BMC Bioinformatics

Software

SNP HiTLink: a high-throughput linkage analysis system employing dense SNP data

Yoko Fukuda[1], Yasuo Nakahara[1], Hidetoshi Date[1], Yuji Takahashi[1], Jun Goto[1,2], Akinori Miyashita[3], Ryozo Kuwano[3], Hiroki Adachi[4], Eiji Nakamura[4] and Shoji Tsuji*[1,2]

Address: [1]Department of Neurology, Graduate School of Medicine, the University of Tokyo, Tokyo 113-8655, Japan, [2]Department of Clinical Genomics, the University of Tokyo Hospital, Tokyo 113-8655, Japan, [3]Department of Molecular Genetics, Bioresource Science Branch, Center for Bioresources, Brain Research Institute, Niigata University, Niigata 951-8585, Japan and [4]Gene Diversity Software Department, Dynacom Co Ltd., Kanagawa 220-0003, Japan

Email: Yoko Fukuda - yokofukuda-tky@umin.ac.jp; Yasuo Nakahara - yn-tky@umin.ac.jp; Hidetoshi Date - hdate-tky@umin.ac.jp; Yuji Takahashi - yutakahashi-ns@umin.net; Jun Goto - gotoj-tky@umin.ac.jp; Akinori Miyashita - miyashi@bri.niigata-u.ac.jp; Ryozo Kuwano - ryosun@bri.niigata-u.ac.jp; Hiroki Adachi - snpalyze-s@dynacom.co.jp; Eiji Nakamura - snpalyze-s@dynacom.co.jp; Shoji Tsuji* - tsuji@m.u-tokyo.ac.jp

* Corresponding author

Published: 24 April 2009

BMC Bioinformatics 2009, **10**:121 doi:10.1186/1471-2105-10-121

Received: 24 December 2008
Accepted: 24 April 2009

This article is available from: http://www.biomedcentral.com/1471-2105/10/121

Abstract

Background: During this recent decade, microarray-based single nucleotide polymorphism (SNP) data are becoming more widely used as markers for linkage analysis in the identification of loci for disease-associated genes. Although microarray-based SNP analyses have markedly reduced genotyping time and cost compared with microsatellite-based analyses, applying these enormous data to linkage analysis programs is a time-consuming step, thus, necessitating a high-throughput platform.

Results: We have developed SNP HiTLink (SNP High Throughput Linkage analysis system). In this system, SNP chip data of the Affymetrix Mapping 100 k/500 k array set and Genome-Wide Human SNP array 5.0/6.0 can be directly imported and passed to parametric or model-free linkage analysis programs; MLINK, Superlink, Merlin and Allegro. Various marker-selecting functions are implemented to avoid the effect of typing-error data, markers in linkage equilibrium or to select informative data.

Conclusion: The results using the 100 k SNP dataset were comparable or even superior to those obtained from analyses using microsatellite markers in terms of LOD scores obtained. General personal computers are sufficient to execute the process, as runtime for whole-genome analysis was less than a few hours. This system can be widely applied to linkage analysis using microarray-based SNP data and with which one can expect high-throughput and reliable linkage analysis.

Background

Recent technological development of high-density SNP chips has made it practical to genotype more than a million SNPs. Because microarray-based dense SNP typing requires less time and typing cost and can provide much more information than PCR-based microsatellite markers, it is now widely recognized as a powerful tool for linkage analysis [1-3]. To apply SNP information to genome-wide high-throughput linkage analysis, however, there are some difficulties as follows. 1) LINKAGE file preparation: Most linkage analysis software accepts LINKAGE format genotype data containing information on each marker for pairwise analysis or that on all markers on each chromosome for multipoint analysis. For example, pairwise analysis of 1000 SNPs on a chromosome using MLINK [4,5], a pairwise linkage analysis program, means preparing 1000 genotype files and 1000 marker information files, followed by running the program 1000 times. In multipoint analysis, information on the 1000 genotypes or marker information containing intermarker distances should be described in one file. Preparation of these files based on the information contained in the CHP file, which is generated by Affymetrix Genotyping Console ™ from firstly created CEL files in genotyping assays, are laborious and time-consuming for researchers. 2) Typing error: In microarray-based SNP detection, typing error is rare but inevitable because several factors such as the quality of genomic DNA, experimental conditions and the number of samples incorporated in the clustering of genotypes, can lead to inaccurate SNP calling [6-9]. This relatively rare miscalling, however, can lead to critical miscalculation in linkage analysis, particularly when par-

