BMC Genomics

BioMed Central

Research

Open Access

ILOOP – a web application for two-channel microarray interwoven loop design

Mehdi Pirooznia[1], Ping Gong[2], Jack Y Yang[3], Mary Qu Yang[4], Edward J Perkins[5] and Youping Deng*[1]

Address: [1]Department of Biological Sciences, University of Southern Mississippi, Hattiesburg, MS, 39406, USA, [2]SpecPro Inc., Vicksburg, MS, 39180, USA, [3]Harvard University, Harvard Medical School, Cambridge, MA 02140-0888, USA, [4]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20852, USA and [5]Environmental Laboratory, U.S. Army Engineer Research and Development Center, Vicksburg, MS, 39180, USA

Email: Mehdi Pirooznia - mehdi.pirooznia@usm.edu; Ping Gong - Ping.Gong@erdc.usace.army.mil; Jack Y Yang - jyang@bwh.harvard.edu; Mary Qu Yang - jyang@bwh.harvard.edu; Edward J Perkins - Edward.J.Perkins@erdc.usace.army.mil; Youping Deng* - Youping.Deng@usm.edu

* Corresponding author

from IEEE 7th International Conference on Bioinformatics and Bioengineering at Harvard Medical School
Boston, MA, USA. 14–17 October 2007

Published: 16 September 2008

BMC Genomics 2008, **9**(Suppl 2):S11 doi:10.1186/1471-2164-9-S2-S11

This article is available from: http://www.biomedcentral.com/1471-2164/9/S2/S11

© 2008 Pirooznia et al; licensee BioMed Central Ltd.
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Microarray technology is widely applied to address complex scientific questions. However, there remain fundamental issues on how to design experiments to ensure that the resulting data enables robust statistical analysis. Interwoven loop design has several advantages over other designs. However it suffers in the complexity of design. We have implemented an online web application which allows users to find optimal loop designs for two-color microarray experiments. Given a number of conditions (such as treatments or time points) and replicates, the application will find the best possible design of the experiment and output experimental parameters. It is freely available from http://mcbc.usm.edu/iloop.

Background

Microarray technology is now widely used to address complex scientific questions and for studies of gene interactions. However, it is associated with a number of technical challenges. The high cost of microarrays plus the complex logistical issues associated with microarray studies, often require that compromises must be made in the number of samples analyzed. Replication of data is a fundamental and widely appreciated principle of design that is often sacrificed. Microarray users now acknowledge that "replication" means different things in the microarray context [1-3]. "Replication" might refer to (A) Spotting genes multiple times per array; (B) Hybridizing multiple arrays to the same RNA samples; and (C) Using multiple individuals of a certain variety or type. Replication types (A) and (B) are sometimes referred to as technical replication while type (C) represents biological replication in the classical statistical sense. Biological replicates can assess biological variability, which is essential, for instance, to surmise that the mean expression of a gene differs in two populations [4]. Three layers can be considered in a design of a two-color microarray experiment. Experimental units are at the top layer of the experiment, two RNA samples obtained from each unit are in the middle layer,

and the arrangement of array elements on the slides would be placed at the bottom layer of the experiment [5].

Certain decisions as to how many microarray slides will be used and which mRNA samples will be hybridized to each slide must be made in preparation of mRNA samples before carrying out a microarray experiment [6,7]. Kerr and Churchill [2] and Glonek and Solomon [8] suggested efficient designs for some common microarray experiments. The most commonly used design is the reference design (Figure 1-A). In this design, each condition of interest is compared with samples taken from a standard reference. This design allows an indirect comparison between the conditions, because the reference is common to all of the arrays. In contrast, a loop design (Figure 1-B) compares two conditions via a chain of other conditions or multiple-pairwise (interwoven loop) fashion [9,10].

