

## Database tool

# ScreenTroll: a searchable database to compare genome-wide yeast screens

Peter H. Thorpe<sup>1,2,†</sup>, John C. Dittmar<sup>3,†</sup> and Rodney Rothstein<sup>1,\*</sup>

<sup>1</sup>Department of Genetics & Development, Columbia University Medical Center, 701 West 168th Street, NY 10032, USA, <sup>2</sup>Division of Stem Cell Biology & Developmental Genetics, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, NW7 1AA, UK and <sup>3</sup>Department of Biological Sciences, Columbia University, NY 10027, USA

\*Corresponding author: Tel: +1 212 305 1733; Fax: +1 212 923 2090; Email: rothstein@columbia.edu

†These authors contributed equally to this work.

Submitted 7 February 2012; Revised 22 March 2012; Accepted 30 March 2012

Systematic biological screens typically identify many genes or proteins that are implicated in a specific phenotype. However, deriving mechanistic insight from these screens typically involves focusing upon one or a few genes within the set in order to elucidate their precise role in producing the phenotype. To find these critical genes, researchers use a variety of tools to query the set of genes to uncover underlying common genetic or physical interactions or common functional annotations (e.g. gene ontology terms). Not only it is necessary to find previous screens containing genes in common with the new set, but also useful to easily access the individual manuscript or study that classified those genes. Unfortunately, no tool currently exists to facilitate this task. We have developed a web-based tool (ScreenTroll) that queries one or more genes against a database of systematic yeast screens. The software determines which genome-wide yeast screens also identified the queried gene(s) and the resulting screens are listed in an order based on the extent of the overlap between the queried gene(s) and the open reading frames (ORFs) characterized in each individual yeast screen. In a separate list, the corresponding ORFs that are found in both the queried set of genes and each individual genome-wide screen are displayed along with links to the relevant manuscript via NIH's PubMed database. ScreenTroll is useful for comparing a list of ORFs with genes identified in a wide array of published genome-wide screens. This comparison informs users whether any of their queried ORFs overlaps a previous study in the ScreenTroll database. By listing the manuscript of the published screen, users can read more about the phenotype associated with that study. Together, this information provides insight into the function of the queried genes and helps the user focus on a subset of them.

## Background

The creation of a comprehensive collection of non-essential open reading frame (ORF) deletions in the yeast *Saccharomyces cerevisiae* has made this organism a primary model for genomics and high-throughput biology (1). The genomics data generated using the gene deletion collection has been central in driving the development of systems biology (2). When analyzing the ORFs identified in a genome-wide yeast screen, it is possible to determine the genetic and physical interactions between them using relatively sophisticated approaches [for example, Biopixie, (3)]. Additionally, using gene ontology term enrichment analysis, it is possible to determine if functional categories

are enriched within a set of ORFs [reviewed in (4)]. There are also tools that provide an overview of multiple phenotypic properties (interactions, localization, etc.) for a given list of genes [for example, FunSpec (5)]. However, what is lacking is a database and search tool that (i) identifies common ORFs between a queried set and ORFs characterized by individual genome-wide studies, (ii) orders the results based upon the likelihood of the overlap and (iii) lists the manuscripts associated with the studies.

## Construction and content

We have assembled a database from published manuscripts of hundreds of groups of ORFs identified in, or derived

from, large-scale yeast screens. We have focused our attention upon datasets that have systematically utilized the non-essential gene deletion collection to assay a specific phenotype. However, there are an increasing number of screens included in ScreenTroll using collections of mutant alleles of essential genes. There are two types of screens that are commonly reported: first, those that list a set of ORFs as affected. For example, Alvaro *et al.* (6) screened each non-essential gene deletion for its ability to increase the frequency of nuclear foci of Rad52, a key DNA repair protein. That study produced a list of 86 ORF deletions with this phenotype. The database entry for that study includes a short description of the screen phenotype: Elevated Rad52 foci, a summary of the manuscript describing the screen: Alvaro *et al.* 2007. Plos Genetics. 3; e228 with its PubMed ID number and finally, a list of the ORFs that were identified in the study. Much of this type of data is not included in other databases and these data are the core of the ScreenTroll database.

