

StRAP: An Integrated Resource for Profiling High-Throughput Cancer Genomic Data from Stress Response Studies

Seth Johnson^{1,2}, Biju Issac³, Shuping Zhao^{1,2}, Mohit Bisht¹, Orieta Celiku¹, Philip Tofilon¹, Kevin Camphausen¹, Uma Shankavaram^{1*}

1 Radiation Oncology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, United States of America, **2** SAIC-Frederick, Inc., National Cancer Institute-Frederick, Frederick, Maryland, United States of America, **3** Division of Bioinformatics, Sylvester Comprehensive Cancer Center, University of Miami, Florida, United States of America

Abstract

The increasing availability and maturity of DNA microarray technology has led to an explosion of cancer profiling studies for identifying cancer biomarkers, and predicting treatment response. Uncovering complex relationships, however, remains the most challenging task as it requires compiling and efficiently querying data from various sources. Here, we describe the Stress Response Array Profiler (StRAP), an open-source, web-based resource for storage, profiling, visualization, and sharing of cancer genomic data. StRAP houses multi-cancer microarray data with major emphasis on radiotherapy studies, and takes a systems biology approach towards the integration, comparison, and cross-validation of multiple cancer profiling studies. The database is a comprehensive platform for comparative analysis of gene expression data. For effective use of arrays, we provide user-friendly and interactive visualization tools that can display the data and query results. StRAP is web-based, platform-independent, and freely accessible at <http://strap.nci.nih.gov/>.

Citation: Johnson S, Issac B, Zhao S, Bisht M, Celiku O, et al. (2012) StRAP: An Integrated Resource for Profiling High-Throughput Cancer Genomic Data from Stress Response Studies. *PLoS ONE* 7(12): e51693. doi:10.1371/journal.pone.0051693

Editor: Sui Huang, Institute for Systems Biology, United States of America

Received: August 9, 2012; **Accepted:** November 5, 2012; **Published:** December 17, 2012

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: This work was supported by Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: SJ and SZ are affiliated to SAIC, Frederick, Inc. Authors KC and PT are PLOS ONE Editorial Board members. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

* E-mail: uma@mail.nih.gov

Introduction

DNA microarrays are successfully being used to classify tumors and identify novel biomarkers associated with cancer (for some recent reviews see [1]). Genetic variants and differences in personal genomes not only impact cancer profiles but are often responsible for how the patient and the cancer respond to treatment. In particular, the response to cellular stress, whether induced by cytotoxic drugs, hypoxia, or ionizing radiation can vary greatly, and its genetic basis is subject of much interest. We are especially interested in elucidating the genetic basis of radiotherapy response in search of highly-predictive genetic signatures. Radiotherapy is a core component of cancer treatment [2] but has been relatively under-studied: a glimpse at public resources like Pubmed or array databases shows that radiotherapy studies constitute less than 1% of the total number of records.

Typically, each individual study involves a number of statistical and quantitative analysis steps (see [3] for a summary of typical steps), and can point to gene and gene products that are crucial for disease and treatment. However, the sparse, high-dimensional nature of the microarray data space [4], and the large number of genes involved in often subtle and complex pathways, necessitate meta analyses for comparing and aggregating results from different

studies. Cross-platform compatibility can only be achieved once within-platform consistency issues have been fully addressed and the results of such studies are as good as the gene identification method. MAQC consortium has generally found that proper sample preparation is sufficient to dramatically enhance multilab and multiplatform correlations [5]. The utility of such analyses was documented in the implementation of the CellMiner tool, a web based program for the integration of molecular profiling data at DNA, RNA, protein, and pharmacological levels on the widely studied NCI-60 cancer cells [6]. Several other studies found added complexity for meta analysis due to considerable diversity in source, sample, and platform types [7–9]. The two major technologies of microarrays differ in the basic design, cDNA microarrays use full-length transcripts printed onto the slides and oligonucleotide based arrays constitute a shorter- oligonucleotides synthesized in situ. A major design question is whether to measure the expression levels from each sample on a different microarray (using single-color, or single-channel, arrays), or instead to compare relative expression levels between a pair of samples on each microarray (two-color or two-channel arrays). There are tradeoffs between the two approaches. Single-color arrays allow for more flexibility in analysis, while two-color arrays can control for some technical issues by allowing a direct comparison in a

single hybridization [10]. A recent comparison of single- and two-color methods on the same platforms found good overall agreement in the data produced by the two methods [11]. The Z score transformation procedure for normalizing data is a familiar statistical method in both neuroimaging and psychological studies and recently been used in the meta analysis of microarray datasets from different platforms [12], and is especially suited for database development [13].

The wealth of data has also brought about the creation of a wide range of resources. On one end of the spectrum, data repositories like Gene Expression Omnibus (GEO) [14] provide access to raw experimental data; on the other end, tools like ONCOMINE [15] more ambitiously, but typically at a cost, provide facilities for meta analysis of array data. However, to our knowledge, none of the existing free resources focus on stress response or radiotherapy studies combined with visualization outputs.

We develop StRAP, a free web-accessible resource to address the need to query, compare, profile, and visualize results from different microarray experiments. StRAP hosts data from diverse cancer studies (currently from 12 different tissue types), and will be further extended in the future. We used Z scoring method to standardize data, since the internally normalized values do not change with subsequent addition of new datasets. All data are mapped to Entrez Gene identifiers for consistency in comparison. The user-friendly interface facilitates exploration by a wide-range of researchers, including those with little expertise in bioinformatics.

