paul Clemons
Broad Institute



We have made CTD² data and methods publicly available

Diseases or	Conv	Small Molecule (SM)	Profiling	Contact
Home	Resources	Publications	Data	Contact
				Kenneth Smith
<p>Targeting Cancer Dependencies with Small Molecules</p> <p>Small-Molecule High-Throughput Screening Data</p> <p>Click to navigate to the PubChem BioAssay page.</p> <p>Small-molecule microarray projects:</p> <ul style="list-style-type: none"> Small-molecule microarray assay for binding transcription factor STAT3 Small-molecule microarray assay for binding transcription factor NF-kappa B1 p50 subunit Small-molecule microarray assay for binding mutant tumor suppressor p53 Small-molecule microarray assay for binding helix-loop-helix transcription factor ID4 <p>High-throughput screening projects:</p> <ul style="list-style-type: none"> Plate-reader assay for inhibition of isocitrate dehydrogenase IDH1 mutant R132H. <p>Small-Molecule Cancer Cell-Line Profiling Data</p> <p>Left click to view, right click to download.</p> <p>Broad.CTD2_AUCData.txt Small-molecule profiling of genetically characterized cancer cell lines</p> <p>Broad.CTD2_ranData.txt Small-molecule profiling of genetically characterized cancer cell lines</p> <p>All CTD² data can be found here.</p> <p>Data bulk-download instructions.</p>				Barbara Weir
<p>(CSHL)</p> <p>Cancer Dependencies: Targets and Cell Lines (Broad)</p>				Chun Bu
<p>Data Files</p>				Jinyu Li
<p>SM Profiling</p> <p>SM Screening</p> <p>Data Files</p>				Paul Clemons Aly Shamji

2013-03-28 http://ctd2.nci.nih.gov/DataMatrix/CTD2_DataMatrix.html 2



Many web environments contain both data and tools

Galaxy: Data intensive biology for everyone. Galaxy is an open, web-based platform for performing bioinformatics analyses across a wide variety of data types, ranging from raw sequencing reads to complex integrative analyses. Whether on the fly, public server, or your own private infrastructure, Galaxy allows you to reproduce, and share complete analyses.

GenePattern: A powerful genomics tool for gene expression analysis, providing common data processing tasks. It allows the creation of multi-step workflows.

Office of Cancer Genomics: CTD² Data Matrix. Directions for navigating the Data Matrix: Each name listed within the data matrix is linked to more information. To learn more about: **Projects** - click on a specific name in the Diseases or Cell Lines column. **Centers** - click on the name of the center located directly beneath its corresponding project in the Disease or Cell Lines column. **Experimental methods** - click on the name of the method in the middle column. **Data** - click on "Data File" in the middle column to download the data. **Questions/Comments** - send an email by clicking on the Contact name.

Cancer Genomics Portal: Provides analysis, datasets, and more.

2013-03-28 5

CTD² Network Centers work with NCI on data sharing

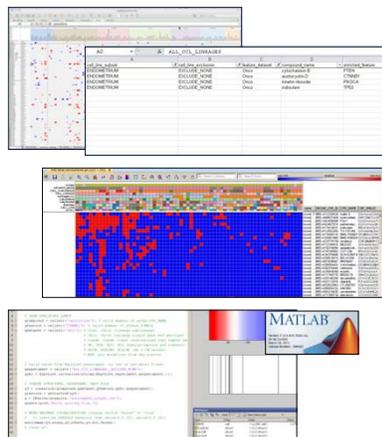
Broad Institute: Targeting Cancer Dependencies with Small Molecules. Small-Molecule High-Throughput Screening Data. Click to navigate to the PubChem BioAssay page. Small-molecule microarray projects: Small-molecule microarray assay for binding transcription factor STAT3. Small-molecule microarray assay for binding transcription factor NF-kappa B1 p50 subunit. Small-molecule microarray assay for binding nuclear tumor suppressor p53. Small-molecule microarray assay for binding helix-loop-helix transcription factor ID4. High-throughput screening projects: Plate-reader assay for inhibition of isocitrate dehydrogenase IDH1 mutant R132H. Small-Molecule Cancer Cell-Line Profiling Data. Left click to view, right click to download. Broad.CTD2_AUCData.txt Small-molecule profiling of genetically characterized cancer cell lines. Broad.CTD2_nwData.txt Small-molecule profiling of genetically characterized cancer cell lines. All CTD² data can be found here. Data bulk-download instructions.

evolving NCI 'Data Matrix' Portal: what can we do quickly to enhance the usability of CTD² Data Matrix?

NCI: Small-molecule profiling of genetically characterized cancer cell lines from temozolomide/methotrexate treatment of cultured cells in SW620 cell line.

