

# LinkageView: a powerful graphical tool for integrating statistical data with Ensembl

Judith E. Stenger<sup>1,2,3</sup>, Hong Xu<sup>1,2</sup>, Carol Haynes<sup>1,2</sup>, Elizabeth R. Hauser<sup>1,2</sup>, Margaret A. Pericak-Vance<sup>1,2</sup>, Pascal J. Goldschmidt-Clermont<sup>2</sup> and Jeffery M. Vance<sup>1,2</sup>

**Keywords:** bioinformatics, linkage analysis, genome annotation, database integration, complex disease susceptibility genes.

## 1 Introduction.

To facilitate efficient prioritization of candidate disease susceptibility genes for association analysis, increasingly comprehensive tools are essential to smoothly integrate vast quantities of disparate private and public data generated by genomic screens, association studies, whole genome sequencing and annotation data and ancillary biological databases. For this purpose, we have developed “LinkageView”, a tool that enables statistical data to be viewed in the context of Ensembl’s [1] Contig- and Cyto- View pages. Drawn under the ideogram, LinkageView illustrates lod score values for markers from a particular study plotted against the position that is determined in base pairs by mapping the primers to the chromosome using electronic PCR (e-PCR)[2]. The graphs are transparently integrated into the Contig- and Cyto-View pages and feature a sliding selection box to display a region of interest enlarged in the Overview and the Detailed View. Genomic features mapping to regions with evidence of linkage are accentuated when LinkageView is used with the Distributed Annotation System (DAS) [3] to display supplemental laboratory information as tracks containing information such as differentially expressed genes pertaining to the disease under investigation. LinkageView is a powerful visual feature that enhances the use of Ensembl as a resource for integrative genomic-based approaches for identifying candidate disease susceptibility genes. We are currently developing a panel to display association data in a region of interest.

## 2 Software and files.

Linkage data is traditionally displayed as a graph in which lod scores are plotted along the ordinate against the genetic location (in cM) of the markers used in the genetic screen and any subsequent analysis. To integrate such data with the physical genome sequence, it is imperative that the abscissa be expressed in base pairs so that the position of markers along the abscissa correctly align with, and strictly correspond to, the horizontal illustration of the ideogram that is displayed immediately above the linkage study graph in the Ensembl Contig View (Figure 1, second panel).

First we downloaded the human genome assembly (NCBI build 34) from the UCSC genome site [4] and the most recent version of the Ensembl annotation system. Then, we mapped the markers used as probes to the actual sequence of the human genome. We used the NCBI UniSTS database (<ftp://ftp.ncbi.nlm.nih.gov/repository/UniSTS/>) as the source of marker probe sequence information to convert the map units into base pair coordinates by locating the position of the STSs in human

---

<sup>1</sup> Center for Human Genetics, Duke University Medical Center, P.O. Box 3445, Durham, North Carolina 27710, USA. E-mail: [center@chg.duhs.duke.edu](mailto:center@chg.duhs.duke.edu)

<sup>2</sup> Department of Medicine, Duke University Medical Center, P.O.Box 3703, Durham, North Carolina 27710, USA. E-mail: [pascal.goldschmidt@duke.edu](mailto:pascal.goldschmidt@duke.edu).

<sup>3</sup> Presenter and corresponding author, E-mail: [jstenger@chg.duhs.duke.edu](mailto:jstenger@chg.duhs.duke.edu)

genome build 34 using e-PCR [2] and finally populated the statistical results table in the database.

We developed four BioPerl modules to draw the linkage plot in Contig View using the data uploaded in the DAS server for a particular study. The Linkage.pm module is a constructor for the linkage object, essentially a table encapsulating the record for a single linkage point. This allows linkage data to be added to the database. Next, LinkageAdaptor.pm, provides the functionality for accessing linkage data from the DAS database. A “slice” object is created which defines the region from which linkage data will be retrieved and returns the records from the linkage objects that are mapped to the chromosome between the coordinates of the slice. The GlyphSet::lodplot.pm module contains all the information for drawing the graph of the statistical data. The WebUserConfig::chrplot.pm module is then added to configure the lod plot so that it is displayed in the LinkageView panel in Contig- and Cyto- View. Finally the plot image was scored into the Ensembl Contig View and Cyto View by modifying the cytocide code so that the linkage plot can be displayed into both Contig- and Cyto- View.

### 3 Figure(s).

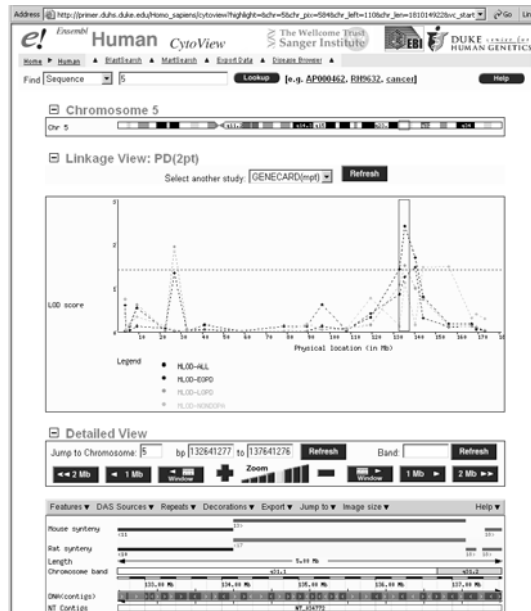


Figure 1: A screen shot of linkage data displayed in Contig View on a local server

### 3 References.

- [3] Dowell R.D., Jokerst R.M., Day A., Eddy S.R., and Stein L. 2001. The Distributed Annotation System. *BMC Bioinformatics*. 2:7.
- [1] Hubbard T., Barker D., Birney E., Cameron G., Chen Y., Clark L., Cox T., Cuff J., Curwen V., Down T., et al. 2002. The Ensembl genome database project. *Nucleic Acids Res*. 30:38-41.
- [4] Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M., and Haussler D. 2002. The human genome browser at UCSC. *Genome Res*. 12:996-1006.
- [2] Schuler GD. 1997. Sequence mapping by electronic PCR. *Genome Res*. 7:541-50.