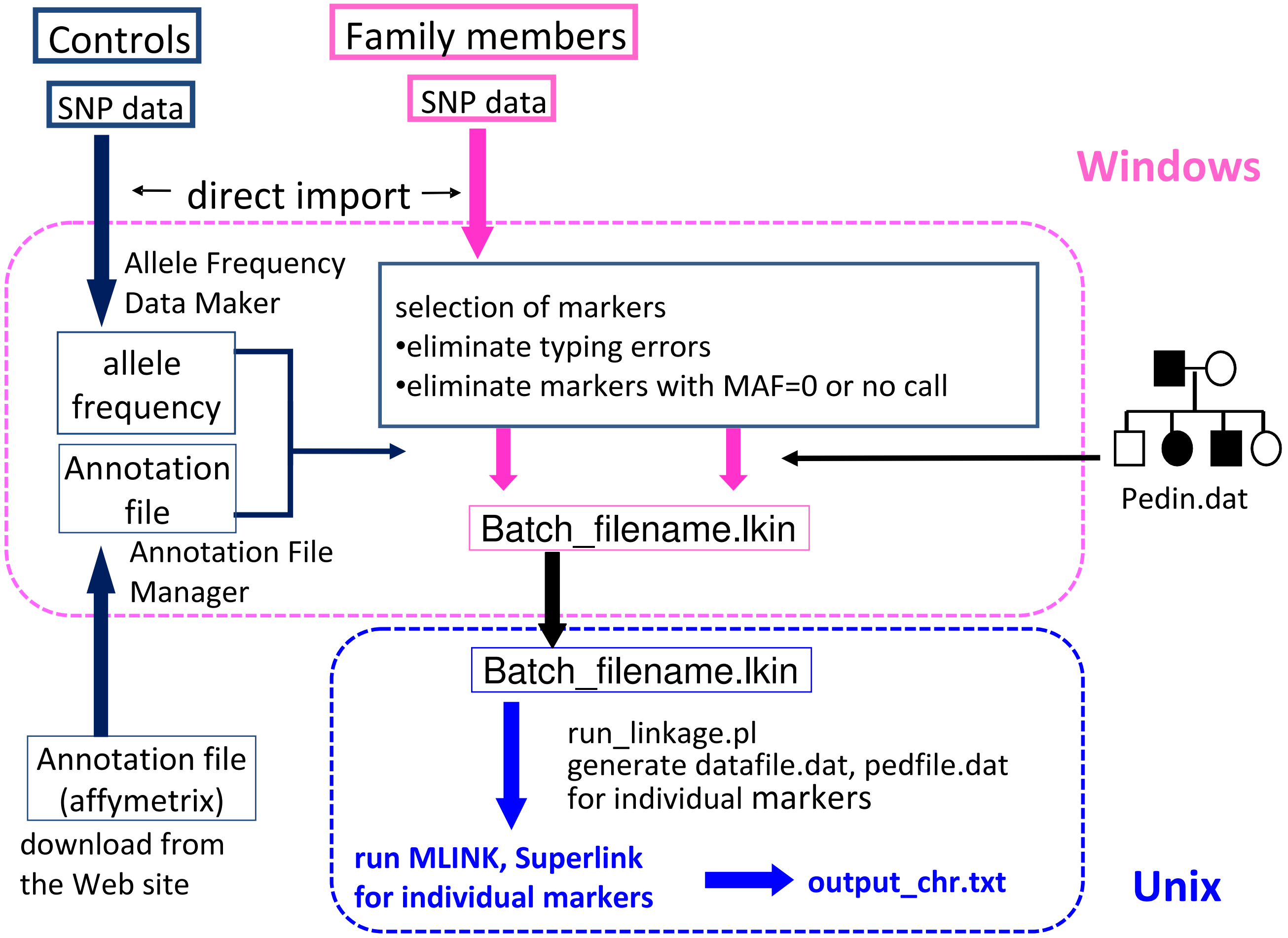


Figure 1

Flowcharts of data processing for pair-wise linkage analysis employing MLINK or Superlink by SNP HiTLink. In Windows OS, import of SNP data, generation of allele frequency file, annotation file and lkin file are conducted along with selection of markers. After lkin file is transported to Unix OS, run_linkage.pl carries out continuous run of MLINK or Superlink by rewriting pedin.pre and pedin.dat files for each marker.