The computation of variance in a loop design depends on the design and the number of samples. In the loop design, each sample is compared directly with other samples in a multiple-pairwise (circular) way. Most studies on microarray design suggest that the loop design of microarray experiments is more efficient than the reference design [11-13]. This approach has stronger statistical power than

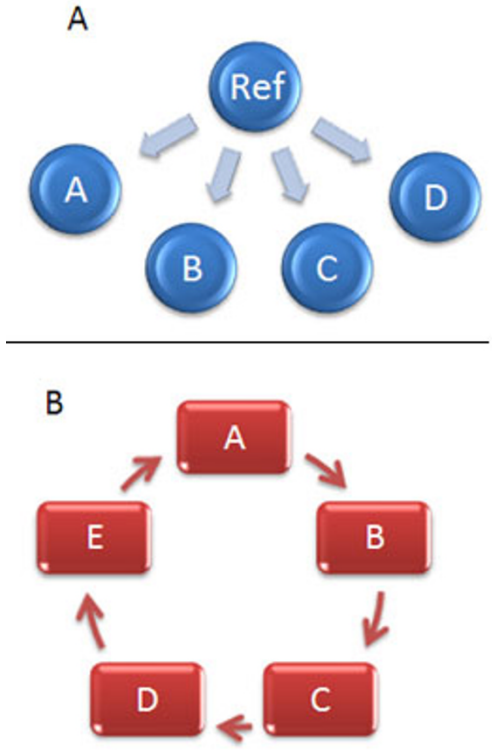


Figure 1
Combination of varieties with dyes for the reference (A) vs. loop design (B).

the reference design. Also, the entire dataset produced in a loop design is useful experimental information, while half the data produced in a reference design experiment is redundant.

In a cDNA microarray experiment, the foreground red and green intensities can be considered as *Rf* and *Gf* for each spot and the background intensities *Rb* and *Gb*. The background-corrected intensities will be *R* and *G* where *R* = *Rf*-*Rb* and *G* = *Gf*-*Gb*. M and A can be calculated as $M = \log R/G$ *and* $A = 1/2 \log RG$. It is convenient to use base 2 logarithms for *M* and *A* so that *M* is units of 2-fold change. On this scale, *M* = 0 represents equal expression, *M* = 1 represents a 2-fold change between the RNA samples, *M* = 2 represents a 4-fold change, and so on. If treatment A is on array 1 and treatment B is on array 2, the contrast A-B is estimated by M_{i1} - M_{i2} with variance $4\sigma_{i\varepsilon}^2$. With k replicates, the estimated contrast would have variance $4\sigma_{i\varepsilon}^2/\mathrm{k}$ [14-17].

In loop design, using the optimal weighting, the variance of the contrast between adjacent treatments is $\sigma_{i\varepsilon}^2 + \sigma_{ia}^2\sigma_{i\varepsilon}^2 / 2(\sigma_{ia}^2 + \sigma_{i\varepsilon}^2)$ while the variance of the contrast between diagonally opposite treatments is $\sigma_{i\varepsilon}^2 + \sigma_{ia}^2\sigma_{i\varepsilon}^2 / (\sigma_{ia}^2 + \sigma_{i\varepsilon}^2)$. Comparing these variances with variance of the contrast from a reference design with *K* replicates, $2\sigma_{i\varepsilon}^2 / \mathrm{k} + 2\sigma_{ia}^2\sigma_{i\varepsilon}^2 / \mathrm{k}(\sigma_{ia}^2 + \sigma_{i\varepsilon}^2)$, it is clear that both of these variances in loop design are smaller than the variance of the contrast from a reference design with the same number of conditions and arrays, primarily because there are two replicates per sample, rather than one [17]. However one disadvantage of this method is that ratios observed across different pairwise comparisons are not immediately comparable and visualizations are more difficult [18,19].

Kerr and Churchill [2] noticed that a loop design stops being optimal when there are more than eight conditions. Therefore it has been suggested that the optimal design could be a form of an interwoven loop design. Figure 2 shows an example of interwoven loop design for an experiment with nine conditions (or time points) and 18 array slides [2,9,10].

Wit et al. [20] have developed an optimization algorithm that searches for the loop design which minimizes the A-optimality criterion. This is in fact an interwoven design. The interwoven design guarantees that each condition is measured equally often by either dye [21]. The Wit et al.

