Detailed Sequence Analysis of the Distal Down Syndrome Critical Region and the SOD1/AML1 Region on Chromosome 21

Todd Taylor [1]
taylor@hgc.ims.u-tokyo.ac.jp
Masahira Hattori [1]
hattori@ims.u-tokyo.ac.jp
Kiyoteru Noguchi [2]
k-noguchi@comp.hitachi.co.jp
Tomomi Shobu [1]
shobu@ims.u-tokyo.ac.jp
Kazuo Ishii [1]
ishii@jet.sci.kitasato-u.ac.jp
Atsushi Toyoda [1]
toyoda@jet.sci.kitasato-u.ac.jp
Yoshiyuki Sakaki [1]
sakaki@ims.u-tokyo.ac.jp

[1] Human Genome Center, Institute of Medical Science, University of Tokyo
4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
[2] Hitachi, Ltd., Hitachi Omori 2nd Bldg., 27-18, Minami Oi 6-chome, Shinagawa-ku, Tokyo 140-8572, Japan

1 Introduction

Down syndrome (MIM 190685) is the most common genetic form of mental retardation. It is caused by trisomy of all or a portion of chromosome 21. The 2.5-Mb region [3, 4] between DNA markers D21S17 and ERG (MIM 165080) is associated with the main features of Down syndrome and is termed the Down syndrome critical region.

As part of the human genome project, our laboratory is studying the physical structure of human chromosome 21. To identify the genes and functional units mapped on chromosome 21, we are focusing on the physical mapping and sequencing of several regions. One region we have sequenced is the distal 1.6-Mb end of the Down syndrome critical region. The 4-Mb region from the SOD1 gene (MIM 147450) to the AML1 (MIM 151385) gene will soon be completely sequenced. This study describes the detailed sequence analysis of these two regions of chromosome 21.

2 Methods

Completed sequences are analyzed using several programs available on the Internet. Many of the programs use different input formats and require that analyses be run one at a time. Most of these programs can be run by cutting the large sequences into small, overlapping fragments and then saving them in FASTA format. For many of the analyses, the Baylor College of Medicine Search Launcher Batch Client is used to interface with their WWW server (http://gc.bcm.tmc.edu:8088/search-launcher/launcher.html) [5]. The batch client reads in multiple input files, performs several different searches (i.e., BLAST [1]), and stores the results on the user's computer as individual HTML files. The output files are parsed using several Perl scripts [2], and then the data is cleaned up and transferred to a database. The results are displayed graphically and conclusions are drawn. If necessary, sub-regions of the sequence are re-BLASTed under different parameters.

3 Results

A summary of the chromosome 21 analysis will be presented.

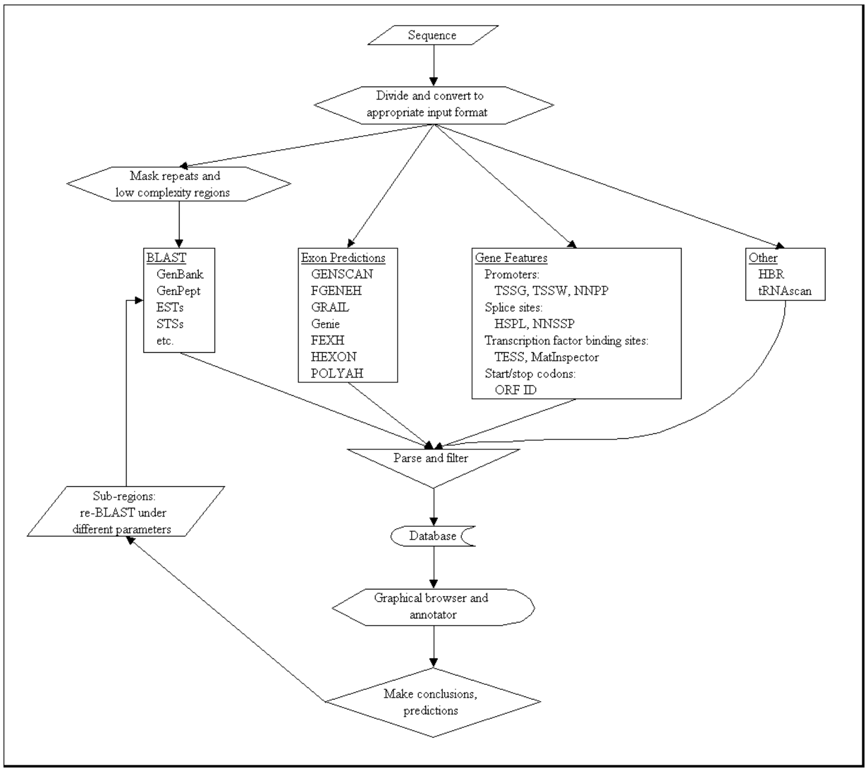

Figure 1: Sequence analysis flowchart.

References

[1] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.*, 25: 3389–3402, 1997.

[2] Wall, L., Christiansen, T. and Schwartz, R.L., *Programming Perl (Second Edition)*, O'Reilly & Associates, Inc., Sebastopol, 1996.

[3] Lucente, D., Chen, H.M., Shea, D., Samec, S.N., Rutter, M., Chrast, R., Rossier, C., Buckler, A., Antonarakis, S.E. and McCormick, M.K., Localization of 102 exons to a 2.5 Mb region involved in Down syndrome, *Hum. Mol. Genet.*, 4:1305–1311, 1995.

[4] Peterson, A., Patil, N., Robbins, C., Wang, L., Cox, D.R. and Myers, R.M., A transcript map of the Down syndrome critical region on chromosome 21, Hum *Mol. Genet.*, 3:1735–1742, 1994.

[5] Smith, R.F., Wiese, B.A., Wojzynski, M.K., Davison, D.B. and Worley, K.C., BCM Search Launcher–an integrated interface to molecular biology data base search and analysis services available on the World Wide Web, *Genome Res.*, 6:454–462, 1996.