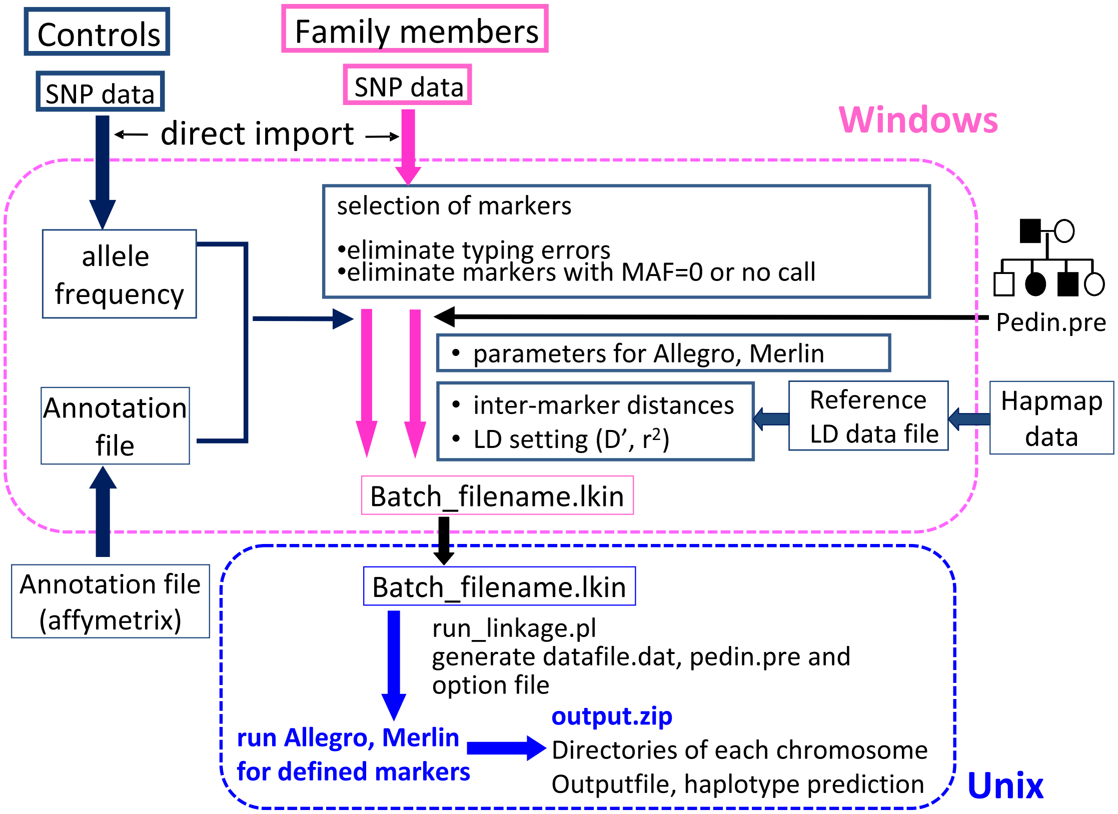


Figure 2

Flowcharts of data processing for multipoint linkage analysis employing Merlin or Allegro with SNP HiTLink. Procedures are basically similar to those by pair-wise analysis except that the model setting, selection of intermarker distances are executable here. run_linkage.pl carries out a run of Allegro or Merlin with all selected markers by writing whole information in pedin.pre, datain.dat.