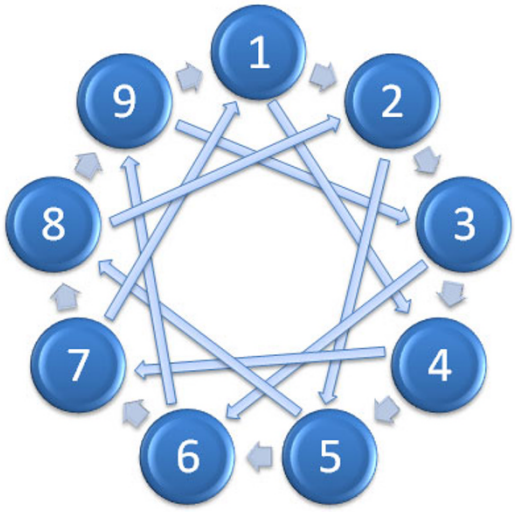


Figure 2
An example interwoven loop design with 18 arrays and 9 conditions.

optimization algorithm in fact allows one to input the number of conditions and the number of arrays one can afford to hybridize.

Currently biologists take a considerable amount of time to develop loop designs manually and the final design may not be optimized. To date there is no available tool for biologists to automatically design and visualize the interwoven loop for a microarray experiment. Development of such a tool will permit biologist to quickly generate different array hybridization loops, compare the cost and experiment design and efficiently design microarray studies so that robust statistical conclusion can be made.

Implementation

Here we calculate the most optimal loop by considering the number of replicates and conditions. The main point is to generate an optimal number of arrays based on combination of conditions and replicates for two-dye microarray experiment.

The web application has been developed using PHP language on an open source Apache web server. It is freely available from http://mcbc.usm.edu/iloop (Figure 3).

Given a number of conditions (treatments/timepoints) and replicates, the program generates the optimal interwoven loop design. The start menu has two drop boxes, one for "Number of Conditions" and another for "Number of Replicates". By selecting the number of con-

