BMC Bioinformatics

Software

TagSmart: analysis and visualization for yeast mutant fitness data measured by tag microarrays

Chulyun Kim[1,6], Sangkyum Kim[1], Russell Dorer[3], Dan Xie[2], Jiawei Han[1] and Sheng Zhong*[1,2,4,5]

Address: [1]Department Computer Science, University of Illinois at Urbana Champaign, Urbana, IL, USA, [2]Department of Bioengineering, University of Illinois at Urbana Champaign, Urbana, IL, USA, [3]Department of Pathology, Virginia Mason Medical Center, 1100 Ninth Avenue, Seattle, WA, USA, [4]Department of Statistics, University of Illinois at Urbana Champaign, Champaign, IL, USA, [5]Institute for Genomic Biology, University of Illinois at Urbana Champaign, Urbana, IL, USA and [6]School of Electrical Engineering and Computer Science, Seoul National University, South Korea

Email: Chulyun Kim - ckim5@uiuc.edu; Sangkyum Kim - kim71@uiuc.edu; Russell Dorer - russell.dorer@vmmc.org; Dan Xie - danxie2@uiuc.edu; Jiawei Han - hanj@cs.uiuc.edu; Sheng Zhong* - szhong@uiuc.edu

* Corresponding author

Abstract

Background: A nearly complete collection of gene-deletion mutants (96% of annotated open reading frames) of the yeast *Saccharomyces cerevisiae* has been systematically constructed. Tag microarrays are widely used to measure the fitness of each mutant in a mutant mixture. The tag array experiments can have a complex experimental design, such as time course measurements and drug treatment with multiple dosages.

Results: TagSmart is a web application for analysis and visualization of *Saccharomyces cerevisiae* mutant fitness data measured by tag microarrays. It implements a robust statistical approach to assess the concentration differences among S. cerevisiae mutant strains. It also provides an interactive environment for data analysis and visualization. TagSmart has the following advantages over previously described analysis procedures: 1) it is user-friendly software rather than merely a description of analytical procedure; 2) It can handle complicated experimental designs, such as multiple time points and treatment with multiple dosages; 3) it has higher sensitivity and specificity; 4) It allows users to mask out "bad" tags in the analysis.

Two biological tests were performed to illustrate the performance of TagSmart. First, we generated titration mixtures of mutant strains, in which the relative concentration of each strain was controlled. We used tag microarrays to measure the numbers of tag copies in each titration mixture. The data was analyzed with TagSmart and the result showed high precision and recall. Second, TagSmart was applied to a dataset in which heterozygous deletion strain mixture pools were treated with a new drug, Cincreasin. TagSmart identified 53 mutant strains as sensitive to Cincreasin treatment. We individually tested each identified mutant, and found 52 out of the 53 predicted mutants were indeed sensitive to Cincreasin.

Conclusion: TagSmart is provided "as is" to analyze tag array data produced by Affymetrix and Agilent arrays. TagSmart web application is assessable by Windows, Mac, and Linux users. It also has a downloadable version for execution on PCs running Windows. TagSmart is available for academic use at: http://biocomp.bioen.uiuc.edu/tagsmart

1. Background

A nearly complete collection of gene-deletion mutants (96% of annotated open reading frames) of the yeast *Saccharromyces cerevisiae* has been systematically constructed [1,2]. Each deletion is marked with two unique oligonucleotide tags, making it possible to use microarrays and the tag arrays [3] to measure the relative abundance of each mutant strain in a mixture. The relative fitness of every individual gene deletion mutant can be compared across multiple intercellular environments. A general question of interest is which gene-deletion and environment interaction is most lethal or most viable.

Every mutant is barcoded with two tags, namely the uptag and the downtag. The two tags are deletion-specific. They are synthesized into the genome of the deletion strain at the location of the deleted gene. For Affymetrix tag arrays, four probes on the microarray are designed to hybridize to the sense and antisense strands of each tag. These probe sets are indicated by Perfect Match (PM), Mis-Match (MM), complementary Perfect Match (cPM), and complementary Mis-Match (cMM). In summary, every mutant strain is represented by eight probe readouts: uptag-PM, uptag-cPM, uptag-MM, uptag-cMM, downtag-PM, downtag-cPM, downtag-MM, downtag-cMM [Additional file 1]. Please refer to [4] for details of Agilent tag arrays.

A general experiment design is a two-environment, multiple-time-point design [1,2,5]. Two mixtures of gene-deletion mutants are grown under two different environmental conditions: a drug-treated condition and a control condition. Mutant samples are collected from both collections at a series of time points, e.g., 4, 8, and 16 cell generations. DNA of these samples are retrieved, amplified, and hybridized to tag arrays. We summarize the experimental designs and analytical procedures in published literature at below and in Table 1.