In the remainder of this paper we briefly describe StRAP's construction and core features.

Materials and Methods

Architecture

The runtime architecture of StRAP is described in Figure 1. The architecture is 3-tiered. The basic design of the architecture is an enhancement of our previously published CellMiner tool [6]. The bottom tier represents the sources of experimental (microarray), meta (cell line) data, and external tools that are invoked to visualize the data. The middle tier represents how the data are processed, stored, and made available to the user. The pre-processing steps were performed before deployment. At this stage, data from the lower tier were accessed, processed (using R scripting), and stored in the StRAP data repository (comprised of a MySQL database, and other files stored on the server file system). The right hand side of the middle tier represents the analysis "services" that are available at runtime to the user. These include filtering of data (according to user constructed queries), visualization of results, and the options to download the data. These services are made available as web-services and are hosted on an Apache server. The top tier represents the user interface (implemented using PHP, Javascript, AJAX, and HTML), and is organized around three main modules (Genes, Cell lines, and Arrays).

Data Repositories

Four main data repositories reside at the backend of StRAP: (1) Gene associated annotation information derived from the National Center for Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>), (2) Pre-processed gene expression microarray molecular profile data (including pre-computed statistics), (3) Metadata on cell lines, and (4) Metadata on platform-associated information.

The structured layout of the tables promotes efficient querying and integration of phenotypic data, metadata and molecular

profile information from various studies. The database supports multiple concurrent query sessions.

The repositories are stored as a MySQL relational database (<http://www.mysql.com>).

Data Preparation

The microarray data were obtained as raw files whenever available or else as author deposited normalized files from the GEO database [14], ArrayExpress [16], or in-house experiments. Two platform types are predominantly used in these studies: cDNA two-color (National Cancer Institute- ROSP 8K Human Array and Agilent whole human genome microarrays), and single color arrays (currently we house Affymetrix and Illumina gene chip data).

The raw data were assessed for quality and normalized by the Lowess [17], or MAS5 [18] methods for cDNA, and Affymetrix arrays, respectively. Z-score transformation was used to obtain a uniform scale across different studies and platforms, which is necessary for comparing data from different studies. Pre-computed statistical tests were performed at three nested-level complexity.

- At the top level, each study is subjected to ANOVA analysis performed between all controls and cases to give an overall significance of the study design.
- A tissue level ANOVA analysis is implemented as a second tier of comparison between all the controls and cases for each tissue type in a study.
- At the experiment level, for each cell-line/sample, a case-control comparison is performed by t-test analysis.

Pre-processing and computation of statistical tests are performed in the R environment (<http://www.r-project.org/>).

Interface

The front end interface is a web-based application implemented using R, PHP (<http://www.php.net/>) and Python (<http://www.python.org/>). The application is deployed on an Apache HTTP server (<http://httpd.apache.org/>) at the National Cancer Institute (NCI).

Core Features

Data access and presentation is organized around three main concepts or modules: (1) Genes, (2) Cell lines, and (3) Arrays. Flexible user-defined data queries can be initiated from any of the modules; the data visualization options for the results are displayed in integrated views and may, depending on the query, involve cross-talk between modules. Several links to external resources promote a systems biology approach. Table 1 gives a summary of core features for each module. Pre-computed statistics (as described in the previous section) enable display of efficient and intuitive graphs.

Genes

The genes module enables gene-centric queries of the StRAP microarray studies. Queries can be based on gene or protein identifiers, synonyms, gene descriptions, or chromosome location. The results include associated arrays and studies, and a compilation of gene-annotation information, spatial localization within the genome visualized in the UCSC Genome browser [19], and network neighborhood maps generated from protein-protein interaction networks [20]. Queries can also be constructed using gene lists defined by the user or generated, for example, from Gene Ontology (GO) terms [21].

