

VisGenome: visualisation of single and comparative genome representations

Joanna Jakubowska^{a,*}, Ela Hunt^b, Matthew Chalmers^a, Martin McBride^c
and Anna F. Dominiczak^c

^aDepartment of Computing Science, University of Glasgow, G12 8QQ, Scotland,

^bDepartment of Computer Science, ETH Zurich, 8092 Zurich, Switzerland,

^cBHF Glasgow Cardiovascular Research Centre, University of Glasgow, G12 8TA, Scotland

ABSTRACT

Summary: VisGenome visualises single and comparative representations for the rat, the mouse, and the human chromosomes at different levels of detail. The tool offers smooth zooming and panning which is more flexible than seen in other browsers. It presents information available in Ensembl for single chromosomes, as well as homologies (orthologue predictions including ortholog one2one, apparent ortholog one2one, ortholog many2many) for any two chromosomes from different species. The application can query supporting data from Ensembl by invoking a link in a browser.

Availability: <http://www.dcs.gla.ac.uk/~asia/VisGenome>

Contact: asia@dcs.gla.ac.uk

1 INTRODUCTION

VisGenome was designed to match the visualisation needs of the BHF Blood Pressure Group at the University of Glasgow which uses a rat model of cardiovascular disease. We analyse genomic data at different levels of detail, to dissect rat and human QTLs (McBride et al., 2006) in the search for candidate disease genes. QTLs are analysed in three species: the rat, the human and the mouse, and the work uses genotyping, and micro array and proteomics techniques. VisGenome supports QTL analysis by showing QTLs and the genes within each QTL in two species in one display, along with the supporting experimental data. It also shows data for a single chromosome in one display and supports zooming at an arbitrary level of detail. Current software release connects to Ensembl (Hubbard et al., 2007) and can be used as an alternative viewer for a subset of that data. Extensions are being tested and an ongoing user study will guide further development.

2 FEATURES

VisGenome loads QTLs, genes, micro array probes, bands, and markers, and displays pairs of homologies (orthologue predictions) from Ensembl. It welcomes the user with a view of all rat, mouse and human chromosomes, see Fig. 1. In the single chromosome representation, after choosing a chromosome of interest by clicking, a new view with detailed data about the chromosome is created. In the comparative representation the user clicks on two chromosomes from different species and a new view representing homologies

between the chosen chromosomes is created. After choosing a chromosome, the users can manipulate the view by mouse and keyboard interaction.

Navigation. VisGenome offers mouse and keyboard interaction. The users choose the representation by selecting a menu item or by typing Ctrl+S (single) or Ctrl+C (comparative). A choice is offered for marking the chromosome region of interest. The users can drag or enlarge the box enclosing the region marked on the chromosome or enter the region coordinates in the top panel.

Zooming and Panning. We offer smooth zooming which supports the visual exploration of the chromosome space. Zooming and panning uses Piccolo (Bederson et al., 2004). The zooming technique allows the users to keep an area of interest in focus during interaction with the data. Zooming is manipulated by the right mouse button by moving it to the right (zoom in) or to the left (zoom out). Panning uses the left mouse button. Both interactions are easy to use and the users quickly become familiar with them.

Marking a region of interest. The users can choose a chromosome region of interest. Then, the main view shows only the data for this region and the users manipulate only the data in the selected area.

Additional Information. VisGenome supports the display of additional information about presented data in an info panel.

Supporting data. Access to Ensembl is provided via clicking on a feature of interest which invokes Ensembl web pages in the user's browser.

3 IMPLEMENTATION

VisGenome was written and tested on Windows XP with Java 1.5, on a 2.39 GHz Pentium 4 with 512 MB RAM. VisGenome connects via JDBC to Ensembl databases for the rat, the mouse, and the human. During the tests we visualised chromosome bands, micro array probe sets, QTLs and genes. Eclipse, www.eclipse.org, was used as a development environment. The application is packaged as a jar file. The user has to install Java prior to invoking the code. 1280 x 1024 screen resolution was used for testing.

4 DISCUSSION AND CONCLUSIONS

Visualisation of genome comparisons is an important research tool in biology and medicine. None of the tools we studied (Jakubowska et al., 2006) fulfilled the requirement of showing in detail and overview all the data needed for our work in one display, and allowed us to study QTLs and other relevant data. VisGenome extends

