ALOHOMORA: a tool for linkage analysis using 10K SNP array data

Franz Rüschendorf1,2,* and Peter Nürnberg2,3

1Bioinformatics Department and 2Gene Mapping Center, Max Delbrück Center (MDC) for Molecular Medicine, Berlin-Buch, Germany and 3Cologne Center for Genomics, University of Cologne, Cologne, Germany

Received on December 4, 2004; accepted on January 6, 2005
Advance Access publication January 10, 2005

ABSTRACT

Summary: ALOHOMORA is a software tool designed to facilitate genome-wide linkage studies performed with high-density single nucleotide polymorphism (SNP) marker panels such as the Affymetrix GeneChip® Human Mapping 10K Array. Genotype data are converted into appropriate formats for a number of common linkage programs and subjected to standard quality control routines before linkage runs are started. ALOHOMORA is written in Perl and may be used to perform state-of-the-art linkage scans in small and large families with any genetic model. Options for using different genetic maps or ethnicity-specific allele frequencies are implemented. Graphic outputs of whole-genome multipoint LOD score values are provided for the entire dataset as well as for individual families.

Availability: ALOHOMORA is available free of charge for non-commercial research institutions. For more details, see http://gmc.mdc-berlin.de/alohomora/

Contact: fruesch@mdc-berlin.de

In the past two decades, positional cloning via genome-wide linkage analysis in families has been a powerful approach to the elucidation not only of numerous Mendelian but also of common diseases, provided that high-risk alleles were involved in some of the families (Botstein and Risch, 2003; Carlson et al., 2004). Until recently it was common practice to use a panel of about 400 microsatellites at 10 cM average intermarker distance from the well-defined human genetic map for linkage analysis (Murray et al., 1994; Dib et al., 1996; Kong et al., 2002). However, recent progress in SNP discovery and genotyping provides the opportunity to use this marker type for linkage analysis as well (Collins et al., 1997; Matise et al., 2003). If properly selected only twice as many SNPs have to be analyzed in comparison to the highly polymorphic microsatellites to extract the same amount of linkage information from the studied families (Evans and Cardon, 2004). Moreover, the traditional 10 cM microsatellite scan is likely to miss linkage signals due to an inadequately low information content associated with this sparse map of markers (Evans and Cardon, 2004; Middleton et al., 2004; John et al., 2004). Thus, a high-density SNP panel for linkage analysis should comprise several thousands of markers. A convenient tool to genotype >10000 SNPs approximately equally distributed over the whole genome in a single experiment is the Affymetrix GeneChip® Human Mapping 10K Array (Kennedy et al., 2003; Matsuzaki et al., 2004). Obviously, the substantial increase of markers raises a problem for the analysis of the data as some linkage programs were designed for the requirements of conventional low-number marker sets. For instance, Genehunter 2.1 (Kruglyak et al., 1996) is restricted to 300 markers and Simwalk2 (Sobel and Lange, 1996) to 31 only. Partially, this may be overcome by using recompiled versions of the programs allowing for a higher maximum number of markers. Alternatively, the analysis may be performed with subsets of markers using a sliding window mode.

When we started using the Mapping 10K SNP Array for linkage analysis no software was available to import the data into the common linkage programs. Therefore, we developed our own program ALOHOMORA that easily converts Affymetrix genotype data into linkage and haplotype information. The program is written in Perl/Tk running under Windows and Linux. The current version accepts genotypic data as generated by the GeneChip DNA Analysis Software (GDAS v3.0) from Affymetrix.

With ALOHOMORA, a comprehensive quality control of the data can be performed accessing other freely available programs. Gender of samples is checked by counting the heterozygote SNPs on the X-chromosome and comparing it to the pedigree file information. The correct relationships within the families are checked by the program GRR (Abecasis et al., 2001). PedCheck is used for detection of Mendelian errors (O’Connell and Weeks, 1998). SNPs with Mendelian errors and SNPs that are not informative for any individual of a dataset can be selectively removed from the data. Non-mendelian errors are identified by the Merlin option ‘error’ (Abecasis et al., 2002) and the unlikely genotypes deleted in the individuals in which they occur. Other options are the chip version used, because SNP contents may differ between different versions of chips, the preferred genetic map and the allele frequencies for the appropriate ethnicity (Fig. 1A). For linkage analysis, data may be converted for Allegro (Gudbjartsson et al., 2000), Genehunter, Merlin and Simwalk2. For the chosen program, the user can define the genetic model when parametric analysis is performed, the size of a moving window and furthermore, select linkage program-specific options (Fig. 1B).