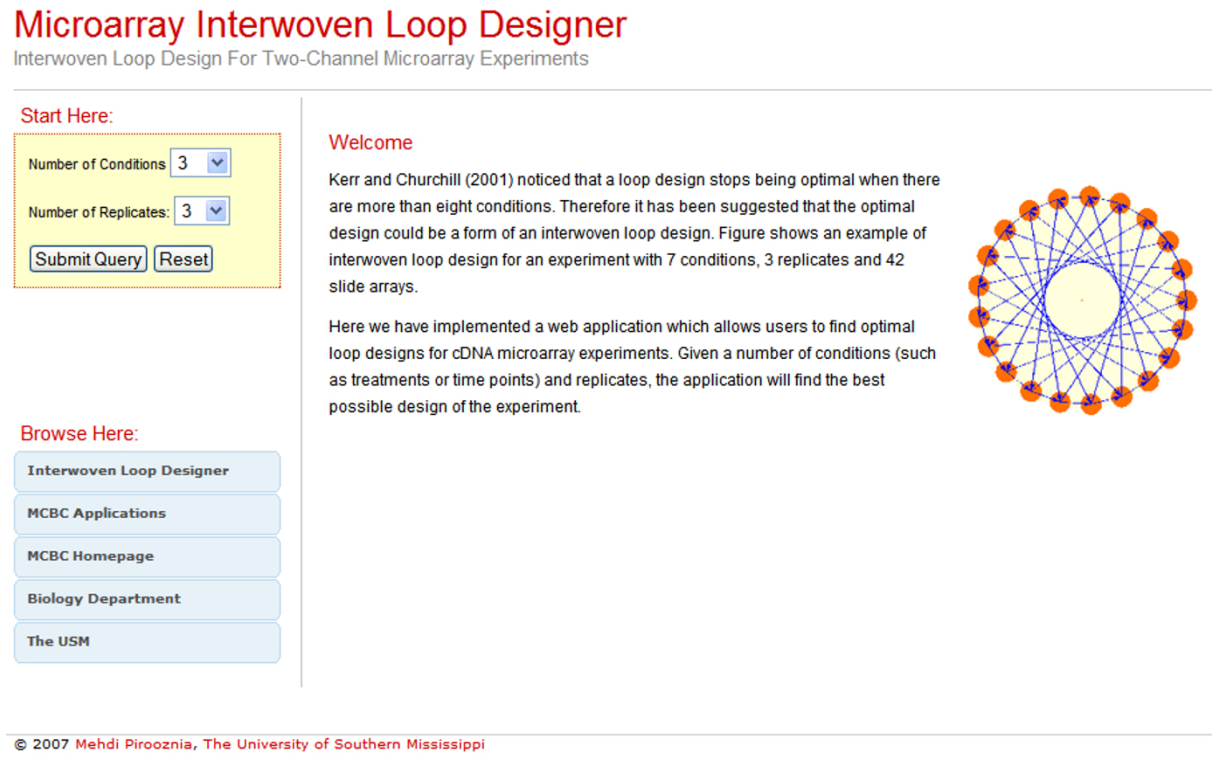


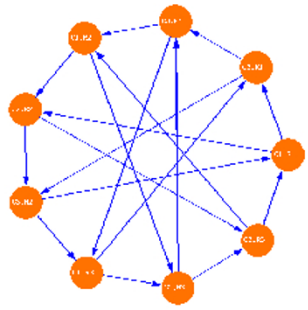
Figure 3
The web application screenshot.

Results

Total number of required samples: = 9 (Show Samples)

Total number of required arrays: = 18

Array Number	Red Channel	Green Channel
1	C1.R1	C2.R1
2	C2.R1	C3.R1
3	C3.R1	C1.R2
4	C1.R2	C2.R2
5	C2.R2	C3.R2
6	C3.R2	C1.R3
7	C1.R3	C2.R3
8	C2.R3	C3.R3
9	C3.R3	C1.R1
10	C1.R1	C2.R2
11	C2.R1	C3.R2
12	C3.R1	C1.R3
13	C1.R2	C2.R3
14	C2.R2	C3.R3
15	C3.R2	C1.R1
16	C1.R3	C2.R1
17	C2.R3	C3.R1
18	C3.R3	C1.R2



Click to enlarge picture

close X

Samples:

C1.R1
C2.R1
C3.R1
C1.R2
C2.R2
C3.R2
C1.R3
C2.R3
C3.R3

Figure 4
A screenshot of optimal interwoven loop table and graph.

ditions and replicates, the application generates an experiment design matrix table (Figure 4 and 5). The following pseudo code represents the algorithm used for array construction from the experiment design matrix table:

$$
\begin{aligned}
&for\ (i = 0; i < t - 1; i++)\{\\
&array\ (sample[i], sample[i+1])\\
&\}\\
&array\ (sample[t-1], sample[0])\\
&for\ (i = 0; i < t; i++)\{\\
&array\ (sample[i], sample[f \bmod (i + 1 + c, t)])\\
&\}
\end{aligned}
$$

Where c is the number of conditions and t is the total number of required sample (conditions × replicates).

In Figure 5, table cells represent a sample of the replicate with the corresponding condition. Here all cells are editable and user is able to change the condition and/or replicate's name. The application generates a visualization of optimal interwoven loop table and graph (Figure 4). The total number of required samples as well as the total number of required arrays is calculated. Any two spots connected with arrow in the graph in Figure 4 represent an array combined from red to green channel.

As an example figure 4 shows a screenshot of an optimal interwoven loop table and graph. In this case with 3 con-

Experiment Design Matrix

Experiment Design Matrix Table: Click on each cell to modify the text then click on "Generate Loop".

	Condition1	Condition2	Condition3
Replicate1	C1.R1	C2.R1	C3.R1
Replicate2	C1.R2	C2.R2	C3.R2
Replicate3	C1.R3	C2.R3	C3.R3

Keep Default Generate Loop

Figure 5
The experiment design matrix table.

ditions and 4 biological replicates, the application calculates the total number of required samples, 12, and the total number of required arrays would be 24. A similar experiment with the reference design would require 24 arrays. However the loop design creates 4 technical replicates per sample and samples are always hybridized to different samples (biological replicates).

Application

We evaluated the utility of the design application using a microarray designed to study the effect of chemical toxicity on earthworm [22]. Earthworms were exposed to three different concentration of TNT (2,4,6-trinitrotoluene). We used five biological replicates for each exposure. The application produced an optimal design of 40 arrays derived from 20 cDNA probes in accordance with an interwoven loop scheme as shown in Figure 6. cDNA samples from each biological replicate were labeled twice with a green channel fluorescence dye (Cy3) and twice with a red fluorescence channel dye (Alexa 647).

The hybridization experimental design tool proved simple to use and facilitated execution of the complex sample pairing required by this approach. A similar reference design would require 80 arrays to achieve 5 biological replicates over 4 conditions with 4 technical replicates compared to 40 required arrays in the optimal loop design in this experiment.

