Wilhelm Johannsen Centre for Functional Genome Research

University of Copenhagen

Visualization of 2260 chromosomal breakpoints in Mendelian Cytogenetics Network database associated with mental retardation

Mads Hjorth[1,3], Elizabeth Engenheiro[1], Søren Holstebroe[1], Iben Bache[1], Merete Bugge[1], Zeynep Tümer[1], Niels Tommerup[1] and MCN[2]

[1]Wilhelm Johannsen Center for Functional Genome Research, Copenhagen University[2]Mendelian Cytogenetic Network Consortium[3]madsh@mcndb.org

Introduction

Mental retardation, affecting 2-3% of the population, is an extremely heterogeneous condition where probably hundreds if not thousands of genes might be involved. Mendelian Cytogenetics Network is a collaboration of >300 cytogenetic laboratories that submit disease-associated balanced chromosomal rearrangements (DBCRs) to a central database [3]. Presently, there are 2908 DBCRs in MCNdb and by far the largest traitgroup is mental retardation (n=2260, 77.7%).

The great number of DBCRs in MCNdb is giving raise to many novel questions about statistical methods and visualisations for the initial selection of novel disease related genes for further investigation. This poster will present some of the tools and methods that are currently being developed and used at the Wilhelm Johannsen Centre for Functional Genome Research. All of the illustrations on this poster is created from MCNdb data with the use of tools available online for members of the network.

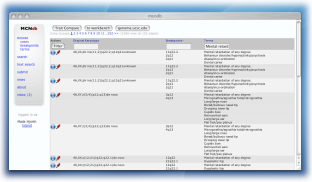

Figure 1. *The online interface to the browser function in MCNdb. The user can filter the available cases based on breakpoint location (i.e. 1p21), clinical terms (i.e. Mental retardation) and other criteria. The resulting set of cases can be shown as a track in NCBI and ENSEMBL genome browsers.*

NCBI and Ensembl

One important aspect of the digital annotation of DBCRs are the possibility to compare DBCRs with other positional annotations in common databases. Below is an illustration of the MCNdb feature for showing a subset of cases as genome annotation tracks. Each element on the track is hyperlinked back to the online representation in MCNdb.

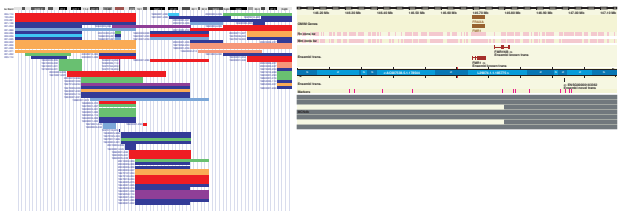

Figure 2. *A user-made track from the MCNdb shown in the UCSC [2] (left) with coloring according to different phenotypic variations and in the ENSEMBL [1] (right) genome browsers showing available markers and contigs used for sequencing around the possible chromosomal breakpoints.*

Statistics and comparison

One interesting aspect of the collective organisation of DBCRs is the possibility to make statistical considerations about the co-occurrence of clinical terms in subsets of DBCRs. In this case we have selected to focus on cases with mental retardation (n=1060) and find, not surprisingly, that other general terms is among the common terms ('Facies significantly abnormal' n=140, 13%; 'Seizures of any type' n=126, 12%), but also more specific terms unique to this subset has been identified ('Paresis of ocular muscles/squint' n=79, 8%; 'Hypotonia' n=115, 11%).

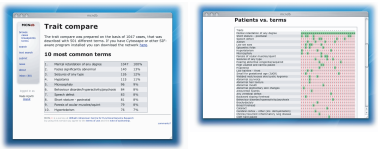

Figure 3. *Screen shots from the output of the* Trait Compare *on the subset of DBCRs related to 'Mental Retardation'*

Genomic distribution

To give a quick genome-wide overview of the distribution of selected chromosomal breakpoints various visualisation tools has been developed. Below is the output of the distribution tools from the MCNdb on the subset of cases with mental retardation. The overrepresentation on chromosome 7 is likely to be caused by recurrent translocation, whereas the chromosome 2 shows similar significans to chromosome 1 as found in the subset of infertile men [5].

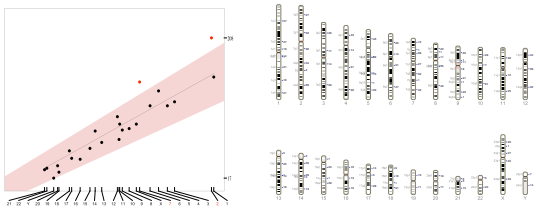

Figure 4. *The genomic distribution of breakpoints from cases with mental retardation found in MCNdb. On the left the numbers of cases are shown vertically for each chromosome. The shaded area represents the 95% confidence interval for a linear distribution according to the genomic length of each chromosome. On the right each chromosome band on a 86 chromosomal regions level is shown with the corresponding numbers of cases.*

Clustering

We have started developing clustering algorithms to identify significant clusters that might correspond to known or novel syndromes. This work is still in its initial phase and we have still to prove the concept of clustering of clinical terms in MCNdb cases statistically. Below is shown the first results from two different approaches for clustering terms with relations to mental retardation. Each node in the graphs represents a single clinical term souch as 'Hypotonia' or 'Clinodachtly' and each edge represents the existence of one or more case showing the two connected terms.

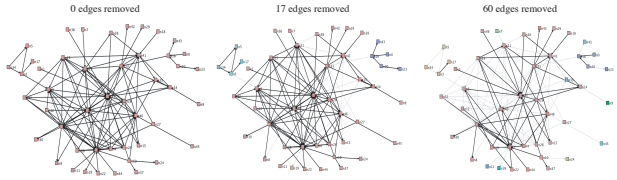

Figure 5. *Three steps of clustering by removing the edges with highest betweenness [6].*

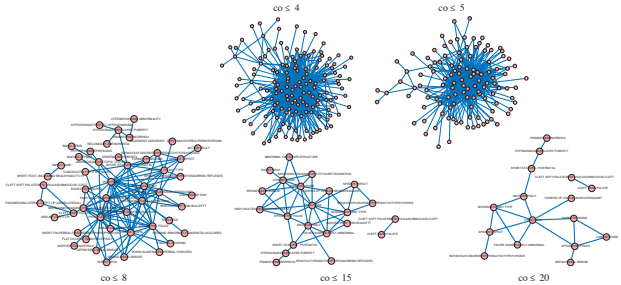

Figure 6. *Five different levels of cut-offs for the number of cases showing two clinical terms. Graphs with higher cut-off shows less connectivity, and only terms which are connected to others are shown.*

References

[1] http://www.ensembl.org

[2] The Human Genome Browser at UCSC. Kent, W.J., Sugnet, C. W., Furey, T. S., Roskin, K.M., Pringle, T. H., Zahler, A. M., and Haussler, D. *Genome Res. 12(6), 996-1006* (2002).

[3] Mendelian Cytogenetics Network database. Avaliable http://www.mcndb.org

[4] International System for Human Cytogenetic Nomenclature (1995) http://www.iscn1995.org

[5] An excess of chromosome 1 breakpoints in male infertility. Iben Bache et.al. *European Journal of Human Genetics 12, 9931000* (2004).

[6] The structure and function of complex networks. Mark Newman, *SIAM Review 45, 167-256* (2003).

Acknowledgements

Wilhelm Johannsen Centre for Functional Genome Research is established by the Danish National Research Foundation.

This poster is available as a PDF-file at
www.wjc.ku.dk