

Poster presentation

Open Access

## A systematic strategy for the discovery of candidate genes responsible for phenotypic variation

Paul Fisher\*<sup>1</sup>, Cornelia Hedeler<sup>1</sup>, Katherine Wolstencroft<sup>1</sup>, Helen Hulme<sup>1</sup>, Harry Noyes<sup>2</sup>, Stephen Kemp<sup>2</sup>, Robert Stevens<sup>1</sup> and Andrew Brass<sup>1,3</sup>

Address: <sup>1</sup>School of Computer Science, Kilburn Building, University of Manchester, Oxford Road, Manchester, M13 9PL, UK, <sup>2</sup>School of Biological Sciences, Biosciences Building, University of Liverpool, Crown Street, Liverpool, L69 7ZB, UK and <sup>3</sup>Faculty of Life Science, Michael Smith Building, University of Manchester, Oxford Road, Manchester, M13 9PT, UK

Email: Paul Fisher\* - fisherp@cs.manchester.ac.uk

\* Corresponding author

from Third International Society for Computational Biology (ISCB) Student Council Symposium at the Fifteenth Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) Vienna, Austria. 21 July 2007

Published: 20 November 2007

BMC Bioinformatics 2007, 8(Suppl 8):P7 doi:10.1186/1471-2105-8-S8-P7

This abstract is available from: <http://www.biomedcentral.com/1471-2105/8/S8/P7>

© 2007 Fisher et al; licensee BioMed Central Ltd.

### Introduction

The use of Quantitative Trait Loci (QTL) data is increasingly used to aid in the discovery of candidate genes involved in phenotypic variation. Tens to hundreds of genes, however, may lie within even well defined QTL. It is therefore vital that the identification, selection and functional testing of candidate Quantitative Trait genes (QTg) are carried out systematically, and without bias [1]. With the advent of microarrays, researchers are able to directly examine the expression of all genes on a genome wide scale, including those underlying QTL regions.

The scale of data being generated by such high-throughput experiments has led some investigators to follow a hypothesis-driven approach [2]. Although these techniques for candidate gene identification are valid, they run the risk of overlooking genes that have less obvious associations with the phenotype. By making selections based on prior assumptions of what processes may be involved, the genes that may actually be involved in the phenotype can be overlooked. A further complication is that the use of *ad hoc* methods for candidate gene identification are inherently difficult to replicate and are compounded by poor documentation of the methods used to generate and capture the data from such investigations in published literature.

With an ever increasing number of institutes offering programmatic access to their resources in the form of web services, however, experiments previously conducted manually can now be replaced by automated experiments, capable of processing a far greater volume of data. By reconstructing the original investigation methods in the form of workflows, we are now able to pass data directly from one service to the next. This enables us to process the data in a much more systematic, un-biased, and explicit manner.

### Methods

We propose a data-driven methodology that identifies the known pathways that intersect a QTL and those derived from a set of differentially expressed genes from a microarray study. This methodology is implemented systematically through the use of web services and workflows. For the purpose of implementing this systematic pathway-driven approach, we have chosen to use the Taverna workflow [3].

### Results and Discussion

Preliminary studies into the modes of resistance to African Trypanosomiasis were carried out for the mouse model organism. These studies illustrated how the large-scale analysis of microarray gene expression and QTL data,

investigated at the level of biological pathways, enables links between genotype and phenotype to be successfully established [4]. This approach was implemented systematically through the use of explicitly defined workflows.

## References

1. Glazier A, Nadeau J, Aitman T: **Finding genes that underlie complex traits.** *Science* 2002, **298**:2345-2349.
2. Kell D, Oliver S: **Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era.** *Bioessays* 2004, **26**:99-105.
3. Oinn T, Addis M, Ferris J, Marvin D, Senger M, Greenwood M, Carver T, Glover K, Pocock M, Wipat A, et al.: **Taverna: a tool for the composition and enactment of bioinformatics workflows.** *Bioinformatics* 2004, **20**:3045-3054.
4. Fisher P, Hedeler C, Wolstencroft K, Hulme H, Noyes H, Kemp S, Stevens R, Brass A: **A Systematic Strategy for Large-Scale Analysis of Genotype-Phenotype Correlations: Identification of candidate genes involved in African Trypanosomiasis.** *Nucleic Acids Research* 2007 in press.

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](http://www.biomedcentral.com/info/publishing_adv.asp)