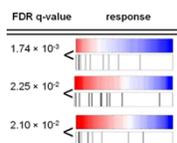
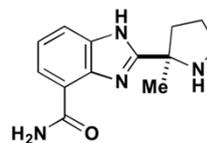
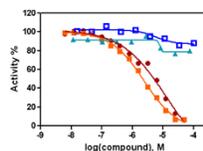
2013-03-28 6

CTD² Network Centers work with NCI on data sharing



**new datasets, pre-computed
result sets, and code**

FUTURE PLANS for NCI Portal



**new visualizations for web-
based result display**

2013-03-28

7 BROAD INSTITUTE

CTD² Dashboard aims to present stories from the Network

The screenshot shows the CTD² Dashboard interface. The top navigation bar includes links for CTD², Dashboard, Centers, Resources, Publications, Data Matrix, and About. The main content area is titled 'Stories' and features a large text block with a sample target and drug, followed by a 'read more' link. Below the main text are four columns: Stories, Targets, Compounds, and Search, each with a brief description and a 'read more' link.

BROAD INSTITUTE

CTD² Dashboard aims to present stories from the Network

Hypotheses from the CTD² : Candidate Targets and Biomarkers

List last updated: 2012-MM-DD

BCR-ABL (Entrez) translocators	CML	target YY inhibitors active in cells (link to Portal dataset)	project description page	compoundXX active in humans (PubMedID)		
target X (Entrez)	ovarian cancer		project description page	Inducible shRNA active in mice (PubMed ID)	links to other PMIDs	
eIF-4A		eIF-4A as biomarker for mTOR therapy	project description page	compoundXX active in cell lines (link to Portal dataset and CCLE)	links to other PMIDs	Dr. X (email)

etc.

different levels of validation evidence will be defined by the Network



CTD² Dashboard aims to present stories from the Network

Target(s):
 CEBPB [link to NCBI](#), [link to GeneCards](#)
 CEBPD [link to NCBI](#), [link to GeneCards](#)
 STAT3 [link to NCBI](#), [link to GeneCards](#)
(multiple gene target - from homepage)

Tumor Context: GBM glioblastoma multiforme
Genomic context: SNB19
(links out to genomic context for relevant system)

CTD² Network Stories (may be multiple, one example shown)

Co-ectopic expression of CEBPD and STAT3 [reprograms neural stem cells](#) along an aberrant mesenchymal lineage. Conversely, co-silencing of CEBPB and STAT3 in mesenchymal GBM cells induces loss of [mesenchymal markers](#) and [abrogates tumorigenesis in vivo](#).

provided by investigator (edited by editor)

Publications:
 Carro, M et al., Nature 2010
 Chen, J et al. in review

network context

(Click for interactive - widget e.g., Cytoscape)

CTD² Network Analyses and Datasets related to this compound

Tier I Data:
 Discovery using Master Regulator Analysis. ✓

Tier II Data:
 Validation in multiple Mesenchymal GBM lines ✓

Tier III Data:
 In vivo validation in mouse xenografts (links out to the Portal Plus or Tier Detail views) ✓

Target Validated ✓ Not validated ✗

Small Molecules: 5-FU, Etoposide, ... (links to other Dashboard pages)

Biomarkers: CEBPB, pSTAT3, ...



CTD² Dashboard aims to present stories from the Network

The screenshot shows a web browser displaying the CTD² Dashboard for a specific small molecule. At the top left is the chemical structure of a small molecule. To its right, the "small molecule name" is listed along with synonyms and links to external databases like DrugBank, PubChem, and ChemBank. Below this, "known protein target(s)" are listed with links to HCB1. The main section, "CTD² Network Stories", features a grid of story cards, including one about "Tier 2/3 evidence" and another about "XX mutations in gene ZZ". A graph titled "Figure Legend (popup)" is also visible. Below the stories are sections for "CTD² Network Analyses and Datasets related to this compound" and "Other resources related to this compound", each with several links to external resources.



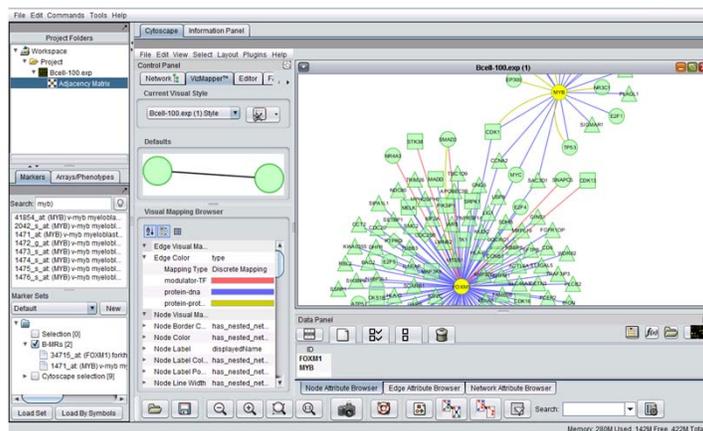
CTD² Dashboard leverages capabilities of participating groups