A few analytical procedures have been proposed to analyze tag microarray data. (see Additional file 2 for a detailed review of these procedures.) Most of these procedures were designed to handle a specific dataset generated by a specific experiment, and therefore are not applicable to analyzing other data generated from a different experimental design, with exceptions to the procedure described by Giaever et al. [2] and the procedure we recently proposed [5]. The latter procedure is more general than the Giaever procedure because it can handle treatment with multiple dosages. The TagSmart software implements the latter procedure (hereafter referred to as the TagSmart procedure). When there is a complex experimental design (e.g., multiple time points, or multiple dosages), TagSmart will take the most advantage of the comprehensive data available. When data is generated from a simple experimental design, the TagSmart procedure will automatically degenerate into a simpler procedure.

2. Implementation

2.1 TagSmart software

TagSmart [3] is a web application that can be operated by web browsers, such as Mozilla Firefox, Safari, and Internet Explorer. TagSmart also has a downloadable version for execution on PCs running on Windows. TagSmart has three modules: data preprocessing, computation, and visualization.

Data preprocessing module
The data preprocessing module integrates data files and annotation files into one easily interpretable data file. The following files are required as input files to the preprocessing module: 1) a series of tag array data files, in either CEL or TXT format; 2) a chip description file (CDF file); 3) an array description file; 4) a tag mask file; and 5) a user-supplied experiment description file. Except that the tag array data files and experiment description file should be provided by the user, all the other files can be downloaded from the TagSmart website. The CEL format data files are direct outputs of an Affymetrix scanner. If users have applied Affymetrix software such as GCOS [3] to process the data, they may have data files in TXT format. TagSmart allows users to supply data files in either CEL or TXT format. The CDF file records the coordinates of each tag on the tag microarray. The array description file links each tag to its corresponding open reading frame and gene name. Because the same tag was used to make both the homozygous deletion mutant [1] and the heterozygous deletion mutant [6], the analysis of heterozygous and homozygous mutants share the same array description file (the analysis of homozygous mutants only uses a subset of this file). The tag mask file records a list of tags that do not show "responsiveness" to the concentration change of their corresponding mutants. These tags are considered to be bad (the procedure of detecting bad tags will be described later). The user can choose to mask out the "bad" tags in the subsequent analysis by clicking on the "Bad tag filtering" checkbox (Additional file 3). Finally, the experiment description file is a user-supplied file, recording the experimental condition (e.g., treated/control, dosage, time, etc.) for each array data file. Users should follow the instruction on the TagSmart help-page to construct this file. The output of the preprocessing module is an easily interpretable data file in tabular format. Instead of averaging the multiple signals of a mutant (e.g., four PM signals in the Affymetrix platform), the preprocessing module retains them individually in the output.

Table 1: Summary of experimental design and data structure in previous studies.

	Multiple time points	Multiple treatment dosages	Replicates in treatment	Replicates in control
Winzeler et al [1]	Y	N	N	N
Ooi et al [15]	N	N	Y	Y
Ooi et al [16]	N	N	Y	Y
Giaever et al [2]	Y	N	N	Y
Warren et al [17]	N	N	Y	Y
Lee et al [18]	N	N	Y	Y
Pan et al [19]	N	Y	Y	Y
Lum et al [6]	N	N	N	Y
Birrell et al [20]	N	N	Y	Y
Dorer et al [5]	Y	Y	Y	Y
Yuan et al [4]	N	N	Y	Y
Peyser et al [21]	N	N	Y	Y

Computation module

By choosing the radio button of "Analyze a preprocessed data file" in the main page, users activate the computation module. Users should specify desired criteria for selecting mutants in the subsequent webpage (Additional file 4). Fold Change (FC) and Q-value (equivalent to false discovery rate) are allowed. If the user has array data for a common pool of mutant mixture, such as a time 0 sample before the separation of mutant growth in treatment and control, she/he can choose to use such data to get more precise estimate of mutant growth rates. This is achieved by checking the "Generation-0 correction" check box (The procedure to handle a common mixture pool will be described later). Advanced users are also allowed to tune a parameter called the number of permutations. As in the Statistical Procedure section described below, TagSmart employs a matched permutation method to obtain background distribution. The number of permutation is positively correlated with the accuracy of computation but also computation time. We suggest a default number of 500 permutations, which is a balance between accuracy and time. Additional file 5 gives an example text output of a computation. Mutants that satisfied the user-defined thresholds are listed. Their related information, including open reading frame's name (ORF), gene name, the two associated tags, p-value, q-value, and fold change are provided. Users can sort the output by any information with a click on the corresponding column name. A more comprehensive report, including the actual data and experimental conditions, can be saved as a text file by clicking the disk icon on the output webpage.

Visualization module

An interactive graphical display of the computation result is accessible by clicking the "heatmap" icon after the computation (Figure 1A). Alternatively, the saved report file on a user's local computer can be uploaded onto the server and visualized using the visualization module. TagSmart adopts heatmap as the way to present mutant's relative concentration. The first two rows in the heatmap use a novel color scheme to represent experimental design information. The first row represents the treatment factor (treatment 0, treatment 1, etc.). The second row represents the time factor (Generation 0, Generation 4, etc.). From the third row on, a traditional heatmap is applied to show the relative concentration of each mutant. Red represents higher concentration and green represents lower concentration. Detailed information on treatment, time, and array signal can be monitored by moving mouse cursor over the corresponding color-coded region (Figure 1B and 1C).

