

Development and Evaluation of Web Applications for Investigating Candidate Genes in Rat Models of Complex Diseases

Akademisk avhandling

som för avläggande av filosofie doktorsexamen vid Göteborgs universitet
kommer att offentligas i föreläsningssalen Gösta Sandels, Medicinaregatan 9B
Fredagen den 12 December 2008, kl. 09.00

av

Lars Andersson

Institutionen för Cell- och Molekylärbiologi – Genetik, Göteborgs universitet

Fakultetsopponent: Professor Tomas Olsson, Institutionen för klinisk neurovetenskap,
Karolinska Institutet

Avhandlingen baseras på följande delarbeten:

- I. **Andersson L**, Petersen G, Johnson P, Ståhl F: A web tool for finding gene candidates associated with experimentally induced arthritis in the rat. *Arthritis Res Ther* 2005, **7**(3):R485-492.
- II. **Andersson L**, Petersen G, Ståhl F: Ranking candidate genes in rat models of type 2 diabetes. *Submitted 2008*
- III. **Andersson L**, Ståhl F: CGC-RefLink, a tool for finding functional gene pairs related to experimentally induced arthritis in rats. *Submitted 2008*
- IV. **Andersson L**, Ståhl F: Distribution of candidate genes for experimentally induced arthritis in rats. *Submitted 2008*

Development and Evaluation of Web Applications for Investigating Candidate Genes in Rat Models of Complex Diseases

Lars Andersson, Department of Cell and Molecular Biology – Genetics
University of Gothenburg, Box 462, SE 405 30 Göteborg, Sweden

Abstract

Many human diseases, such as rheumatoid arthritis and type 2 diabetes mellitus, have a very complex development, depending on both environmental and multiple genetic factors. By crossing inbred rat strains susceptible to a genetic disorder with strains resistant to the same disorder, genomic regions associated with the disease can be identified, so called quantitative trait loci (QTLs). A QTL region is often rather large, sometimes covering hundreds of genes. To help selecting the most likely causative candidate genes from such QTLs in rat, we have created a publicly available application, called candidate gene capture (CGC).

The CGC application was primarily applied on experimentally induced arthritis QTLs in rat. CGC uses an array of keywords compared to the reference term “arthritis”. For each keyword, this results in a keyword score that reflects the percentage of PubMed abstracts containing the keyword that also contain the reference term. OMIM records for human genes localized to human regions homologous to rat QTL regions, are scanned for all keywords. The sum of all matching keyword scores is used to rank candidate genes within each QTL. When evaluated, the CGC application is able to rank candidate genes for arthritis-associated QTLs in a manner very similar to what is done manually.

In a second application, CGC was applied on non-insulin dependant diabetes mellitus QTLs in rat. Here, the number of included keywords was dramatically increased. In the CGC-Diabetes application the user can choose from 25 different reference terms, to which the keywords are compared. The reference terms are selected to represent sub-phenotypes of diabetes so that the user can choose which distinct characteristics to analyze. A “phylogenetic tree” was created to give an overview of how much the gene rankings would differ when different reference terms are used. Just like the CGC-Arthritis application, the CGC-Diabetes application proves to be successful in ranking candidate genes in a manner very similar to what is done manually.

In an extended version of the CGC-Arthritis application, CGC-RefLink, candidate genes identified for a QTL using CGC can be functionally connected to candidate genes in other QTLs via hyperlinks in the respective OMIM records. In a comparative study, CGC-RefLink was applied on arthritis QTLs from two distinct rat crosses. In this way, we were able to find functional connections between genes in QTLs from the two crosses that could contribute to a similar arthritis phenotype.

Finally, using the CGC-Arthritis and the CGC-RefLink applications, we analyzed the localization of candidate genes in the rat genome. We concluded that i) certain QTLs from two different rat crosses harbor a number of genes involved in similar functions, which could be associated to arthritis and ii) candidate genes are randomly distributed between QTL and non-QTL regions.

Keywords: rat, complex disease, rheumatoid arthritis, type 2 diabetes mellitus, QTL, candidate genes, web application, text mining.

ISBN 978-91-628-7647-0