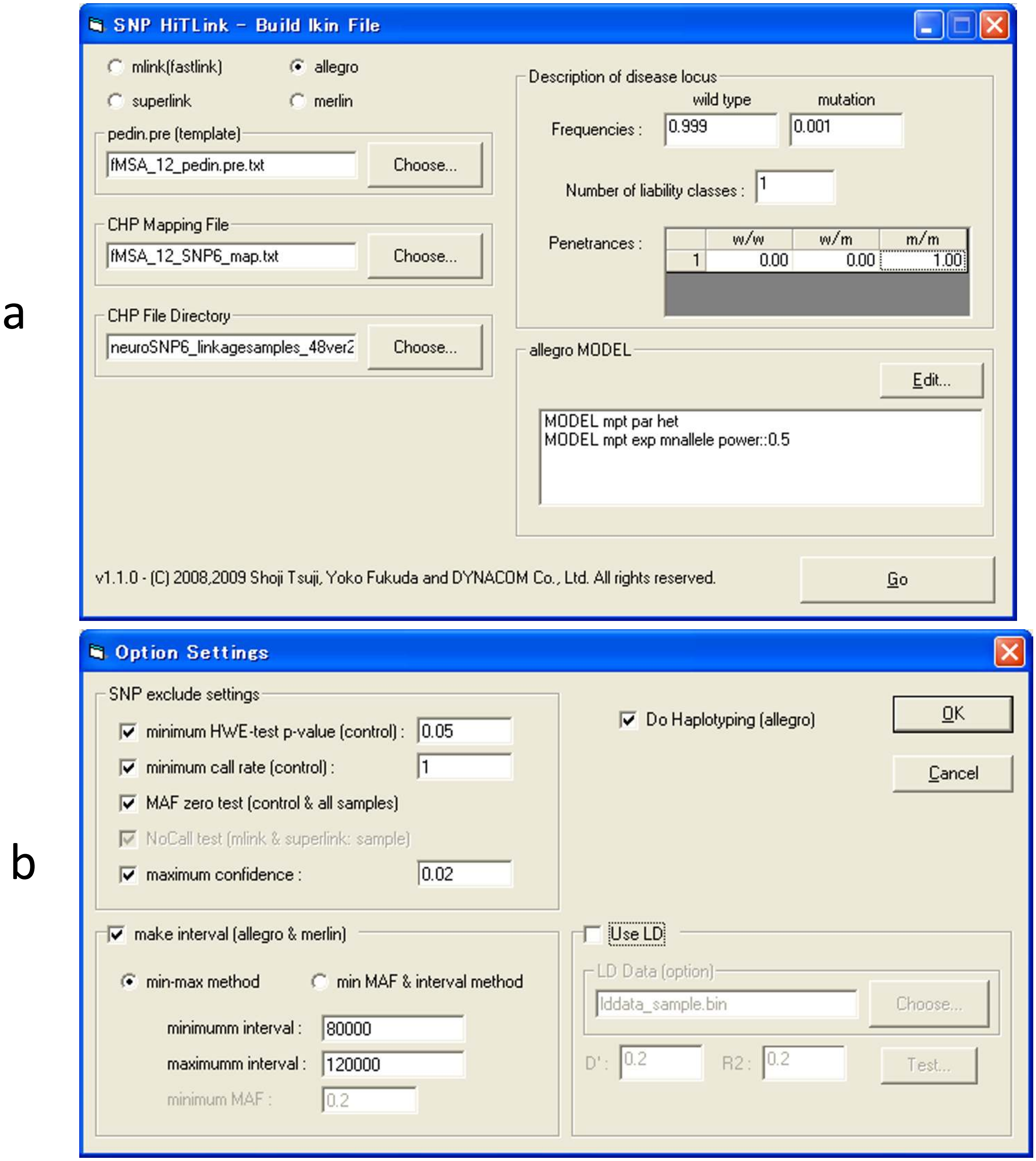
ent genotypes are lacking, or in multipoint analysis. Therefore, estimation and elimination of typing error data would be necessary for reliable results. 3) Linkage disequilibrium (LD) in neighboring markers for multipoint analysis: In algorithms of multipoint linkage analysis, it is usually assumed that all markers are in linkage equilibrium with each other. Markers in LD should be appropriately eliminated to avoid inaccurate calculation, which can be accompanied by inflation of LOD scores [10,11]. This is particularly important when using recently developed high-density SNP chips.

We have herein developed SNP HiTLink that directly accepts Affymetrix SNP CHP files and perform parametric/nonparametric linkage analyses with quite flexible marker selection functionalities.

Implementation

SNP HiTLink works under Windows XP SP2 or later/Vista (Use only 32-bit versions of Windows) and unix (supporting perl 5) OS [Additional files 1 and 2]. MLINK (LINKAGE/fastlink), Superlink, Merlin and Allegro should be installed in Unix OS. MLINK is included in FASTLINK package. Allegro is available from deCODE genetics, Inc. At present, SNP HiTLink accepts files in the CHP file format (filename.chp) of the Affymetrix Mapping 100 k/500 k array set and Genome-Wide Human SNP array 5.0/6.0. SNP HiTLink consists of two processes. The first process creates necessary data files by the program described in the Visual Basic programming on Windows OS, and these files are then transferred to Unix OS. The Perl script files invoke necessary linkage programs with necessary data files on Unix OS.