We tested TagSmart with multiple web browsers under Linux, MacOS, and Windows operation systems. TagSmart is implemented with C# programming language and ASP.NET technology. TagSmart is currently hosted on a Dell rack server with dual 3 GHz Intel(R) Xeon(TM) dual-core processors and 6GB RAM. A standalone executable for Windows is also downloadable from the TagSmart website.

2.2. Statistical Procedure

We describe TagSmart's statistical procedure assuming data comes from the most complicated setting (i.e., multiple time points and treatment with multiple dosages). This procedure automatically degenerates into a simpler procedure when data comes from a simpler experimental design.

Array signals are first normalized to make 1% trimmed mean the same across all arrays [7]. Let $y_{i\alpha\beta\gamma\delta\tau}$ be the normalized signal for mutant i, in environmental condition α, at time point β, measured by tag γ (*uptag* and *downtag*) and probe δ (*PM* and *cPM*), on the replicate array τ. For notational simplicity, we will suppress the mutant indicator *i* hereafter. When time-0 data is available, users can use the following metric to represent the growth rate meas-
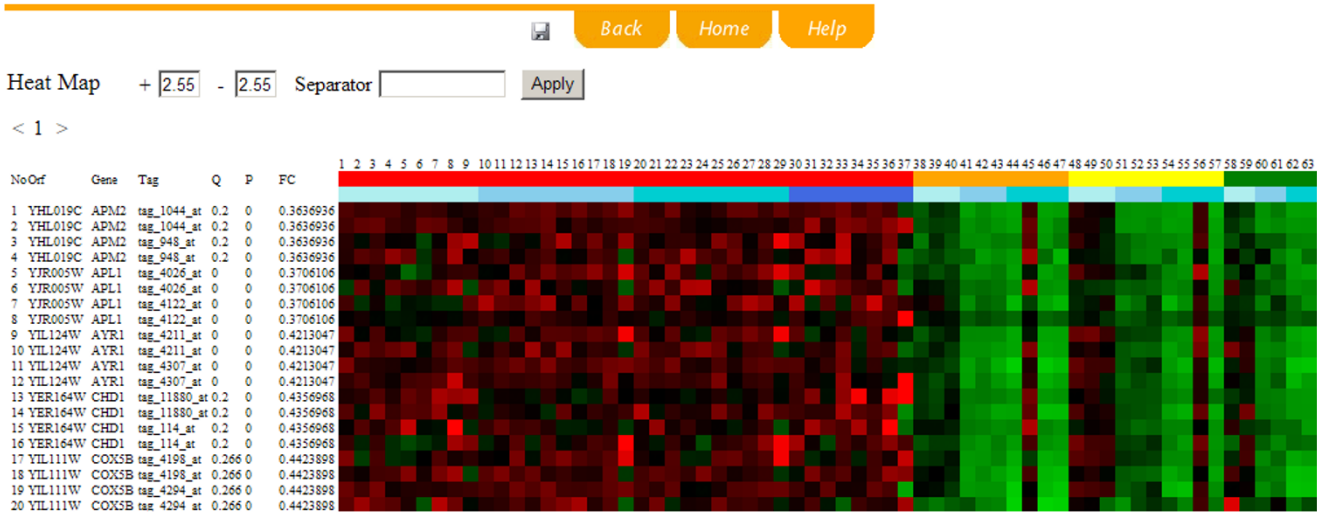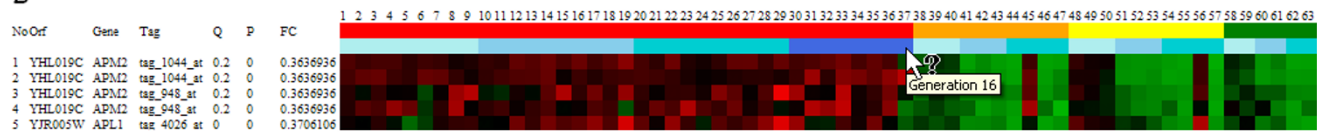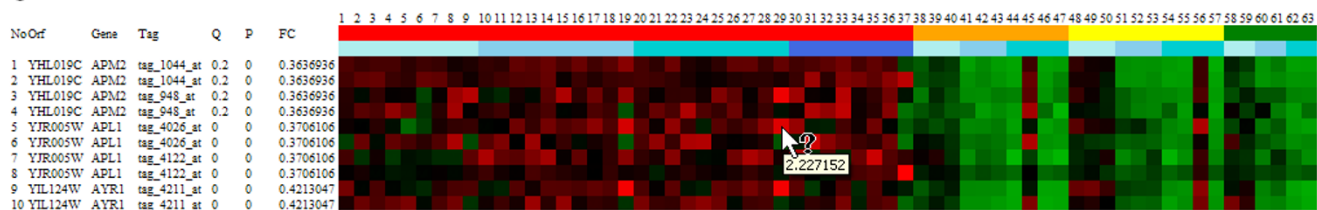
Figure 1
(A) An example TagSmart's graphical output. Every mutant is represented in 4 rows, representing signals from four tag-probe pairs (uptag-PM, uptag-cPM, downtag-PM, and downtag-cPM). The arrays are sequentially marked on the top row. The first two colored lines indicate the treatment and the time under which the mutant mixture was harvested and DNA was retrieved. The heatmap starts from the third colored line. Red indicates good fitness in treatment, and green indicates reduced fitness in treatment. (B) When mouse cursor is moved over the first two colored lines, the corresponding information for treatment and time will appear on screen. (C) When mouse cursor is over a dot in the heatmap, the actual fitness value will appear.