*to whom correspondence should be addressed

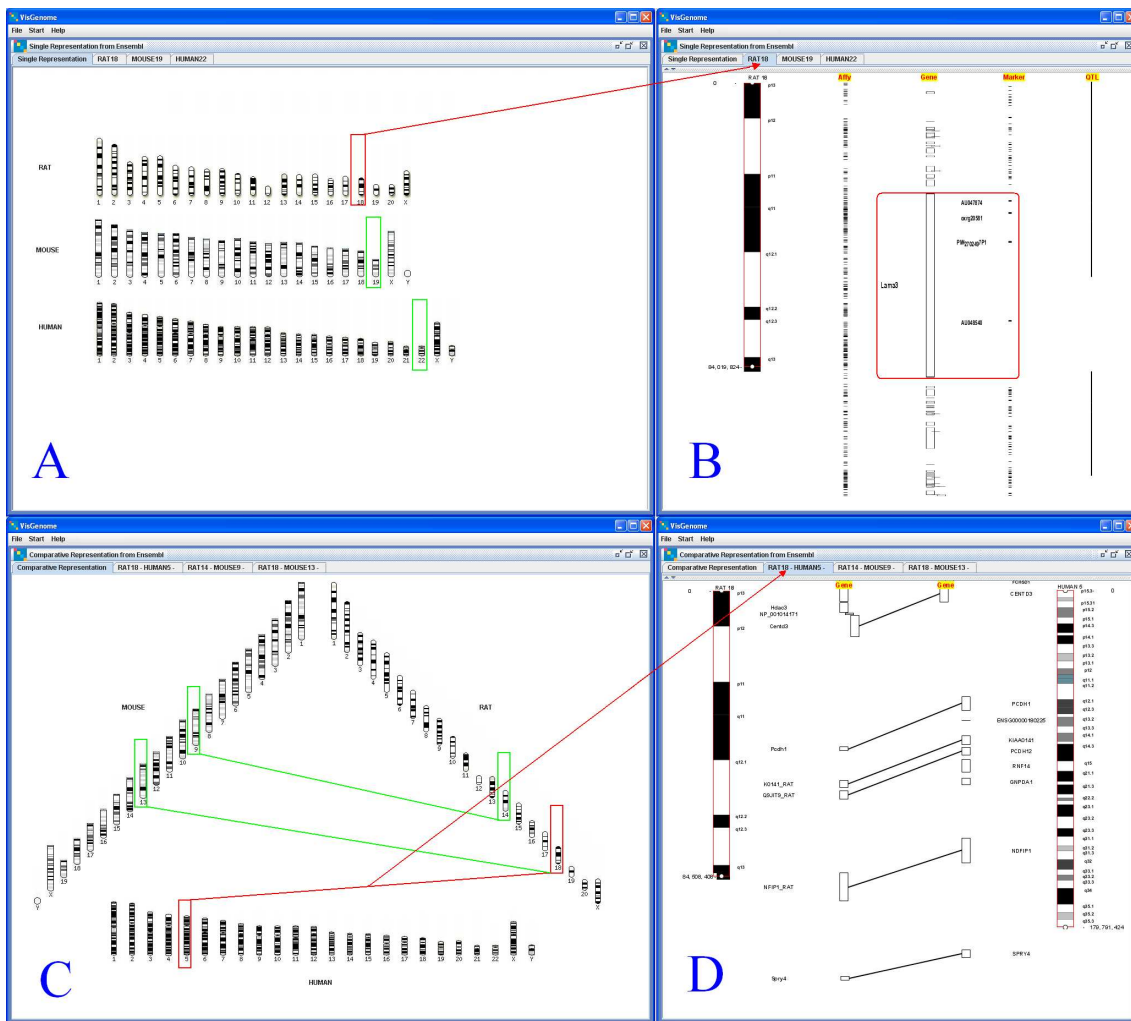


Fig. 1. A and B - VisGenome, single chromosome view. C and D - comparative view. A and C: chromosomes from the mouse, the rat and the human. B: an overview and detail for the rat chromosome 18. D: an overview and detail for the rat chromosome 18 and the human chromosome 5.

SyntenyVista (Hunt et al., 2004) with a single genome display and presents homologies alongside QTLs and micro array probes.

Our plans are as follows. We are conducting a user study comparing VisGenome and Ensembl, to identify future improvements and assess the usability of our solution. We will broaden the application of cartoon scaling which is not part of the current release, and requires algorithmic extensions. We will support the import of additional data from external files or web services, and local caching. We will add data import from DAS servers (Dowell et al., 2001). We are preparing a follow-on release supporting user-driven choice of colours and data sources. The data will be cached, so that the tool is usable even if access to Ensembl is slow, and the user will not have to re-import the same data. An intelligent caching solution is being tested.

ACKNOWLEDGEMENT

This work was funded by the BHF Programme (PG 02012) and Chair (CH98001) Grants and the Wellcome Trust Cardiovascular

Functional Genomics (WT066778/Z/01/Z) all to A.F.D., the MRC (J.J.), and by the MRC and an EU Marie-Curie Fellowship (E.H.).

REFERENCES

Hunt, E., Hanlon, N., Leader, D., Bryce, H., and Dominiczak, A. F. (2004) The Visual Language of Synteny. *OMICS*, **8**(4):289–305.
 Jakubowska, J., Hunt, E. and Chalmers, M. J. (2006) Granularity of genomics data in genome visualisation. *Dept of Comp. Sci., University of Glasgow*, Tech. Report: TR-2006-221, <http://www.dcs.gla.ac.uk/publications/PAPERS/8212/AsiaElaMatthewCHI06.pdf>.
 Hubbard, T. J. P. et al. (2007) Ensembl 2007. *Nucleic Acids Res.* **35**, Database issue: D610-D617.
 Bederson, B. B., Grosjean, J., and Meyer, J. (2004) Toolkit Design for Interactive Structured Graphics. *IEEE Transactions on Software Engineering*, **30** (8):535–546.
 McBride, M. W., Graham, D., Delles, C., Dominiczak, A. F. (2006) Functional genomics in hypertension *Curr Opin Nephrol Hypertens* 2006 Mar;**15**(2):145-51.
 Dowell, R. D. et al. (2001) The Distributed Annotation System. *BMC Bioinformatics*, **2**:7.