Figures 1 and 2 shows a flow-chart representing the process of linkage analysis. "Allele Frequency Data Maker" and "Annotation File Manager" programs are implemented in SNP HiTLink to obtain allele frequencies and SNP information. These are automatically generated from CHP files of control samples and annotation files downloaded from the Affymetrix web page. When analyzing a new family, users need to prepare a "map" file and "pedin.dat"



a

SNP HiTLink - Build lkin File

○ mlink(fastlink) ● allegro
○ superlink ○ merlin

pedin.pre (template): fMSA_12_pedin.pre.txt

CHP Mapping File: fMSA_12_SNP6_map.txt

CHP File Directory: neuroSNP6_linkagesamples_48ver2

Description of disease locus

	wild type	mutation
Frequencies :	0.999	0.001

Number of liability classes : 1

Penetrances :

	w/w	w/m	m/m
1	0.00	0.00	1.00

allegro MODEL

```
MODEL mpt par het
MODEL mpt exp mnallele power::0.5
```

v1.1.0 - (C) 2008,2009 Shoji Tsuji, Yoko Fukuda and DYNACOM Co., Ltd. All rights reserved.

b

Option Settings

SNP exclude settings

☑ minimum HWE-test p-value (control) : 0.05
☑ minimum call rate (control) : 1
☑ MAF zero test (control & all samples)
☑ NoCall test (mlink & superlink: sample)
☑ maximum confidence : 0.02

☑ Do Haplotyping (allegro)

☑ make interval (allegro & merlin)

● min-max method ○ min MAF & interval method

minimumm interval : 80000
maximumm interval : 120000
minimum MAF : 0.2

☐ Use LD

LD Data (option): lddata_sample.bin

D' : 0.2 R2 : 0.2

Figure 3
Interface of first step of "build lkin file"(a) and "option settings"(b) of SNP HiTLink.