ured by a probe at time β: $x_{\alpha\beta\gamma\delta\tau} = \dfrac{y_{\alpha\beta\gamma\delta\tau} - \overline{y_{\beta=0,\gamma\delta}}}{\overline{y_{\beta=0,\gamma\delta}}}$,

where $\overline{y_{\beta=0,\gamma\delta}}$ is the average across all array replicates at time-0. The growth rate measurement $x_{\alpha\beta\gamma\delta\tau}$ was inspired by the Coefficient of Variation (CV) statistic. With the observation that signals with larger magnitude usually have larger variability, $x_{\alpha\beta\gamma\delta\tau}$ can be regarded as a signal for concentration change, with the raw signal intensity penalized. When time-0 data is not available, the normalized signal is directly passed onto the next step: $x_{\alpha\beta\gamma\delta\tau} = y_{\alpha\beta\gamma\delta\tau}$

To compare the difference of concentration changes between different experimental conditions, we first compute a modified T statistic:

$$T = \frac{\overline{x_{\alpha=trt}} - \overline{x_{\alpha=contr}}}{s + s_0}$$

where $\overline{x_{\alpha=trt}}$ and $\overline{x_{\alpha=contr}}$ are average concentration changes under treated and untreated conditions, respectively. If the treatment was conducted with multiple dosages, $\overline{x_{\alpha=trt}}$ is computed with all the data from all

dosages. *s* is the pooled standard deviation of all data. It is computed by:

$$s = \left(\frac{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \times \left\{ \sum_{\alpha=trt,\beta,\gamma\delta\tau} [x_{\alpha=trt,\beta,\gamma\delta\tau} - \overline{x_{\alpha=trt}}]^2 + \sum_{\alpha=contr,\beta,\gamma\delta\tau} [x_{\alpha=contr,\beta,\gamma\delta\tau} - \overline{x_{\alpha=contr}}]^2 \right\}}{n_1 + n_2 - 2} \right)^{\frac{1}{2}}$$

where $\sum_{\alpha=trt}$ denotes sum over all treated signals. If there are multiple dosages, all these dosages should be summed. n_1 and n_2 are the numbers of *x* s under treated and untreated conditions, respectively. s_0 is a small positive constant (0.001) that ensures T not outrageously large.

We use a matched permutation strategy to generate background distribution for the T statistic. To illustrate the matched permutation procedure, we assume that data comes from the following hypothetical experiment. Let A and B denote two mutant mixtures treated with high and low dosages of a drug, respectively. Mutant mixtures C and D are replicates grown under control condition. Mutant samples are collected at 5 and 15 cell generations. In this hypothetical experiment, 8 tag arrays are used to gather data. We permute $x_{\alpha\beta\gamma\delta\tau}$ with the same time (β), tag (γ), and probe (δ). To see this permutation procedure graphically, signals are arranged in Figure 2 and color coded. The signals are permutated under the constraint that a signal can only be switched to a box with the same color as its original one.

With K permutations, we obtain K new statistics T^k, k = 1, 2, ..., K. We compute a q-value (roughly equivalent to false discovery rate [8]) for every mutant. The statistical interpretation for a mutant's q-value is: if the threshold is set so that this mutant is the last mutant to be called significant, the q-value is the estimated percentage of false positives among all the mutants being called significant. We compute the q-value for a mutant by [9,10]:

$$\hat{q} = \frac{\frac{1}{K} \sum_{k=1}^{K} \sum_{j=1}^{N} I\{ |T_j^k| \ge |T| \}}{\max\{1, \sum_{j=1}^{N} I\{ |T_j| \ge |T| \}\}} \hat{\pi}_0(c_0)$$, where j = 1, 2, ..., N

is the index for mutants. T is the T-statistic computed from un-permuted data. $\hat{\pi}_0(c_0)$ is the estimated proportion of mutants with no fitness difference between experimental conditions. $\hat{\pi}_0(c_0)$ is estimated by [9,10]:

$$\hat{\pi}_0(c_0) = \frac{\frac{1}{N} \sum_{j=1}^{N} I\{ |T_j| < c_0 \}}{\frac{1}{NK} \sum_{k=1}^{K} \sum_{j=1}^{N} I\{ |T_j^k| < c_0 \}}$$, and C_0 is a pre-

defined constant (0.5). The choice of C_0 does not affect the estimate of $\hat{\pi}_0(c_0)$ as long as C_0 is reasonably small [8,9].