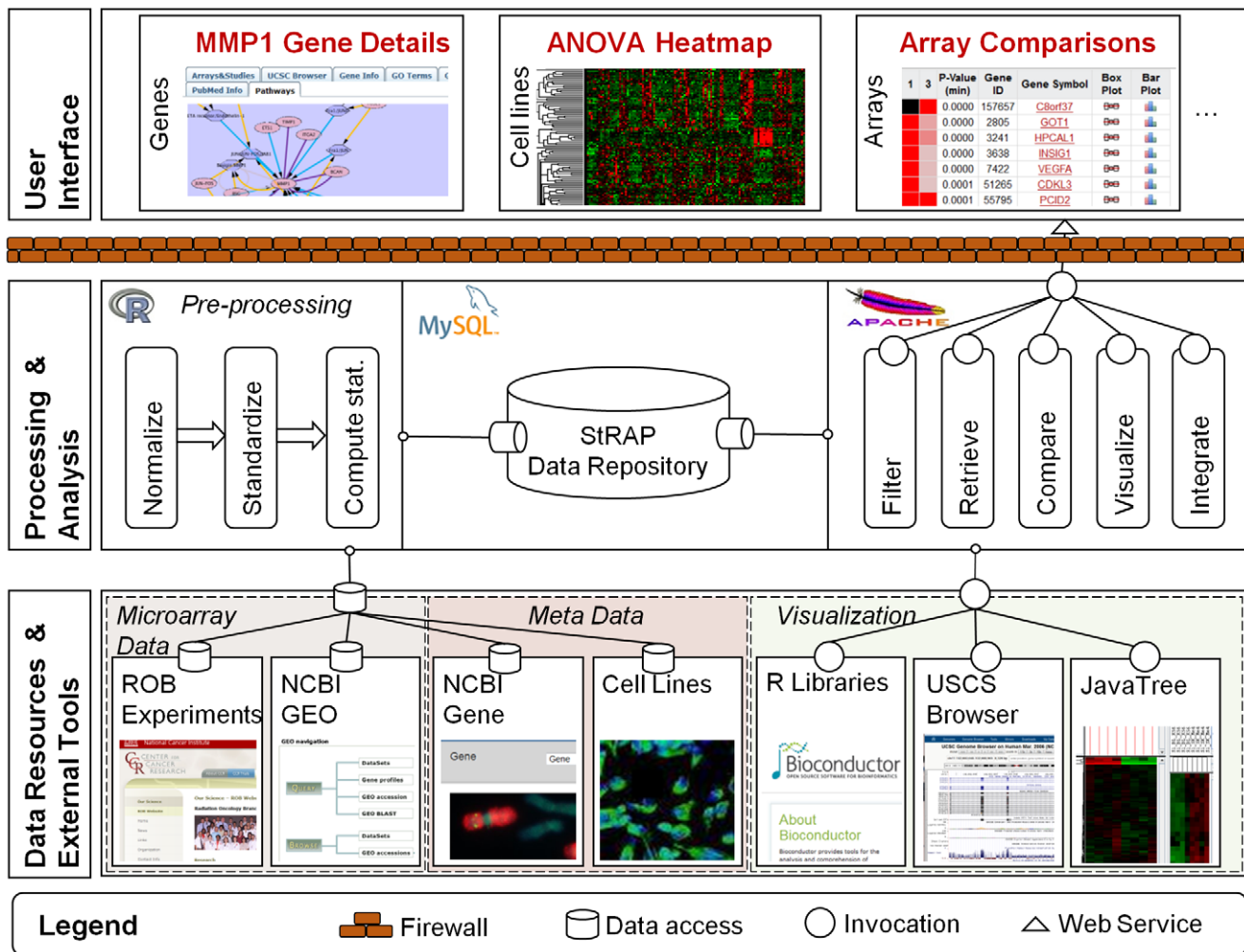


Figure 1. StRAP architecture diagram. The diagram represents a runtime view of the architecture of StRAP. The lower tier represents the sources of experimental data, meta data, and external tools that are invoked to visualize the data. The middle tier represents how the data are processed, stored, and made available to the user. The right hand side of the middle tier represents the analysis “services” that are available at runtime to the user. The higher tier represents the user interface, and is organized around three main modules (Genes, Cell lines, and Arrays). doi:10.1371/journal.pone.0051693.g001

A typical gene-centric query (see Figure 2 for an example workflow) starts by identifying studies profiling the expression of a gene (list) of interest. The expression profiles and their statistical significance are then visualized via boxplots, and barplots (showing study-level, and experimental-level case-control differences). If the input involves a list of genes, an interactive heatmap option enables viewing expressions of genes in selected studies. The heatmap is visualized using the Java Treeview program [22].

As an added convenience, the genes module includes a gene identifier conversion utility, which can be used to map from one type of gene identifier (for example, Entrez gene symbol) to another (for example, Entrez geneid).

Cell Lines

The cell lines module provides metadata on available cell lines and associated studies. Queries in this module are tailored to allow selection of complete studies, by tissue of origin, or individual cell line. Comparisons can be made for samples within a study or across studies. (See Figure 3 for an example workflow.) Differentially expressed genes in studies of interest are identified based on case-control t-test analyses (cell line selection) and ANOVA

analysis (studies with more than one group). The default filter is set to $p \leq 0.05$, but can be customized by the user.

Arrays

The arrays module provides an overview of the current contents of the database, including the number of studies, information on platforms, contributors, and available meta-information. Pre-processed data or data from the original source can be downloaded from this module. Integrated queries from this module allow performing comparison of studies by common samples or union of genes within the selected studies.

An example workflow is shown in Figure 4. Arrays can be filtered by the select stimulus used in the study. Given our interest in effects of ionizing radiation, most of the arrays in the repository have “radiation” as stimulus.

Validation

Radiation therapy is a core component of cancer treatment. However, radiation response often varies considerably among different patients [23]. Therefore, it is important to identify genes predictive of radiation response. Equally important is to validate

Table 1. StRAP modules functionality.

	Gene module	Cell line module	Array module
Description	Gene level comparison of one or more studies queried by gene related keyword search, identifier, or GO terms.	Gene identifier conversion utility	Overview of all studies, or studies by tissue. Enables comparison of studies by common cell lines, or union of genes. Enables downloading of data.
Results	List of associated studies	Gene information Mapping of gene identifiers of one type to another	Option to compare studies by cell line Option to compare studies by gene
Details of meta information	UCSC genome browser, Pubmed references, GO terms, and Pathway Commons networks	Table of mapped gene identifiers	Experimental details, study reference information, contributors, and data download
Visualization of multi-gene query. A multi-study comparison by Metamap, or a single or multi-study comparison by Heatmap	Yes	List of selected studies	Yes
Visualization of Single gene query. A single or multi-study comparison at study level by Boxplot, and at Experiment level by Barplot	Yes	Description of origin and source, experimental details, and source reference information	Yes

doi:10.1371/journal.pone.0051693.t001

Step 1: Enter gene of interest

Gene Search

Option:

Search by Keyword

Enter any keyword, such as Entrez ID Number: "14", or Symbol: "AAMP", or any part of a Description: "Cancer"

Search

Search by List of Terms

Step 1: Select Input Type

- Gene Symbol
- Entrez GeneID
- Gene Description
- Chromosome
- GO description
- GenBank Accession