The screenshot shows the Pipeline Pilot Web Port interface. On the left is a "Protocols/Web Services" tree with a protocol named "Compound Names and Annotations" highlighted. A red circle with the number "1" is next to this protocol. An arrow points from this protocol to a central window titled "Compound Names and Annotations". This window contains a text box with the instruction "This protocol finds names and annotations for a compound specified by BROAD_ID" and a "BROAD_ID" input field. A red circle with the number "2" is next to the input field. Below the input field is a "Submit" button, with a red circle and the number "3" next to it. To the right of the input field is a chemical structure of a small molecule. Below the structure is a table titled "Names and Etc." and another table titled "Annotations".

database housing small-molecule names/synonyms and target annotations



CTD² Dashboard leverages capabilities of participating groups

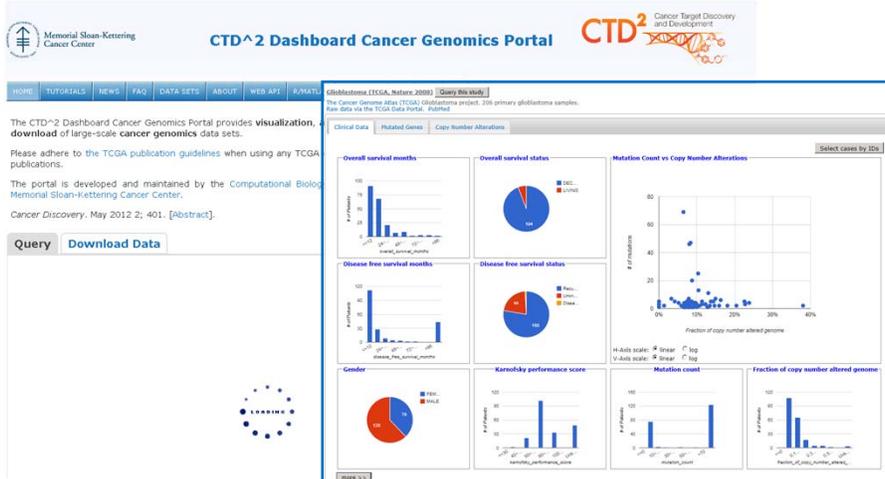


workbench for network biology and visualizations

2013-03-28

15

CTD² Dashboard leverages capabilities of participating groups

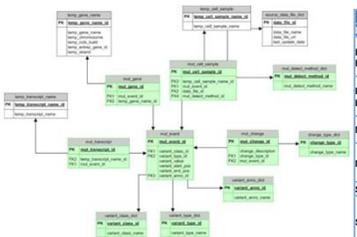


integrated portal for cancer genomics analyses

2013-03-28

16

CTD² Dashboard leverages capabilities of participating groups



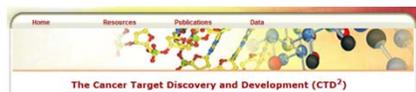
variant_classification			
NCI	CCLL	COSMIC (404864)	CDDB
Frame_Shift_Del Frame_Shift_Ins	Frame_Shift_Del Frame_Shift_Ins	Frameshift Complex (73) Unknown (347)	Frame_Shift
In_Frame_Del In_Frame_Ins	In_Frame_Del In_Frame_Ins	deletion inframe in frame insertion inframe	In_Frame
Missense_Mutation	Missense_Mutation	Missense compound substitution (137)	Missense
Nonsense_Mutation	Nonsense_Mutation	Nonsense compound substitution (18) Unknown (2)	Nonsense
Silent	Silent	coding silent	Silent
Splice_Site	Splice_Site_SNP Splice_Site_DNP Splice_Site_Ins Splice_Site_Del		Splice_Site
Translation_Start_Site	Start_Codon_Del		Translation_Start_Site
Nonstop_Mutation	Nonstop_Mutation	Nonstop extension	Nonstop
5'UTR	5'UTR		5'UTR
3'UTR	3'UTR		3'UTR
5'Flank	5'Flank		5'Flank
Intron	Intron		Intron
De_novo_Start_InFrame	De_novo_Start_InFrame		De_novo_Start_InFrame
De_novo_Start_OutOfFrame	De_novo_Start_OutOfFrame		De_novo_Start_OutOfFrame
	Stop_Codon_DNP Stop_Codon_Ins		Stop_Codon
		Complex (611) Unknown (7675) No detectable mRNA/protein (28)	Unknown
		Whole gene deletion (50)	Whole_Gene_Deletion

variant_type	
COSMIC	CDDB
Substitution	SNP DNP TNP QNP
Insertion	INS
Deletion	DEL
Complex	Complex
Unknown (26535)	Unknown

relational schema and vocabularies for cancer cell mutations

2013-03-28

What does LINCS have to do with CTD²?



has constituted a data working group to discuss formats, metadata, contexts, and data-release policies for sharing

has constituted a data working group to discuss formats, metadata, contexts, and data-release policies for sharing

LINCS and CTD² data working data working groups can work together and learn from each other

2013-03-28