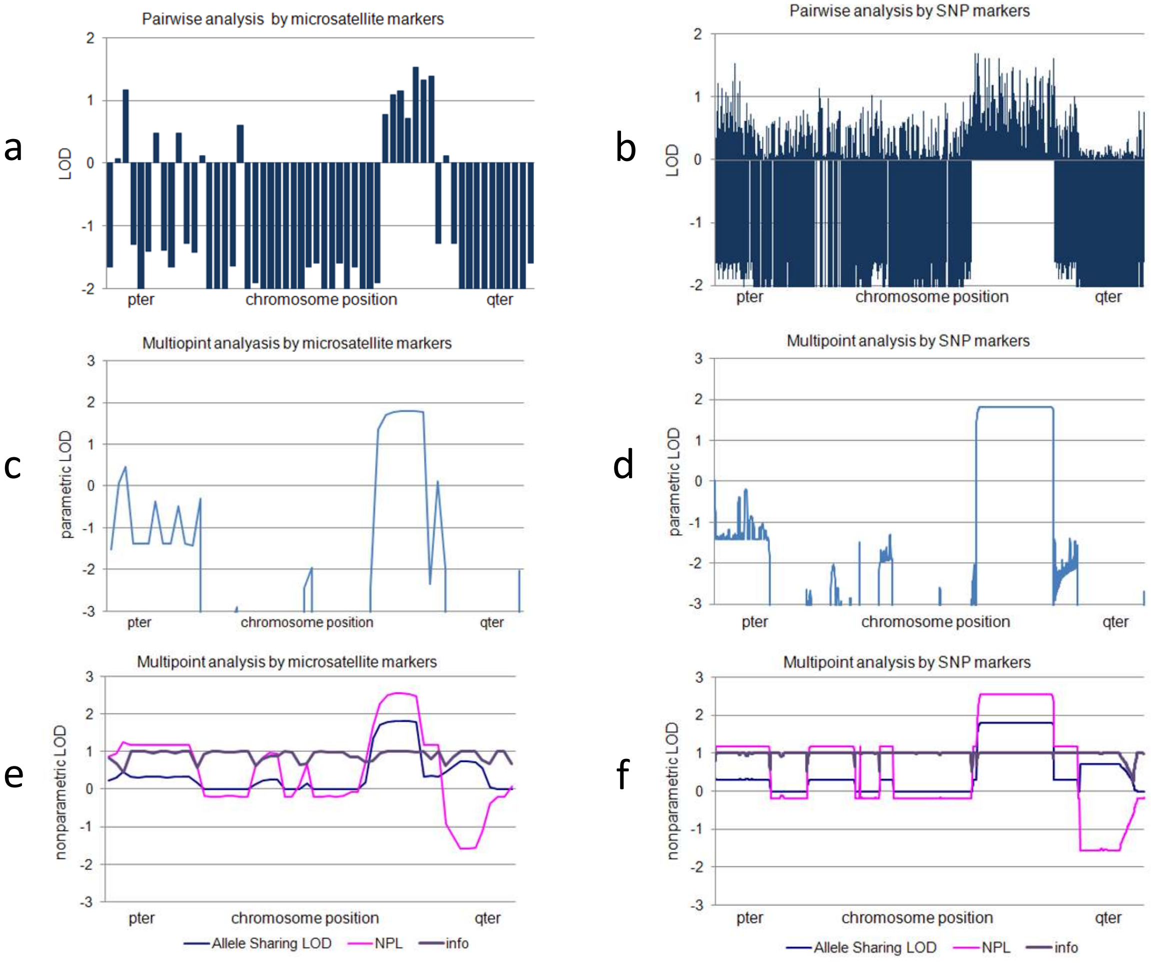


Figure 4

Results of pairwise analysis (a and b) by MLINK, multipoint parametric analysis (c and d), and multipoint nonparametric analysis (e and f) by Allegro employing microsatellite (a, c and e) and 100KSNP (b, d, and f) markers. SNP markers were selected as confidence score < 0.1, HWE > 0.05, call rate > 0.95, and intervals of 100 kb (for multipoint analysis). The x-axis represents the position on each chromosome and the y-axis represents calculated parametric LOD scores (allele sharing LOD), nonparametric linkage scores (NPL), or information measures (info).

(MLINK, Superlink) or "pedin.pre" (Merlin, Allegro) files manually by a text editor [see Additional file 3]. Although "pedin.dat" or "pedin.pre" should be described basically in the standard LINKAGE format (see manuals of each program for detail), no genotype data are required here. "map" files link an individual number described in "pedin.dat" or "pedin.pre" to the name of a "filename .chp" file from each individual.

SNP HiTLink can run four standard linkage analysis programs, MLINK [4,5], Superlink [12], Merlin [13] and Allegro [14,15]. Pair-wised analysis is supported by MLINK, Superlink and Allegro while multipoint analysis can be conducted by Merlin and Allegro in SNP HiTLink. Figure 3 shows the interface of the first step of the "build lkin file" (Figure 3a) and "option settings" (Figure 3b). For the pairwise linkage analysis by MLINK or Superlink, the user chooses pedin.dat and map files then specify the directory containing the CHP files. Disease gene frequency and liability class are defined here. For performing Merlin or Allegro, the user chooses pedin.pre files instead of pedin.dat, and then chooses model options that are identical to those originally implemented in Merlin and Allegro. After selecting programs and models, the user sets the marker-selecting options in which we implemented various parameters to eliminate typing errors and uninformative markers classified as follows.

1) To eliminate markers with typing errors, HWE, call rate, and confidence score are used as the effective indexes because deviations from HWE, lower call rates and higher confidence scores at particular markers sometimes suggest

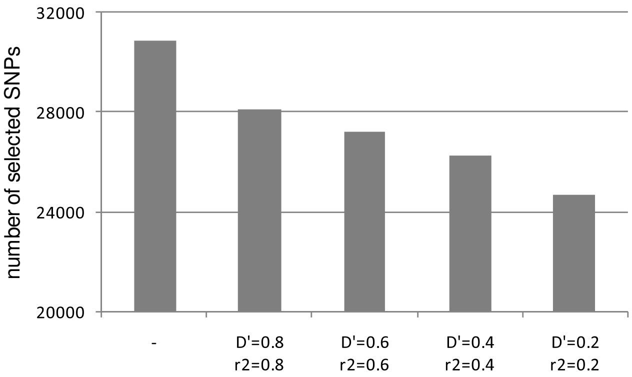


Figure 5
Number of markers on chromosome 1 employed in multipoint analysis (intervals of 100–500 bp, confidence score < 0.02, and HWE > 0.05, call rate = 1) with varied LD settings. DNA obtained from two affected siblings of a family was analyzed using Genome-Wide Human SNP array 6.0.

problems with genotyping. 2) To select informative markers useful for linkage analyses, the 'MAF zero test' and 'No call test' will be performed because these markers are totally uninformative. 3) To avoid employing markers in LD in the multipoint analysis, appropriate intermarker distances or D' and r^2, which are indexes of LD, can be defined by users.