Step 2: Input a List of Terms

Choose File No file chosen

OR

Search Reset

Show/Hide Conversion Tool

Step 2: Select studies of interest

ABL1 gene details

Option:

Arrays&Studies UCSC Browser Gene Info GO Terms GeneRIFs

PubMed Info Pathways

Studies featuring the gene



All	Study #	Name
<input checked="" type="checkbox"/>	1	K562-Me45_RadiationInducedTranscriptomeChanges; 1 gene
<input checked="" type="checkbox"/>	2	MCF7_RadiationInducedMarkersOfPrematureSenescence; 1 gene
<input type="checkbox"/>	3	H9hESCs_RadiationEffect; 1 gene
<input type="checkbox"/>	4	Lymphocytes_LateRadiationTherapyToxicity; 1 gene
<input type="checkbox"/>	5	THP1_RadiationEffectOnProcoagulantActivity; 1 gene
<input type="checkbox"/>	6	Lymphoblastoid_RadiationTherapyToxicity; 1 gene
<input type="checkbox"/>	7	SCC_RadioresistantTumorResponse; 1 gene
<input type="checkbox"/>	8	ALL_RadiationSensitivityAndResistance; 1 gene
<input type="checkbox"/>	9	SCC_TimeCourseRadioresistantTumorResponse; 1 gene
<input type="checkbox"/>	10	MCF10A_EMT-Initiation; 1 gene
<input type="checkbox"/>	11	MulticancerCells_RadiationEffectOnPolysomeRNA; 1 gene
<input type="checkbox"/>	12	Lymphocytes_ATM-regulationOfDNADamageResponse; 1 gene
<input type="checkbox"/>	13	GliomaCells_HypoxiaEffectOnTotalRNA; 1 gene
<input type="checkbox"/>	14	StemCells_HypoxiaEffectOnTotalRNA; 1 gene
<input type="checkbox"/>	15	U251-U87_ComparisonOfGrowthModelsUnderRadiation; 1 gene
<input type="checkbox"/>	16	NCI60_RadiationResponse; 1 gene
<input type="checkbox"/>	17	Lymphocytes_TranscriptionalResponseToUVAandIR; 1 gene
<input type="checkbox"/>	18	MNCs_SignaturesForRadiationBiodosimetry; 1 gene
<input type="checkbox"/>	19	LCL_LowDoseRadiationResponse; 1 gene
<input type="checkbox"/>	20	SCC61_STAT1-modulationOfWarburgEffectInTumorGrowthAndRadioresistance; 1 gene
<input type="checkbox"/>	21	SCC_RadiationInducedCell-deathSignalingPathwayActivation; 1 gene

Compare

Step 3: Select visualization of interest

Visualization Options

Instructions: Please click an image to choose visualization type

Barplot:  Boxplot: 

Show/Hide Help

Steps 4+: Display visualizations

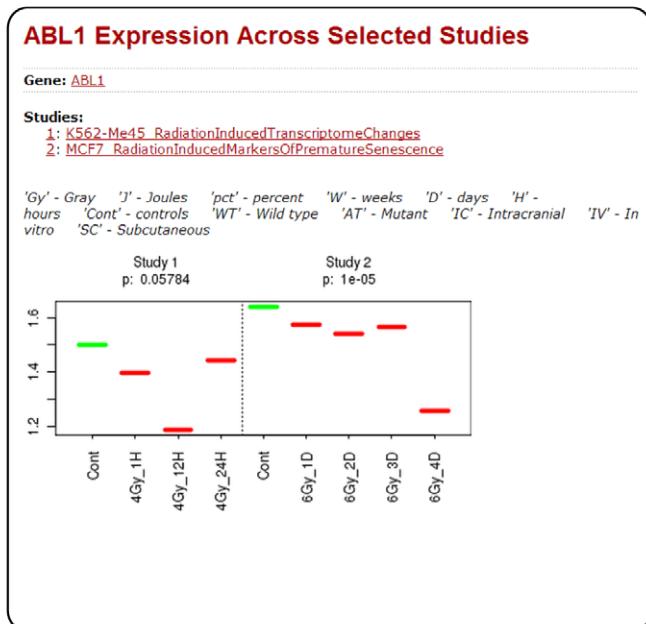
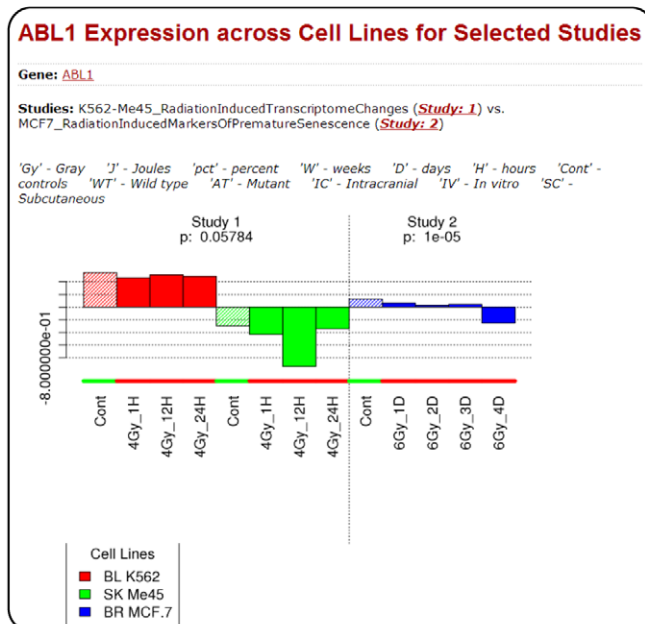


Figure 2. Example of a workflow initiated from the Genes module. Typically, a workflow initiated from the Genes module involves 1) entering a gene of interest (or list of genes), 2) displaying and selecting studies featuring the gene, 3) choosing a visualization option, and 4+) displaying and inspecting the chosen visualization. The example shown is for gene "ABL1". doi:10.1371/journal.pone.0051693.g002

the results of an analysis in independent data with similar experimental design.