Finally, the fold change (FC) between treatment and control is computed by:

$FC = \sum_{\beta} a_\beta \times FC_\beta$. It is a weighted sum of each time point

fold change. We require $\sum_{\beta} a_\beta = 1$. Larger a_β will stress the

importance of that β time point. FC_β is the fold change at time β. It is defined as:

$$FC_\beta = \frac{1}{\#(\gamma,\delta)} \sum_{\gamma} \sum_{\delta} \frac{\sum_{\alpha=trt} \sum_{\tau} x_{\alpha\beta\gamma\delta\tau}}{\sum_{\alpha=contr} \sum_{\tau} x_{\alpha\beta\gamma\delta\tau}}$$

where α, β, γ, δ, and τ are defined the same as above. It is worth noticing that TagSmart does not first average all probe signals and then take the ratio, but rather it first takes ratio on the same probe and then averages over all tags and probes. TagSmart jointly uses q-value and FC to call significant mutants.

3. Results
Titration Experiment
To illustrate TagSmart's performance, we did a titration experiment using homozygous deletion mutants. Eight mutant mixture pools were made, which were denoted as pools A, B, C, D, E, F and G, respectively. The mutants had roughly equal concentrations in mixture pools A and G. One sixth of the mutants were diluted into 1/25 concentration whereas the concentration of the rest mutants were untouched in pool B. Another one sixth, not overlapping with the first one sixth, were diluted to 1/25 concentration in pool C, so did pools D, E, and F. In the end pools B to F each had one sixth of the mutants diluted. DNA from each mutant pool was hybridized to a tag microarray. TagSmart procedure was applied to identify the mutants with lower concentration in pools C to G. A wide range of thresholds for determining the mutants with lower concentration were applied, and for each threshold the computationally identified mutants were compared to the real diluted mutants. We computed the precision and the
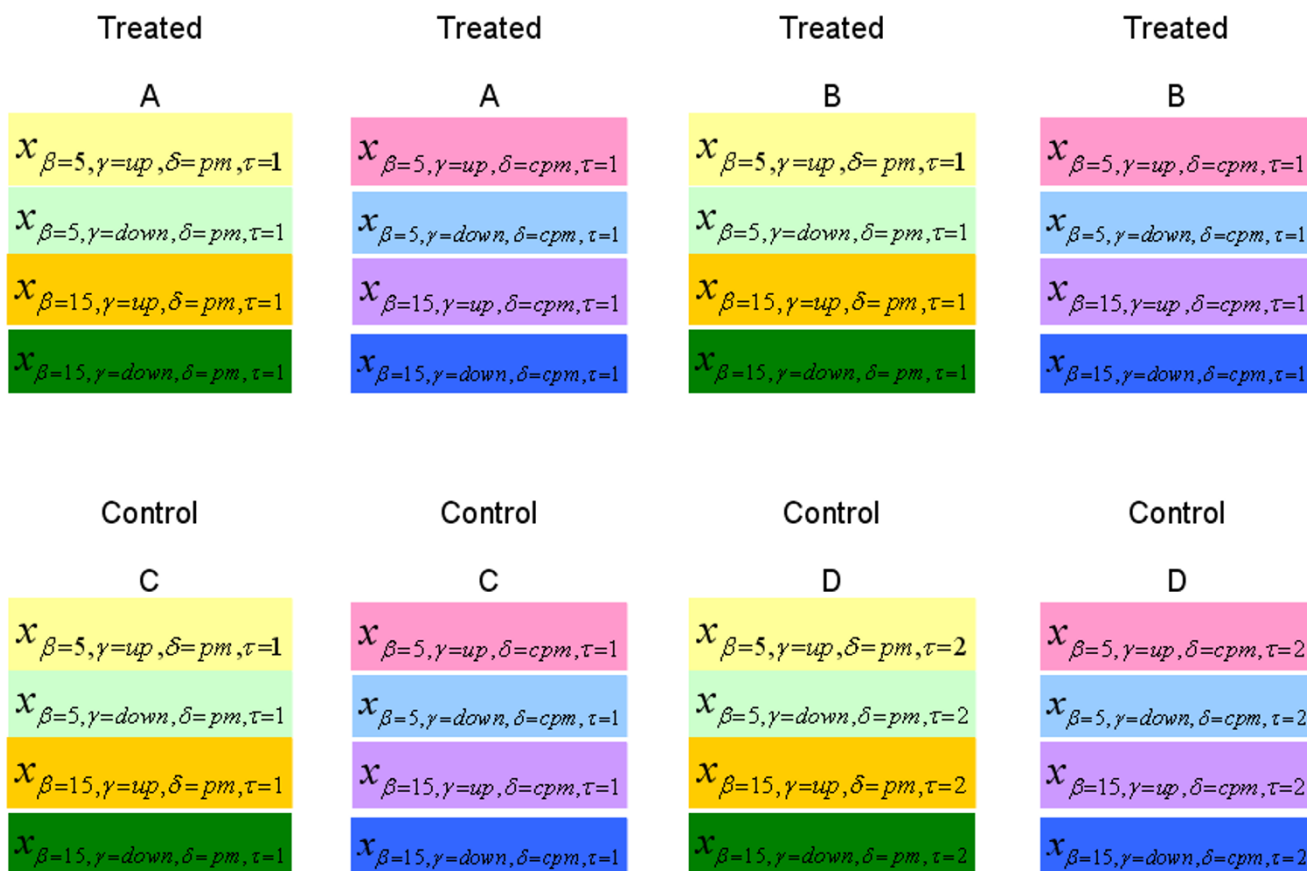
Figure 2
Data table for a mutant in a hypothetical experiment. This hypothetical experiment has two replicate mutant mixtures in control condition, and two treated mixtures under different treatment dosage. Mutant samples were harvested at both cell generations 5 and 15. Eight tag arrays should be used to gather the data. TagSmart's matched permutation procedure switches the signals with the same color.