Discussion

The main significance of this paper is introduction of a web application that implements loop design for microarray experiment. To date no online application has been available to achieve this goal. Such designs should be analyzed by treating the arrays as blocks of size 2 and analyzing the channels as individual observations.

Our web application will allow scientists to design and graph the optimal interwoven loop faster. They can quickly select the number of conditions and replicates and weight the number of samples and arrays in order to minimize the cost and complexity of the experiment as well as maximizing the efficiency of the experiment.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MP and YD initiated the study. MP and YD designed and implemented the application, web server configuration and web application programming. PG and EJP provided the data. MP, YD, PG, and EJP drafted the original manuscript. YD, PG, JY, MY and EJP coordinated and directed

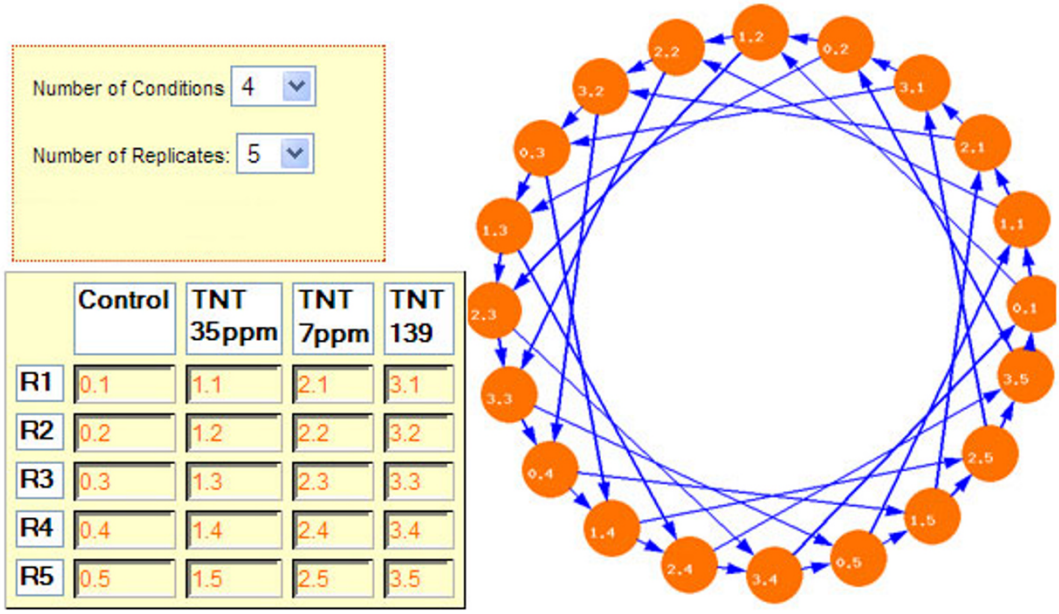


Figure 6
An interwoven loop hybridization schemes for 4 treatments with 5 independent biological replicates. Circles represent treatment samples. Sample code: 0.x = replicate × of solvent control worms; 1.x = replicate × of 10.6 mg TNT/kg soil treated worms; 2.x = replicate × of 2 mg TNT/kg soil treated worms; 3.x = replicate × of 38.7 mg TNT/kg soil treated worms; x = 1–5. Arrows represent array hybridizations between respective samples where the arrowhead indicates Alexa 647 dye labeling and the base of arrows indicate Cy3 dye labeling.

the project. All authors have read and approved the final manuscript.

Acknowledgements

This work was supported by the Mississippi Computational Biology Consortium (NSF EPS-0556308), Army Environmental Quality Program of the US Army Corps of Engineers (under contract #W912HZ-05-P-0145) and the Mississippi Functional Genomics Network (DHHS/NIH/NCRR Grant# 2P20RR016476-04). Permission to publish this information was granted by the Chief of Engineers.