To illustrate the functionality of StRAP, we used a study by Rieger and colleagues [24] on peripheral blood lymphoblastoid cells derived from patients with acute radiation toxicity and

control group of patients with mild toxicity. Using gene expression profiling, the authors reported 24 highly predictive genes of radiation response. We sought to explore the expression of these 24 genes in several independent studies from StRAP database, and found 18 genes significantly changed among the selected studies.

Step 1: Display available cell lines and select cell line of interest

Cell Lines

Options:

(For General Search enter any keyword, such as Cell Line Name: "MCF" or Tissue: "breast" or Histology: "gliosarcoma" or Institution: "Duke" or Contributor: "Lieber" or Reference: "Cell")

All Study #	Available Studies	Available Cell Lines
1	K562-Me45_RadiationInducedTranscriptomeChanges; 2 cell lines	786.Q A498 A549 AGHN ALL
2	MCF7_RadiationInducedMarkersOfPrematureSenescence; 1 cell line	AsPC1 BJ BT.549 BXP3 CAK1
3	H9hESCs_RadiationEffect; 1 cell line	CCRF_CEM COLC.205 DU145 EKVX GM02184
4	Lymphocytes_LateRadiationTherapyToxicity; 1 cell line	GM03332 H9.hESCs HCC.2998 HCT.116 HCT.15
5	THP1_RadiationEffectOnProcoagulantActivity; 1 cell line	HL60 HOP82 HOP92 HS578T HT29
6	Lymphoblastoid_RadiationTherapyToxicity; 2 cell lines	IGR.QV1 K562 KM12 LCL LOXIMVI
7	SCC_RadioresistantTumorResponse; 2 cell lines	M14 MALME.3M MCF.10A MCF.7 MDA.MB.231
8	ALL_RadiationSensitivityAndResistance; 1 cell line	MDA.MB.435 MDAN MDNSC11 MDNSC20 MDNSC23
9	SCC_TimeCourseRadioresistantTumorResponse; 2 cell lines	Me45 MEC MiaPaca.2 MNCs MOLT.4
10	MCF10A_EMT-Initiation; 1 cell line	MRC5 MRC9 NCI.H226 NCI.H23 NCI.H322M
11	MulticancerCells_RadiationEffectOnPolysomeRNA; 18 cell lines	NCI.H460 NCI.H522 NCIADR.RES Nu61 QVCAR.3
12	Lymphocytes_ATM-regulationOfDNADamageResponse; 2 cell lines	QVCAR.4 QVCAR.5 QVCAR.8 PANC.1 PC.3
13	GliomaCells_HypoxiaEffectOnTotalRNA; 2 cell lines	RPMI.8226 RFX.393 SCC61 SF126 SF268
14	StemCells_HypoxiaEffectOnTotalRNA; 3 cell lines	SF295 SF539 SK.MEL.2 SK.MEL.28 SK.MEL.5
15	U251-U87_ComparisonOfGrowthModelsUnderRadiation; 2 cell lines	SK.OV.3 SKCa SN12C SNB.19 SNB.75
16	NCI60_RadiationResponse; 60 cell lines	SR SW.620 T47D THP1 TK.10
17	Lymphocytes_TranscriptionalResponseToUVAandIR; 1 cell line	U118 U251 U373 U87 UACC.257
18	MNCs_SignaturesForRadiationBiodosimetry; 1 cell line	UACC.62 UO.31
19	LCL_LowDoseRadiationResponse; 1 cell line	
20	SCC61_STAT1-modulationOfWarburgEffectInTumorGrowthAnd Radioresistance; 1 cell line	
21	SCC_RadiationInducedCell-deathSignalingPathwayActivation; 1 cell line	

(Dark background indicates cell lines that are not associated with any microarrays)

Step 2: Select studies of interest

LCL Cell Line Details

Option:

Sample ID	74
Cell Line Name	LCL
Disease State	Multi: normal; prostate; PBMC
Patient's Age	21-26
Patient's Sex	
Prior Treatment	Cells immortalized with Epstein Barr Virus
Tissue Type	Blood
Histology	Normal; lymphocyte
p53 Status	
ECAD Methylation	
MDR Function	
Arrays Using This Cell Line	<input type="checkbox"/> Check All <input checked="" type="checkbox"/> Lymphocytes_LateRadiationTherapyToxicity (4) <input checked="" type="checkbox"/> Lymphoblastoid_RadiationTherapyToxicity (6) <input type="checkbox"/> Lymphocytes_TranscriptionalResponseToUVAandIR (17) <input type="checkbox"/> LCL_LowDoseRadiationResponse (19) <input type="button" value="Metamap"/>
Institution	Stanford University School of Medicine
Contributor	Rieger, KE and Chu, G
Reference	Nucleic Acids Res. 32(16):4786-803, 2003

Step 3: Display comparison visualization

Array Comparisons by Genes for LCL Cell Line

Option:

Search by Keyword:

Enter any keyword, such as Entrez ID Number: "14", or Symbol: "AAMP", or any part of a Description: "Cancer"

Study #	Gene ID	Gene Symbol	Box Plot	Bar Plot
Study #4: Lymphocytes_LateRadiationTherapyToxicity	LCL			
Study #6: Lymphoblastoid_RadiationTherapyToxicity	4 6 (min)			
	0.0000	64782	AEN	
	0.0000	307	ANXA4	
	0.0000	23204	ARL6IP1	
	0.0000	10973	ASCC3	
	0.0000	467	ATE3	
	0.0000	468	ATF4	
	0.0000	8704	B4GALT2	
	0.0000	581	BAX	
	0.0000	638	BIK	
	0.0000	9577	BRE	
	0.0000	57103	C12orf5	
	0.0000	10241	CALCOCO2	
	0.0000	865	CBER	
	0.0000	60492	CCDC90B	
	0.0000	900	CCNG1	
	0.0000	970	CD70	
	0.0000	1019	CDK4	
	0.0000	1026	CDKN1A	
	0.0000	1534	CYB5B1	
	0.0000	1643	DDB2	

Genes 1 to 20 of 12801

Steps 4+: Focus on individual genes and display associated visualizations

Array Comparisons by Genes for LCL Cell Line

Option:

Search by Keyword:

Enter any keyword, such as Entrez ID Number: "14", or Symbol: "AAMP", or any part of a Description: "Cancer"

Study #	Gene ID	Gene Symbol	Box Plot	Bar Plot
Study #4: Lymphocytes_LateRadiationTherapyToxicity	LCL			
Study #6: Lymphoblastoid_RadiationTherapyToxicity	4 6 (min)			
	0.7137	25	ABL1	

Genes 1 to 1 of 1

Figure 3. Example of a workflow initiated from the Cell lines module. The Cell lines initiated workflow typically starts with 1) selection of a cell line (or tissue) of interest (here "LCL"), 2) inspection of the cell line metadata, and associated studies, 3) comparison of studies of interest with a metamap showing significance of differential expression of individual genes for the given cell line, and 4+) inspecting individual genes via barplots and boxplots.

doi:10.1371/journal.pone.0051693.g003

Step 1: Display available array studies and select study of interest

Directory of Array Studies

Options:

Select All	Study #	Study Name	Stimulus	Cell Lines	Genes	Type	Institution	Contributor	Reference	Download
<input type="checkbox"/>	1	K562-Me45_RadiationInducedTranscriptomeChanges	Radiation	2	12509	Affymetrix	Silesian University of Technology	Jaksik, R	E-MEXP-2623	Data File
<input type="checkbox"/>	2	MCF7_RadiationInducedMarkersOfPrematureSenescence	Radiation	1	18081	illumina	Korea institute	Kim, BC et al.	20596626	Data File
<input type="checkbox"/>	3	H9hESCs_RadiationEffect	Radiation	1	16046	Agilent	Stanford	Wilson, KD et al.	20530673	Data File
<input type="checkbox"/>	4	Lymphocytes_LateRadiationTherapyToxicity	Radiation	1	12704	Affymetrix	Leiden University Medical Center	Svensson, JP et al.	17076557	Data File
<input type="checkbox"/>	5	THP1_RadiationEffectOnProcoagulantActivity	Radiation	1	19621	Affymetrix	Max Planck Institute	Goldin-Lang, P et al.	17640852	Data File
<input type="checkbox"/>	6	Lymphoblastoid_RadiationTherapyToxicity	Radiation	2	8038	Affymetrix	Stanford	Rieger, KE et al.	15096622	Data File
<input type="checkbox"/>	7	SCC_RadioresistantTumorResponse	Radiation	2	12704	Affymetrix	The University of Chicago	Khodarev, NN et al.	17909027	Data File
<input type="checkbox"/>	8	ALL_RadiationSensitivityAndResistance	Radiation	1	12704	Affymetrix	University of Birmingham	Stankovic, T et al.	18941120	Data File
<input type="checkbox"/>	9	SCC_TimeCourseRadioresistantTumorResponse	Radiation	2	12704	Affymetrix	The University of Chicago	Khodarev, NN et al.	17909027	Data File
<input type="checkbox"/>	10	MCF10A_EMT-initiation	Radiation	1	12704	Affymetrix	Lawrence Berkeley	Andarawewa, KL et al.	17875706	Data File
<input type="checkbox"/>	11	MulticancerCells_RadiationEffectOnPolysomeRNA	Radiation	18	5593	cDNA	Lee Moffitt Cancer Center	Tofilon, PJ	18483266	Data File
<input type="checkbox"/>	12	Lymphocytes_ATM-regulationOfDNADamageResponse	Radiation	2	11430	illumina	NIA	Mazan-Mamczarz, K et al.	21209379	Data File
<input type="checkbox"/>	13	GliomaCells_HypoxiaEffectOnTotalRNA	Hypoxia	2	13147	Affymetrix	Lee Moffitt Cancer Center	Tofilon, PJ	19372578	Data File
<input type="checkbox"/>	14	StemCells_HypoxiaEffectOnTotalRNA	Hypoxia	3	13147	Affymetrix	Lee Moffitt Cancer Center	Tofilon, PJ	19372578	Data File
<input type="checkbox"/>	15	U251-U87_ComparisonOfGrowthModelsUnderRadiation	Radiation	2	4283	cDNA	ROB	Camphausen, K et al.	15928080	Data File
<input type="checkbox"/>	16	NCI60_RadiationResponse	Radiation	60	2826	cDNA	Columbia University Medical Center	Amundson, S et al.	18199535	Data File
<input type="checkbox"/>	17	Lymphocytes_TranscriptionalResponseToUVAandIR	Radiation	1	8818	cDNA	Stanford	Rieger, KE & Chu, G	15356296	Data File