recall of TagSmart procedure (Figure 3). Precision and recall are defined as follows.

$$\text{Precision} = \frac{\{\text{True positive}\} \cap \{\text{ Predicted positive }\}}{\{\text{ Predicted positive }\}}$$

$$\text{Recall} = \frac{\{\text{True positive}\} \cap \{\text{ Predicted positive }\}}{\{\text{ True positive }\}}$$

Figure 3 shows that at the precision of 0.4, TagSmart achieves recalls of 0.7 to 0.9 in the titration data.

The titration experiment allows us to detect the "bad" tags that do not show consistent signal change for the diluted mutants. Each mutant is diluted in one of the eight mixture pools. The diluted concentration is 1/25 of the concentration of the undiluted concentration. We employed the following procedure to detect "bad" tags. For each tag,

its signal from the diluted pool is compared to the average signal of this tag from the other seven undiluted pool (each mutant is only diluted in one of the eight pools). A tag is regarded as "bad" if its signal from the diluted pool is not smaller than its average signal from the undiluted pools. The "bad" tags are recorded into the tag mask file, which, by user's discretion, can be used to eliminate the bad tags from the subsequent analysis (see the preprocessing module). One reason for a tag being "bad" can attribute to the mutations of the synthetic DNA tags introduced during the construction of the deletion strains [11]. We note that a "bad" tag should not be taken literally, because there are many reasons that can contribute to inconsistency between the signal of a tag and the concentration change. For example, cross-hybridization to the probe on the array may contribute to the inconsistency.
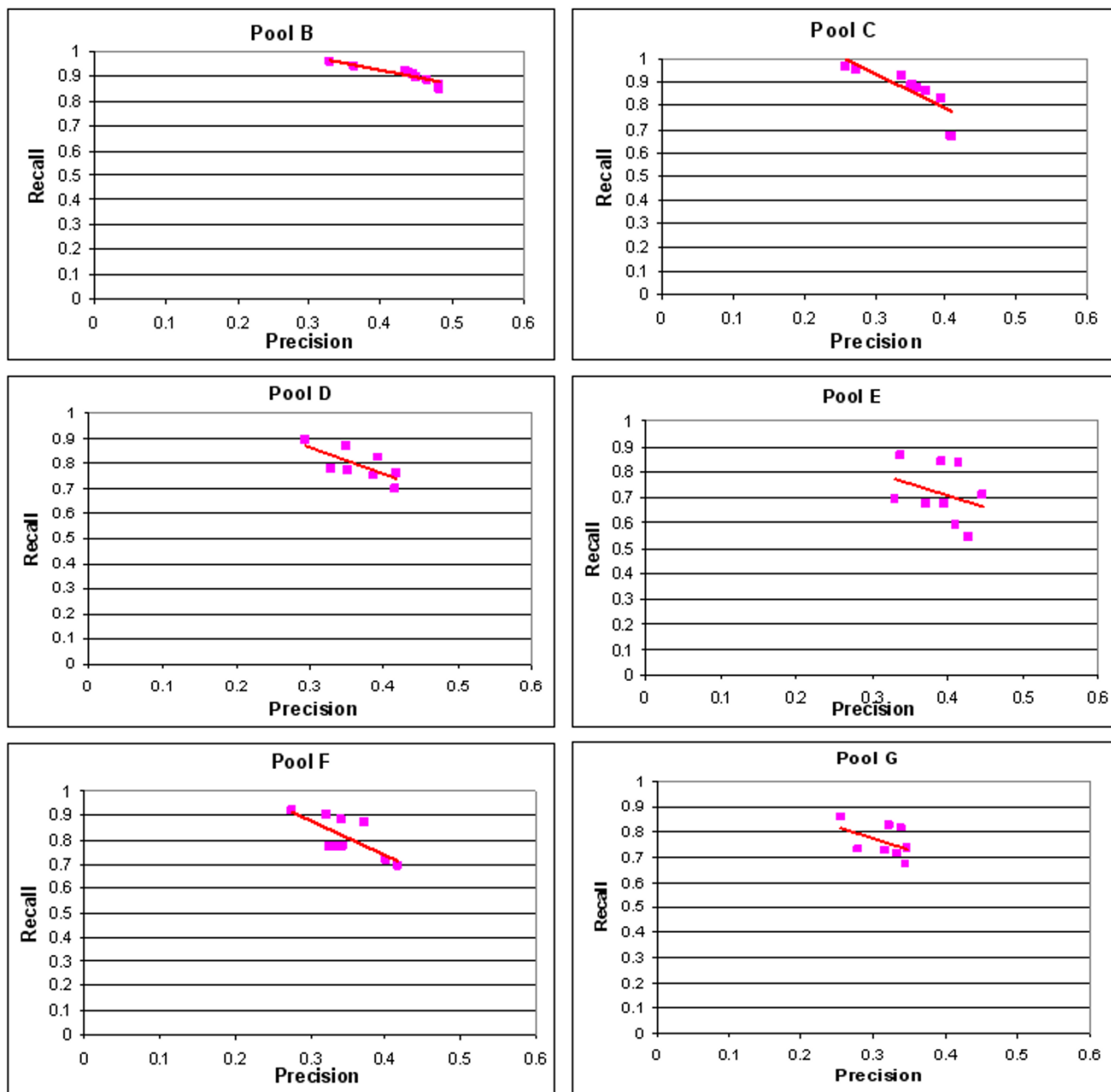
Figure 3
Precision vs. Recall for TagSmart. The six panels represent the mutant mixture pools B-F, respectively. For a wide range of thresholds, the precision and the recall from TagSmart are plotted, and a linear regression line is fitted.