The second type is genome-wide screens that report quantitative data for each deletion, but do not necessarily provide a cutoff value or a defined list of affected ORFs. For example, a growth ratio on the experimental condition versus the control condition is listed for each deletion. To assemble a list of ORF deletions from these quantitative screens, we chose a specific cutoff value to generate a list of ORFs with the strongest phenotype and have indicated this cutoff value in the screen description. For example, the description '1.5 M sorbitol sensitive at 15 generations (competition assay >100 fitness defect)', indicates that the selection of strains chosen for the ScreenTroll database from the study of Giaever *et al.* (7) are sensitive to 1.5 M sorbitol after 15 generations and showed a 'fitness defect' greater than 100. Further details of the definition of 'fitness defect' are clearly explained in the manuscript describing the screen, whose link is accessible directly from the ScreenTroll output.

The ScreenTroll database includes most of the significant interactions from the Costanzo *et al.* (8) large-scale synthetic genetic array (SGA) study, where more than 1700 different query gene deletions were assayed against the entire library. The data from these screens were reported quantitatively and we have included ORFs within an intermediate cutoff value defined by the authors ( $|\epsilon| > 0.08$ ,  $P < 0.05$ ). Thus, the description of the database entry for the synthetic interactions with *cln2Δ* is labeled 'Costanzo SGA Screen, Intermediate Cutoff ( $|\epsilon| > 0.08$  &  $p$ -value < 0.05) - Query: YPL256C (CLN2)'.

## Utility and discussion

To query the ScreenTroll database for commonalities, we have built a web-based search tool that enables users to enter one or more yeast ORFs (the 'query set') into a

single search window. The screens that most closely match the query set are listed in a rank order based upon a 'rank score' (a description of the statistical methods used to evaluate this score is provided on the website and in [Supplementary Data](#)). The rank order is not a precise statistical ranking, but allows the user to focus on screens with extensive overlap as well as screens that identify mutually exclusive sets of ORFs (highlighted in blue). This latter group also provides functional insight since mutual exclusivity likely indicates that the two different phenotypes result from separate molecular pathways. Additionally, when ScreenTroll identifies an overlap, a list of ORFs in common with the query set is provided along with a link to the PubMed reference for the manuscript describing the particular screen. This feature facilitates access to the details of each screen enabling users to evaluate the potential biological significance of the individual ORFs identified.

The ScreenTroll output is ordered by 'rank score', which is a calculation of the hypergeometric  $P$ -value of each comparison. A simple adjustment for multiple comparison testing using the Bonferroni method (9) can be readily applied by multiplying the rank score by the number of screens tested (provided at the top of the results screen). We purposely use the term 'rank score' as opposed to ' $P$ -value' for a number of reasons. First, the rank score assumes that both the user's query set and each screen in the database are derived from the same set of 4800 strains in the viable yeast deletion collection. However, this may not be the case. Second, each published yeast screen has its own, often unknown or unreported, false positive and negative discovery rates, which directly affects the likelihood of an overlap. Further, we cannot predict the user's own false discovery rates. Third, we feel that the biological importance of an overlap between two groups of ORFs is best determined by carefully examining the manuscript describing how the ORF list was derived.

Some screens in the ScreenTroll database are the result of characterizing each deletion strain individually, however, many screens use a competition method. For this latter approach, the entire deletion collection is pooled together and exposed to experimental conditions (7). Subsequent microarray hybridization analysis of 'bar code' sequences specific for each yeast deletion reveals the relative levels of each strain in the pooled population. In this way, strains affected by the experimental condition are identified. In one such study, an exhaustive list of hundreds of different conditions and compounds were tested in both homozygous and heterozygous diploid strains (10) and much of this data is included in the ScreenTroll database. However, competition assays do not directly test each strain separately and some users may prefer to exclude this type of data from their analysis. Consequently, ScreenTroll includes the option of excluding data from competition assays in each search.