[Compare Studies by Cell Line](#) [Compare Studies by Gene](#)

Step 2: View experimental conditions and choose p-value

Experiments In Array Study: "K562-Me45_RadiationInducedTranscriptomeChanges"

Study: K562-Me45_RadiationInducedTranscriptomeChanges Stimulus: Radiation

Dose	Time	Replicates
Cont		2
4Gy	1H	2
4Gy	12H	2
4Gy	24H	2

(filtered by max p-value)

Step 3: View heatmap of genes with significant differential expression

Heatmap of (1410) Genes with p <= 0.05

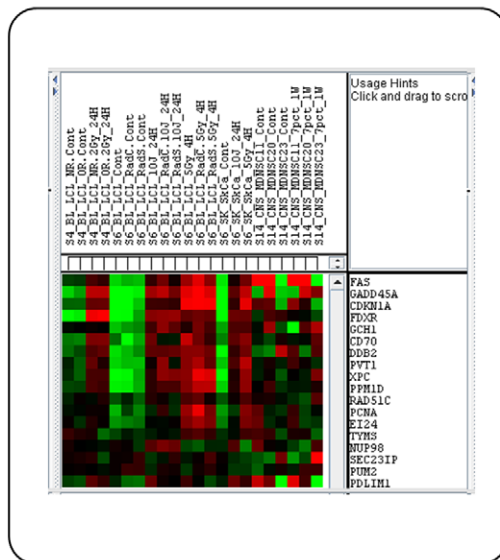
Options: Change p-value max & press [Enter]: 0.05

Steps 4+: Back to overview and compare several studies

Study #	ST	ST	ST	ST	P-Value (min)	Gene ID	Gene Symbol	Box Plot	Bar Plot
Study #1: K562-Me45_RadiationInducedTranscriptomeChanges	1	1	3	4	5				
Study #3: H9hESCs_RadiationEffect	CL	CL	CL	CL	CL				
Study #4: Lymphocytes_LateRadiationTherapyToxicity	30	87	88	74	78				
Study #5: THP1_RadiationEffectOnProcoagulantActivity									
						0.0000	64782	AEN	
						0.0000	307	ANXA4	
						0.0000	23204	ARL6IP1	
						0.0000	10973	ASCC3	
						0.0000	467	ATF3	
						0.0000	581	BAX	
						0.0000	638	BIK	
						0.0000	57103	C12orf5	
						0.0000	25980	C20orf4	
						0.0000	60492	CCDC90B	

Figure 4. Example of a workflow initiated from the Arrays module. The Arrays workflow typically starts with 1) inspection of available arrays and selection of a study of interest, 2) viewing of experimental conditions and selection of a p-value threshold for significance of gene expression differentiation, and 3) study of expressions heatmap. Comparison of several arrays can also be initiated from the overview page. doi:10.1371/journal.pone.0051693.g004

Step 1: View multi-study heatmap



Step 2: Inspect profiles of individual genes

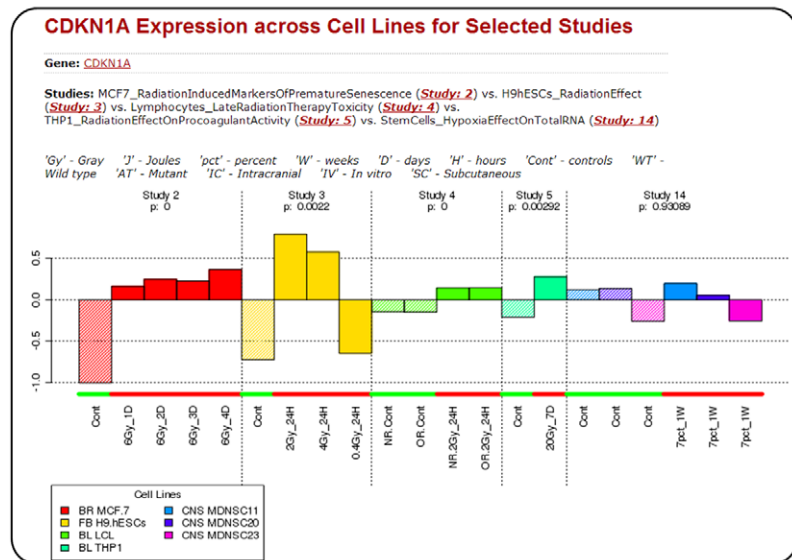


Figure 5. StRAP validation. Differential expression of twenty-four genes identified by Rieger and colleagues [24] to be important for radiation response. In Step 1 shows a multi-study heatmap (for Studies 4, 6, and 14). In Step 2 CDKN1A gene profile was compared in studies with (Studies 2–5), and without radiation as stimulus (Study 14 with hypoxia response). doi:10.1371/journal.pone.0051693.g005

To test if we can reproduce the authors findings, we first selected 3 studies, 2 studies (studies 4 and 6) containing lymphoblastoid cells treated with different doses of radiation, and as a negative control, we chose 1 study (Study 14) with stem cells from CNS tissue with hypoxia stimulus. A multi-study heatmap (Figure 5, Step 1) on the gene subset showed a selective up regulation of the gene subset in studies 4 and 6 but, not in study 14, confirming the role of these genes in response to radiation. Of particular, CDKN1A is a DNA damage response, cell cycle regulating gene reported to be induced by radiation [25,26]. We explored the comparative profiling of CDKN1A gene in a range of studies with diverse cell lines from our database that are treated with (Studies 2–5) or without radiation as stimulus (Study 14). A comparative gene profiling across multiple studies (Figure 5, Step 2) showed a significant induction of the gene selectively in radiation treated studies. In addition the induction is found to have no effect at low dose radiation (0.4 Gy in Study 3) indicating cellular response to radiation is dependent on dose rate used.