Cincreasin experiment
To illustrate the power of TagSmart in a real biological investigation, we applied TagSmart on a tag array dataset [5]. This dataset records the tag array measurements of heterozygous deletion mutants under four experimental conditions, including rich medium (control), 100, 200, and 400 uM treatment of a chemical called Cincreasin. Cincreasin is a newly synthesized molecule that inhibits

the spindle checkpoint process by targeting Mps1 protein [5]. Additional file 6 summarizes this dataset. We restricted our analysis on 200 uM treatment data only.

TagSmart identified 53 mutants as sensitive to Cincreasin treatment (q-value = 1%, FC = 0.5, Additional file 7). Additional file 8 shows the fold changes of top 10 most sensitive mutants. Among theses mutants, Mps1 was

shown to be the direct target of Cincreasin [5]. Mps1 is a dual-specificity kinase required for spindle pole body duplication and spindle checkpoint function [12]. Cincreasin blocks the spindle checkpoint response to a lack of tension on mitotic chromosome by inhibiting Mps1. However, the other mutants besides Mps1 being sensitive to Cincreasin might indicate that there is an aftermath of chained molecular responses to the inhibition of Mps1.

To validate these findings, we re-tested all the 53 heterozygous mutants individually with Cincreasin treatment in colonial growth assay (patch test). Mutant strains were grown in colonies in the same control environment as described in [5] and in 200 uM and 400 uM Cincreasin treated environments. Three wild type colonies were grown under each environment as negative controls. Cin8 homozygous deletion strain was used as positive control (sensitive to Cincreasin treatment), because Cincreasin has been previously shown to cause mis-segregation of chromosomes in cin8 null cells (Figure 3 in reference [5]). In this test, 52 out of the 53 predicted mutants showed significant sensitivity to Cincreasin in this test, comparing to three wild type colonies (Additional file 7 and Additional file 9). This high validation rate demonstrates TagSmart is very resistant to false positive reports. We recognize that due to the limited amount of mutant colonies we could test, it is infeasible for this test to address the amount of false negative reports. The titration experiment described previously in this paper did address both false positive and false negative reports.

The validated mutants fell into three classes: i) mutants with lesions in known components of the spindle, ii) mutants in genes of known function which lack any described role in chromosome segregation, and iii) mutants in genes of unknown function. The first functional category is highly relevant to the function of Cincreasin, an inhibitor of buddy yeast spindle checkout. It is worth further investigation whether the second class reflects additional molecular targets of Cincreasin that lie outside the spindle checkpoint, or previously undiscovered roles in spindle function for this class of genes.

4. Conclusion

Tag microarray data has inspired various research, including identification of gene function [1,13], identification of drug targets [5,6], and evolution and genetic robustness [14]. TagSmart is an interactive online software tool for the analysis of tag microarray data. It is freely available for non-commercial use at [3]. Our future work is to expand TagSmart for integrated analysis with other genomics data, such as expression data and double deletion mutant data.

Availability and requirements

Project name: TagSmart

Project home page: http://biocomp.bioen.uiuc.edu/tagsmart

Operating systems: Platform independent

Programming language: C#, ASP.NET

Other requirements: None

License: None

Restrictions to use by non-academics: licence needed

Authors' contributions

SZ conceived the study, designed the analytical procedure, the software and the biological experiments. CK, SK and JH implemented the software. SZ, CK, SK and DX analyzed the data. RD performed the biological experiments. SZ coordinated the study and wrote the paper. All authors read and approved the final manuscript.