This article has been published as part of *BMC Genomics* Volume 9 Supplement 2, 2008: IEEE 7th International Conference on Bioinformatics and Bioengineering at Harvard Medical School. The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2164/9?issue=S2

References

1. Kerr MK, Martin M, Churchill GA: **Analysis of variance for gene expression microarray data.** *J Comput Biol* 2000, **7(6):**819-837.
2. Kerr MK, Churchill GA: **Experimental design for gene expression microarrays.** *Biostatistics* 2001, **2(2):**183-201.
3. Yang YH, Buckley MJ, Speed TP: **Analysis of cDNA microarray images.** *Brief Bioinform* 2001, **2(4):**341-349.
4. Garosi P, De Filippo C, van Erk M, Rocca-Serra P, Sansone SA, Elliott R: **Defining best practice for microarray analyses in nutrigenomic studies.** *Br J Nutr* 2005, **93(4):**425-432.
5. Churchill GA: **Fundamentals of experimental design for cDNA microarrays.** *Nat Genet* 2002, **32(Suppl):**490-495.
6. Wit E, McClure J: **Statistical adjustment of signal censoring in gene expression experiments.** *Bioinformatics* 2003, **19(9):**1055-1060.
7. Peng X, Wood CL, Blalock EM, Chen KC, Landfield PW, Stromberg AJ: **Statistical implications of pooling RNA samples for microarray experiments.** *BMC Bioinformatics* 2003, **4:**26.
8. Glonek GF, Solomon PJ: **Factorial and time course designs for cDNA microarray experiments.** *Biostatistics* 2004, **5(1):**89-111.
9. Vinciotti V, Khanin R, D'Alimonte D, Liu X, Cattini N, Hotchkiss G, Bucca G, de Jesus O, Rasaiyaah J, Smith CP, *et al.*: **An experimental evaluation of a loop versus a reference design for two-channel microarrays.** *Bioinformatics* 2005, **21(4):**492-501.
10. Kerr MK, Churchill GA: **Statistical design and the analysis of gene expression microarray data.** *Genet Res* 2001, **77(2):**123-128.
11. Oleksiak MF, Churchill GA, Crawford DL: **Variation in gene expression within and among natural populations.** *Nat Genet* 2002, **32(2):**261-266.
12. Rosa GJ, de Leon N, Rosa AJ: **Review of microarray experimental design strategies for genetical genomics studies.** *Physiol Genomics* 2006, **28(1):**15-23.
13. Bueno Filho JS, Gilmour SG, Rosa GJ: **Design of microarray experiments for genetical genomics studies.** *Genetics* 2006, **174(2):**945-957.
14. Tempelman RJ: **Assessing statistical precision, power, and robustness of alternative experimental designs for two color microarray platforms based on mixed effects models.** *Vet Immunol Immunopathol* 2005, **105(3–4):**175-186.
15. Chen YA, McKillen DJ, Wu S, Jenny MJ, Chapman R, Gross PS, Warr GW, Almeida JS: **Optimal cDNA microarray design using expressed sequence tags for organisms with limited genomic information.** *BMC Bioinformatics* 2004, **5:**191.
16. McKillen DJ, Chen YA, Chen C, Jenny MJ, Trent HF 3rd, Robalino J, McLean DC Jr, Gross PS, Chapman RW, Warr GW, *et al.*: **Marine genomics: a clearing-house for genomic and transcriptomic data of marine organisms.** *BMC Genomics* 2005, **6(1):**34.
17. Altman NS, Hua J: **Extending the loop design for two-channel microarray experiments.** *Genet Res* 2006, **88(3):**153-163.
18. Bolstad BM, Collin F, Simpson KM, Irizarry RA, Speed TP: **Experimental design and low-level analysis of microarray data.** *Int Rev Neurobiol* 2004, **60:**25-58.
19. Landgrebe J, Bretz F, Brunner E: **Efficient two-sample designs for microarray experiments with biological replications.** In *Silico Biol* 2004, **4(4):**461-470.
20. Khanin R, Wit E: **Design of large time-course microarray experiments with two channels.** *Appl Bioinformatics* 2005, **4(4):**253-261.
21. Cui X, Churchill GA: **Statistical tests for differential expression in cDNA microarray experiments.** *Genome Biol* 2003, **4(4):**210.
22. Gong P, Guan X, Inouye LS, Pirooznia M, Indest KJ, Athow RS, Deng Y, Perkins EJ: **Toxicogenomic analysis provides new insights into molecular mechanisms of the sublethal toxicity of 2,4,6-trinitrotoluene in Eisenia fetida.** *Environ Sci Technol* 2007, **41(23):**8195-8202.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

BioMedcentral