**SCREENTROLL RESULTS**

[Click here to download your results as an Excel file](#)

There are 86 ORFs in your dataset: "Alvero et al 2007".  
These were compared with 97 screens on file.

The lower the rank score the more significant the representation

#	Screen	ORFs in screen	# common hits	Rank Score
1	Elevated Rad52 foci	86	86	1.35e-186
2	Chromosome instability (CIN) genes	130	25	4.92e-20
3	MMS Sensitive	103	18	9.48e-14
4	Synthetic lethal/sick with siz1 siz2 double null	238	25	1.46e-13
5	Most (+++) MMS Sensitive	78	16	1.89e-13
6	Camptothecin Sensitive	82	16	4.35e-13
7	Synthetic lethal/sick with ulp1-333sgg	332	26	4.20e-11
8	Hydroxyurea Sensitive	133	17	9.46e-11
9	Synthetic lethal/sick with ubc9-2	313	24	4.50e-10
10	2-dimethylaminoethyl chloride sensitive (all statistically significant hits)	60	11	5.34e-09
11	Increased LOH in at least one assay	61	11	6.42e-09

The output was truncated here in the interest of space

**OVERLAP SUMMARY**

Gene names in parenthesis, if they exist.

**1**

There are 86 ORFs in the Elevated Rad52 foci screen.

[Alvaro et al. 2007. Plos Genetics. 3; e228.](#)

86 ORFs overlapped with your dataset: "Alvero et al 2007":

YJL142C (IRC9), YPL024W (RMI1), YGR078C (PAC10), YDL070W (BDF2), YOL002C (IZH2), YFL053W (DAK2), YOL115W (PAP2), YOR135C (IRC14), YMR073C (IRC21), YLL002W (RTT109), YEL001C (IRC22), YOL025W (LAG2), YJL030W (MAD2), YDL059C (RAD59), YGL163C (RAD54), YKR082W (NUP133), YPL194W (DDC1), YMR190C (SGS1), YDR332W (IRC3), YDR289C (RTT103), YJL047C (RTT101), YMR129W (POM152), YLR399C (BDF1), YFR055W (IRC7), YDR386W (MUS81), YMR027W ( ), YHR154W (RTT107), YJL003W (COX16), YMR075W (RCO1), YLR247C (IRC20), YDR116C (MRPL1), YDR446W (ECM11), YOL015W (IRC10), YPR164W (MMS1), YKR010C (TOF2), YJL037W (IRC18), YDR112W (IRC2), YDR540C (IRC4), YFR043C (IRC6), YAR002W (NUP60), YLR021W (IRC25), YFL023W (BUD27), YOL076W (MDM20), YER095W (RAD51), YOR144C (ELG1), YDR138W (HPR1), YOR025W (HST3), YOR375C (GDH1), YPL017C (IRC15), YFR049W (YMR31), YPR135W (CTF4), YOR073W (SGO1), YER116C (SLX8), YDR485C (VPS72), YJL013C (MAD3), YFR038W (IRC5), YPR038W (IRC16), YPL018W (CTF19), YJL051W (IRC8), YKL003C (MRP17), YPR070W (MED1), YPL013C (MRP516), YIR036C (IRC24), YHR134W (WSS1), YKL208W (CBT1), YOR328W (PDR10), YML116W (ATR1), YOR023C (AHC1), YGL175C (SAE2), YOL052C-A (DDR2), YMR063W (RIM9), YDR004W (RAD57), YGL086W (MAD1), YOR044W (IRC23), YOR013W (IRC11), YMR055C (BUB2), YDR363W (ESC2), YML041C (VPS71), YMR167W (MLH1), YLR260W (LCB5), YLL033W (IRC19), YHR031C (RRM3), YOL049W (GSH2), YJL095W (BCK1), YDR439W (LRS4), YOR235W (IRC13)

**2**

There are 130 ORFs in the Chromosome instability (CIN) genes screen.