Conclusions

StRAP is an open-access resource developed primarily to support research on the effects of stress with major emphasis on ionizing radiation on cancer in a systems-biology context.

References

- Chon HS, Lancaster JM (2011) Microarray-based gene expression studies in ovarian cancer. *Cancer Control* 18: 8–15.
- Burnet NG, Benson RJ, Williams MV, Peacock JH (2000) Improving cancer outcomes through radiotherapy. Lack of UK radiotherapy resources prejudices cancer outcomes. *BMJ* 320: 198–199.
- Zhang Y, Szustakowski J, Schinke M (2009) Bioinformatics analysis of microarray data. *Methods Mol Biol* 573: 259–284.
- Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, et al. (2008) The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. *Nat Rev Cancer* 8: 37–49.
- Stafford P, Brun M (2007) Three methods for optimization of cross-laboratory and cross-platform microarray expression data. *Nucleic Acids Res* 35: e72.
- Shankavaram UT, Varma S, Kane D, Sunshine M, Chary KK, et al. (2009) CellMiner: a relational database and query tool for the NCI-60 cancer cell lines. *BMC Genomics* 10: 277.
- Jarvinen AK, Hautaniemi S, Edgren H, Auvinen P, Saarela J, et al. (2004) Are data from different gene expression microarray platforms comparable? *Genomics* 83: 1164–1168.
- Mah N, Thelin A, Lu T, Nikolaus S, Kuhbacher T, et al. (2004) A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol Genomics* 16: 361–370.
- Tan PK, Downey TJ, Spitznagel EL, Jr., Xu P, Fu D, et al. (2003) Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* 31: 5676–5684.

Currently data from twenty one studies have been integrated and made accessible through extensive query options, and a user-friendly web-based interface. Supported by statistical and quantitative analysis methods in the background, the resource overcomes the limits of databases dedicated to raw data exploration, making it possible to infer nontrivial knowledge (such as the differentially expressed genes in multiple studies).

Currently because of the limitation of the number of studies available, it may have limited biological significance. However, the framework of the database is flexible and would allow extensions with data from other types of cancer studies that will help in novel findings.

The database will be periodically updated with new studies and features. We plan, for example, to enable construction of interaction networks using literature text-mining, and information from the Human Protein Reference Database (HPRD) [27] and gene set enrichment analyses and visualizations.

Author Contributions

Conceived and designed the experiments: US KC PT. Performed the experiments: SJ BI MB SZ OC. Analyzed the data: SJ BI SZ OC. Contributed reagents/materials/analysis tools: SJ BI OC MB. Wrote the paper: OC US.

10. Altman NS, Hua J (2006) Extending the loop design for two-channel microarray experiments. *Genet Res* 88: 153–163.
11. Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, Collins PJ, Chu TM, et al. (2006) Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat Biotechnol* 24: 1140–1150.
12. Bevilacqua V, Pannarale P, Abbrescia M, Cava C, Paradiso A, et al. (2012) Comparison of data-merging methods with SVM attribute selection and classification in breast cancer gene expression. *BMC Bioinformatics* 13 Suppl 7: S9.
13. Bargaje R, Hariharan M, Scaria V, Pillai B (2010) Consensus miRNA expression profiles derived from interplatform normalization of microarray data. *RNA* 16: 16–25.
14. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30: 207–210.
15. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, et al. (2004) ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* 6: 1–6.
16. Parkinson H, Sarkans U, Kolesnikov N, Abeygunawardena N, Burdett T, et al. (2011) ArrayExpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res* 39: D1002–1004.
17. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, et al. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 30: e15.
18. Hubbell E, Liu WM, Mei R (2002) Robust estimators for expression analysis. *Bioinformatics* 18: 1585–1592.
19. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. *Genome Res* 12: 996–1006.
20. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, et al. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res* 39: D685–690.
21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25: 25–29.
22. Saldanha AJ (2004) Java Treeview—extensible visualization of microarray data. *Bioinformatics* 20: 3246–3248.
23. Chon BH, Loeffler JS (2002) The effect of nonmalignant systemic disease on tolerance to radiation therapy. *Oncologist* 7: 136–143.
24. Rieger KE, Hong WJ, Tusher VG, Tang J, Tibshirani R, et al. (2004) Toxicity from radiation therapy associated with abnormal transcriptional responses to DNA damage. *Proc Natl Acad Sci U S A* 101: 6635–6640.
25. Amundson SA, Grace MB, McLeland CB, Epperly MW, Yeager A, et al. (2004) Human in vivo radiation-induced biomarkers: gene expression changes in radiotherapy patients. *Cancer Res* 64: 6368–6371.
26. Rodningen OK, Overgaard J, Alsner J, Hastie T, Borresen-Dale AL (2005) Microarray analysis of the transcriptional response to single or multiple doses of ionizing radiation in human subcutaneous fibroblasts. *Radiother Oncol* 77: 231–240.
27. Prasad TS, Kandasamy K, Pandey A (2009) Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods Mol Biol* 577: 67–79.