Additional material

Additional file 1
A generic gene-deletion cassette module. Supplementary figure 1
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-128-S1.doc]

Additional file 2
Summary of previous methods. Supplementary document
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-128-S2.doc]

Additional file 3
Screenshot of the preprocessor module. Supplementary figure 2
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-128-S3.doc]

Additional file 4
Screenshot of the computation module. Supplementary figure 3
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-128-S4.doc]

Additional file 5
An example text output from TagSmart server. Supplementary figure 4
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-128-S5.doc]

Additional file 6

Summary of the Cincreasin dataset. Supplementary table 1
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-128-S6.doc]

Additional file 7

Summary of test results by colonial assays. Supplementary table 2
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-128-S7.doc]

Additional file 8

Fold changes of the top 10 mutants sensitive to Cincreasin treatment. Supplementary figure 5
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-128-S8.doc]

Additional file 9

Colonial assays for testing S.cerevisiae mutants in the presence of DMSO (control) and Cincreasin (treatment). Supplementary figure 6
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-128-S9.doc]

Acknowledgements

References

1. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Davis RW, *et al.*: **Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis.** *Science* 1999, **285(5429):**901-906.
2. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kotter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M: **Functional profiling of the Saccharomyces cerevisiae genome.** *Nature* 2002, **418(6896):**387-391.
3. **DNA TAG3, Affymetrix, Santa Clara, CA.** .
4. Yuan DS, Pan X, Ooi SL, Peyser BD, Spencer FA, Irizarry RA, Boeke JD: **Improved microarray methods for profiling the Yeast Knockout strain collection.** *Nucleic Acids Res* 2005, **33(12):**e103.
5. Dorer RK, Zhong S, Tallarico JA, Wong WH, Mitchison TJ, Murray AW: **A small-molecule inhibitor of Mps1 blocks the spindle-checkpoint response to a lack of tension on mitotic chromosomes.** *Curr Biol* 2005, **15(11):**1070-1076.
6. Lum PY, Armour CD, Stepaniants SB, Cavet G, Wolf MK, Butler JS, Hinshaw JC, Garnier P, Prestwich GD, Leonardson A, Garrett-Engele P, Rush CM, Bard M, Schimmack G, Phillips JW, Roberts CJ, Shoemaker DD: **Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes.** *Cell* 2004, **116(1):**121-137.
7. Miller RGJ: **Beyond ANOVA: Basics of Applied Statistics.** Chapman & Hall; 1997.
8. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci U S A* 2003, **100(16):**9440-9445.
9. Storey JD, Taylor JE, Siegmund D: **Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach.** *Journal of the Royal Statistical Society, Series B* 2004, **66:**187-205.
10. Zhong S, Tian L, Li C, Storch KF, Wong WH: **Comparative analysis of gene sets in the Gene Ontology space under the multiple hypothesis testing framework.** *Proc IEEE Comput Syst Bioinform Conf* 2004:425-435.
11. Eason RG, Pourmand N, Tongprasit W, Herman ZS, Anthony K, Jejelowo O, Davis RW, Stolc V: **Characterization of synthetic DNA bar codes in Saccharomyces cerevisiae gene-deletion strains.** *Proc Natl Acad Sci U S A* 2004, **101(30):**11046-11051.
12. Lauze E, Stoelcker B, Luca FC, Weiss E, Schutz AR, Winey M: **Yeast spindle pole body duplication gene MPS1 encodes an essential dual specificity protein kinase.** *Embo J* 1995, **14(8):**1655-1663.
13. Giaever G, Flaherty P, Kumm J, Proctor M, Nislow C, Jaramillo DF, Chu AM, Jordan MI, Arkin AP, Davis RW: **Chemogenomic profiling: identifying the functional interactions of small molecules in yeast.** *Proc Natl Acad Sci U S A* 2004, **101(3):**793-798.
14. Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li WH: **Role of duplicate genes in genetic robustness against null mutations.** *Nature* 2003, **421(6918):**63-66.
15. Ooi SL, Shoemaker DD, Boeke JD: **DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray.** *Nat Genet* 2003, **35(3):**277-286.
16. Ooi SL, Shoemaker DD, Boeke JD: **A DNA microarray-based genetic screen for nonhomologous end-joining mutants in Saccharomyces cerevisiae.** *Science* 2001, **294(5551):**2552-2556.
17. Warren CD, Eckley DM, Lee MS, Hanna JS, Hughes A, Peyser B, Jie C, Irizarry R, Spencer FA: **S-phase checkpoint genes safeguard high-fidelity sister chromatid cohesion.** *Mol Biol Cell* 2004, **15(4):**1724-1735.
18. Lee MS, Spencer FA: **Bipolar orientation of chromosomes in Saccharomyces cerevisiae is monitored by Mad1 and Mad2, but not by Mad3.** *Proc Natl Acad Sci U S A* 2004, **101(29):**10655-10660.
19. Pan X, Yuan DS, Xiang D, Wang X, Sookhai-Mahadeo S, Bader JS, Hieter P, Spencer F, Boeke JD: **A robust toolkit for functional profiling of the yeast genome.** *Mol Cell* 2004, **16(3):**487-496.
20. Birrell GW, Giaever G, Chu AM, Davis RW, Brown JM: **A genome-wide screen in Saccharomyces cerevisiae for genes affecting UV radiation sensitivity.** *Proc Natl Acad Sci U S A* 2001, **98(22):**12608-12613.
21. Peyser BD, Irizarry RA, Tiffany CW, Chen O, Yuan DS, Boeke JD, Spencer FA: **Improved statistical analysis of budding yeast TAG microarrays revealed by defined spike-in pools.** *Nucleic Acids Res* 2005, **33(16):**e140.