[Yuen et al. 2007 PNAS 104; 3925-3930.](#)

25 ORFs overlapped with your dataset: "Alvero et al 2007":

YPL024W (RMI1), YPL017C (IRC15), YGR078C (PAC10), YOR073W (SGO1), YPR135W (CTF4), YDR004W (RAD57), YLL002W (RTT109), YGL086W (MAD1), YER116C (SLX8), YPR164W (MMS1), YDR363W (ESC2), YJL030W (MAD2), YDL059C (RAD59), YKR082W (NUP133), YGL163C (RAD54), YPL194W (DDC1), YHR031C (RRM3), YMR190C (SGS1), YPL018W (CTF19), YDR386W (MUS81), YHR134W (WSS1), YHR154W (RTT107), YER095W (RAD51), YOR025W (HST3), YOR144C (ELG1)

**3**

There are 103 ORFs in the MMS Sensitive screen.

[Chang et al PNAS 2002. 99;16934-16939.](#)

18 ORFs overlapped with your dataset: "Alvero et al 2007":

YPL024W (RMI1), YGL175C (SAE2), YPR135W (CTF4), YDR004W (RAD57), YLL002W (RTT109), YPR164W (MMS1), YKR082W (NUP133), YDL059C (RAD59), YGL163C (RAD54), YPL194W (DDC1), YMR190C (SGS1), YJL047C (RTT101), YLR399C (BDF1), YPR070W (MED1), YDR386W (MUS81), YHR154W (RTT107), YER095W (RAD51), YDR138W (HPR1)

**4**

There are 238 ORFs in the Synthetic lethal/sick with siz1 siz2 double null screen.

[Makhnevych et al. 2009. Molecular Cell 33; 124-135](#)

25 ORFs overlapped with your dataset: "Alvero et al 2007":

The output was truncated here in the interest of space

**Figure 1.** Example of a ScreenTroll Output. Portions of the output from a ScreenTroll search using the set of 86 ORFs identified by Alvaro *et al.* (6) are shown. ScreenTroll displays a table of the screens in the database ordered by a rank score of the overlap with the query set (top). Below the table, the 'overlap summary' includes more details of each screen including the specific ORFs that overlap with the query set and a link to the PubMed entry for the manuscript that describes the screen (bottom). Both the competition and the Costanzo *et al.* (8) SGA screens were excluded from this search.

ScreenTroll was initially designed to highlight screens that include significant analysis of the associated phenotype. These screens, although genome wide, generally focus on a specific mechanism or phenotype (e.g. Rad52 focus formation, methyl methanesulfonate (MMS) sensitivity, chromosome instability, etc.) and the manuscripts associated with them provide considerable detail about their findings. Nevertheless, as noted above, we have also

included most of the very large scale screening data from the Costanzo *et al.* (8) SGA screens. However, some users may prefer to restrict their search to the core ScreenTroll data set, since the massive amount of data from the SGA screens may overwhelm the output and mask overlaps with the more focused screens. Hence, we have included an option to exclude the data from these large-scale synthetic genetic array experiments.



Primarily, users will enter a set of ORFs identified in a new screen, 'the query set,' to compare with those identified from other screens. If a strong match is found, it suggests that both the user's screen and the published screen share a common feature. For example, we entered a set of ORFs identified by our laboratory as being important to prevent high levels of Rad52 foci (6). ScreenTroll identifies screens that assay for chromosomal instability (11), sensitivity to methyl methanesulfonate (12) and the sumoylation pathway (13), as those that most closely match the query set (a portion of the ScreenTroll output is shown in Figure 1). These matches confirm the shared pathway of DNA damage repair for all of these screens and highlight potentially new insights into the role of sumoylation in regulating the DNA damage response. In addition, having the complete list of overlapping screens is useful since some of the individual ORFs further down the list, which are common to a particular screen, may be of interest to the user. For example, a screen for propanol sensitivity identified *IRC15* and *IRC25*, two previously uncharacterized ORFs from the Alvaro screen (14).