Non-parametric LOD score calculations are preferably performed with Merlin or Allegro, chromosome by chromosome using all SNPs on a chromosome simultaneously for a multipoint analysis. No limitation regarding the number of markers was observed up to 945 SNPs as known to be available for chromosome 2. Parametric linkage analysis was performed with Allegro v1.2, Genehunter 2.1v5 and with Simwalk2 v2.89. Due to the limitations of Genehunter and
Simwalk2 with respect to the number of markers, the analysis was done with subsets of markers in the way of a non-overlapping moving window. For Genehunter window sizes of 50–300 were used and pedigrees limited to max bits $\leq 20$. Simwalk2 was recompiled for using up to 255 markers in one run. In cases of large pedigrees, when Genehunter drops individuals and both Allegro and Merlin skip the pedigree, we split the pedigree to appropriate sizes to run Allegro, Merlin or Genehunter and used Simwalk2 to calculate the pedigree as a whole. NPL and parametric LOD scores may be plotted for all chromosomes.

All four programs, Allegro, Merlin, Genehunter and Simwalk2 generate haplotypes. Mostly we used the haplotyping from Genehunter’s haplo.dump file. For visualization purposes, we developed HaploPainter as a user-friendly tool for the handling of haplotype information in extended pedigrees (Thiele and Nürnberg, 2005).

In conclusion, linkage mapping with high-density SNP arrays is expected to tremendously expedite positional cloning studies. While genotyping with SNP arrays is straightforward, difficulties in data analysis seem to have prevented a broader application so far. Here, we present a graphical user interface, ALOHOMORA, as an open source software for the scientific community to facilitate linkage analysis with chip data. Using this program we successfully analyzed several genome scans performed with the Affymetrix GeneChip® Human Mapping 10K SNP array. These projects were not restricted to autozygosity mapping rather included recessive and dominant traits (Kaindl et al., 2004; Uhlenberg et al., 2004; Janecke et al., 2004). Altogether, Mapping 10K, ALOHOMORA, and HaploPainter are suggested to form a perfect tool box for high-speed gene mapping.

ACKNOWLEDGEMENTS

We would like to thank our colleagues and guests of the Gene Mapping Center at the MDC for their advice and software testing. This work was funded by the Federal Ministry of Science and Education of Germany through the National Genome Research Network.

REFERENCES

ALOHOMORA—using SNP data for linkage analysis

Kong,A., Gadjev,D.F., Saitz,J., Jonsdottir,G.M., Gudjonsson,S.A.,

Kruglyak,L., Daly,M.J., Reeve-Daly,M.P. and Lander,E.S. (1996) Parametric and non-
parametric linkage analysis: a unified multipoint approach. Am. J. Hum. Genet.,
58, 1347–1363.

Matise,T.C., Sachidanandam,R., Clark,A.G., Kruglyak,L., Wijsman,E., Kakol,J.,
Genet., 73, 271–284.

Matsuzaki,H., Loi,H., Dong,S., Tsai,Y.Y., Fang,J., Law,J., Di,X., Liu,W.M., Yang,G.,
Liu,G. et al. (2004) Parallel genotyping of over 10,000 SNPs using a one-primer

linkage analysis of bipolar disorder by use of a high-density single-nucleotide-
polymorphism (SNP) genotyping assay: a comparison with microsatellite marker
assays and finding of significant linkage to chromosome 6q22. Am. J. Hum. Genet.,
74, 886–897.

Murray,J.C., Buetow,K.H., Weber,J.L., Ludwigsen,S., Scherpber-Heidema,T., Man-
ion,F., Quillen,J., Sheffield,V.C., Sunden,S., Duyk,G.M. et al. (1994) A compre-
hensive human linkage map with centimorgan density. Cooperative Human Linkage
Center (CHLC). Science, 265, 2049–2054.

of genotype incompatibilities in linkage analysis. Am. J. Hum. Genet.,
63, 259–266.

Sobel,E. and Lange,K. (1996) Descent graphs in pedigree analysis: applications to hap-
lotyping, location scores, and marker sharing statistics. Am. J. Hum. Genet.,
58, 1323–1337.


Uhlenberg,B., Schuelke,M., Rischendorf,F., Ruf,N., Kaindl,A.M., Henneke,M.,
75, 251–260.