If the user is interested in exploring a new or existing pathway, ScreenTroll can be used to query the ORFs that encode that pathway to determine whether they were enriched in previous screens. For example, the spindle assembly checkpoint (SAC) is a key regulator of mitosis and it is possible to query the database with *MAD1*, *MAD2*, *MAD3*, *BUB1* and *BUB3*, each of which encode key non-essential components of the SAC [see (15) and references therein]. The ScreenTroll output from this query can be viewed by selecting the example provided on the ScreenTroll website. At the time of publication, the first four screens [excluding the Costanzo *et al.* (8) SGA data] that most closely match this query set are (i) gene deletions that are synthetic lethal with kinetochore mutants (16), (ii) deletions that fail to maintain an 'originless' plasmid (17), (iii) deletions that are sensitive to the microtubule poison benomyl (18) and (iv) deletions that result in chromosome instability (11). Since the SAC proteins are located at the kinetochore and help to direct chromosome segregation, these data are consistent with the known mechanism of the SAC. However, the fifth screen listed is a screen for increased Rad52 DNA repair centers (6), reinforcing a role for the SAC in preventing DNA damage (19,20).

If a user is interested in a single gene, ScreenTroll can list all of the screens that identified it. For example, if a user enters *RAD50*, the results show that this gene was identified in numerous genome-wide screens for DNA damage sensitivity, consistent with its known role in DNA repair.

Finally, there are a wealth of gene-gene and protein-protein interaction data available for yeast (8,16,18,21–25) and excellent tools to query these data [for examples, see (26,27)]. The ScreenTroll webpage provides a link to access many of these tools.

## Conclusions

Using this simple tool, similarities between screens are revealed and listed in rank order. The results of ScreenTroll are useful for deciding which ORFs identified in a new screen are of specific interest due to a shared phenotype. Moreover, identifying the 'screen phenotype' conferred by deletion of a specific ORF, or set of ORFs, can illuminate the biological function of the encoded protein(s) and aid in the design of new assays to test its function. We envision that ScreenTroll will be of use to anyone interested in analyzing the results of yeast genomic data. The package with documentation is available at <http://www.rothsteinlab.com/tools/apps/screenTroll>. This website includes the option to separately download the entire database, the source code for the application and information about the statistics used to generate the rank score. There are no access restrictions. In addition, ScreenTroll is available through individual ORF pages on the Saccharomyces Genome Database ([yeastgenome.org](http://yeastgenome.org)).

## Supplementary Data

Supplementary data are available at Database Online.

## Acknowledgments

We would like to thank our colleagues Michael Chang, Bob Reid, Kara Bernstein and Steven Pierce for helpful discussion. J.C.D. and P.H.T. assembled the database from the published literature and created the search software. J.C.D., P.H.T. and R.R. wrote the manuscript.

## Funding

National Institutes of Health (GM50237 and GM67055 to R.R. and GM008798 and CA009503 to J.C.D.). Funding for open access charge: National Institutes of Health (GM50237 and GM67055).

*Conflict of interest.* None declared.

## References

1. Winzeler, E.A., Shoemaker, D.D., Astromoff, A. *et al.* (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, **285**, 901–906.
2. Snyder, M. and Gallagher, J.E. (2009) Systems biology from a yeast omics perspective. *FEBS Lett.*, **583**, 3895–3899.
3. Myers, C.L., Robson, D., Wible, A. *et al.* (2005) Discovery of biological networks from diverse functional genomic data. *Genome Biol.*, **6**, R114.
4. Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.

5. Robinson,M.D., Grigull,J., Mohammad,N. and Hughes,T.R. (2002) FunSpec: a web-based cluster interpreter for yeast. *BMC Bioinformatics*, **3**, 35.
6. Alvaro,D., Lisby,M. and Rothstein,R. (2007) Genome-wide analysis of Rad52 foci reveals diverse mechanisms impacting recombination. *PLoS Genet.*, **3**, e228.
7. Giaever,G., Chu,A.M., Ni,L. et al. (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387–391.
8. Costanzo,M., Baryshnikova,A., Bellay,J. et al. (2010) The genetic landscape of a cell. *Science*, **327**, 425–431.
9. Bonferroni,C.E. (1936) Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, **8**, 3–62.
10. Hillenmeyer,M.E., Fung,E., Wildenhain,J. et al. (2008) The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science*, **320**, 362–365.
11. Yuen,K.W., Warren,C.D., Chen,O. et al. (2007) Systematic genome instability screens in yeast and their potential relevance to cancer. *Proc. Natl Acad. Sci. USA*, **104**, 3925–3930.
12. Chang,M., Bellaoui,M., Boone,C. and Brown,G.W. (2002) A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage. *Proc. Natl Acad. Sci. USA*, **99**, 16934–16939.
13. Makhnevych,T., Sydorskyi,Y., Xin,X. et al. (2009) Global map of SUMO function revealed by protein-protein interaction and genetic networks. *Mol. Cell*, **33**, 124–135.
14. Auesukaree,C., Damernsawad,A., Kruatrachue,M. et al. (2009) Genome-wide identification of genes involved in tolerance to various environmental stresses in *Saccharomyces cerevisiae*. *J. Appl. Genet.*, **50**, 301–310.
15. Skibbens,R.V. and Hieter,P. (1998) Kinetochores and the checkpoint mechanism that monitors for defects in the chromosome segregation machinery. *Annu. Rev. Genet.*, **32**, 307–337.
16. Measday,V., Baetz,K., Guzzo,J. et al. (2005) Systematic yeast synthetic lethal and synthetic dosage lethal screens identify genes required for chromosome segregation. *Proc. Natl Acad. Sci. USA*, **102**, 13956–13961.
17. Theis,J.F., Irene,C., Dershowitz,A. et al. (2010) The DNA damage response pathway contributes to the stability of chromosome III derivatives lacking efficient replicators. *PLoS Genet.*, **6**, e1001227.
18. Parsons,A.B., Brost,R.L., Ding,H. et al. (2004) Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat. Biotechnol.*, **22**, 62–69.
19. Garber,P.M. and Rine,J. (2002) Overlapping roles of the spindle assembly and DNA damage checkpoints in the cell-cycle response to altered chromosomes in *Saccharomyces cerevisiae*. *Genetics*, **161**, 521–534.
20. Kim,E.M. and Burke,D.J. (2008) DNA damage activates the SAC in an ATM/ATR-dependent manner, independently of the kinetochore. *PLoS Genet.*, **4**, e1000015.
21. Ito,T., Chiba,T., Ozawa,R. et al. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA*, **98**, 4569–4574.
22. Uetz,P., Giot,L., Cagney,G. et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature*, **403**, 623–627.
23. Venkatesan,K., Rual,J.F., Vazquez,A. et al. (2009) An empirical framework for binary interactome mapping. *Nat. Methods*, **6**, 83–90.
24. Wang,H., Kakaradov,B., Collins,S.R. et al. (2009) A complex-based reconstruction of the *Saccharomyces cerevisiae* interactome. *Mol. Cell Proteomics*, **8**, 1361–1381.
25. Valente,A.X., Roberts,S.B., Buck,G.A. and Gao,Y. (2009) Functional organization of the yeast proteome by a yeast interactome map. *Proc. Natl Acad. Sci. USA*, **106**, 1490–1495.
26. Hu,Z., Hung,J.H., Wang,Y. et al. (2009) VisANT 3.5: multi-scale network visualization, analysis and inference based on the gene ontology. *Nucleic Acids Res.*, **37**, W115–W121.
27. Lopes,C.T., Franz,M., Kazi,F. et al. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.