- HWE test: the user sets p-value which is calculated from genotype frequencies in control samples. SNPs with a p-value below the settings are eliminated.

- Minimum call rate: the user sets the minimum call rate, which is calculated from "no call/call" ratio in all control samples, to avoid markers with lower call rates suggesting difficulties in genotyping.

- MAF zero test: markers where MAFs are zero can be eliminated.

- NoCall test (MLINK, Superlink): markers that are not called in any samples analyzed will be eliminated.

- Maximum confidence: confidence scores that are reliabilities of signal calling from hybridization can be set here. When the user skips this setting, the default value (for example 0.5 in BRLMM algorithm [16] as a default) defined in Genotyping Console™, which is Affymetrix genotyping software, will be used.

- Interval (Merlin, Allegro): minimum intermarker distances will be set. There are two marker-selecting methods, the min-max method and min MAF and interval method. In the min-max method, the user sets minimum and maximum intervals, then SNP with the highest MAF in the region defined by these intervals will be adopted. On the other hand, the min MAF and interval method select SNPs with MAFs higher than defined, and one SNP locating nearest to the minimum interval from the former SNP will be adopted.

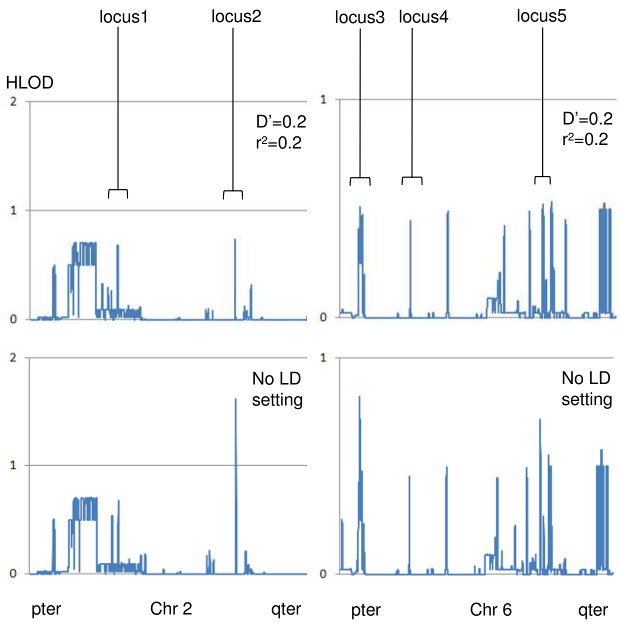
- LD: the user sets the maximum D' and r^2 scores to eliminate neighboring markers in LD with D' or r^2 scores higher than the threshold. The reference LD data file containing all D' and r^2 data obtained from the Hapmap database [17] can be downloaded from our WEB sites. Information of four ethnic populations (CEU, CHB, JPT, and YRI) has been provided as LD data files thus far. Users can make LD data files from their own samples by using LD Data Maker in the Main Menu. Click on LD Data Maker and specify the directory where chip files located.

SNP HiTLink produces a binary file (.lkin file) containing the marker and pedigree information with parameter settings, and this file is transported from Windows OS to Unix OS. Perl programming (run_linkage.pl) performs MLINK, Superlink, Merlin or Allegro against a specified '.lkin' file. Whole genome analysis will be carried out automatically but the user can also specify a chromosome number by option when analyzing only the chromosome of interest. Outputs of haplotype prediction by Allegro in a specific text format are easily visualized on the windows system by using the haplotype viewer implemented in this system. Data are shown in columns and can be copied to an Excel sheet for further use [see also the manual of Additional file 4].

Result and discussion

Figure 4 shows results of pairwise and multipoint analysis of a pedigree using the Affymetrix Mapping 100 K array set along with results obtained using microsatellite (ABI PRISM® Linkage Mapping Set) data. SNPs and microsatellite markers showed similar results in both pairwise and multipoint analyses but a higher resolution and a clearer border of regions where comparably high LOD scores were expected were achieved using SNP markers. These results indicated that SNP data were comparable or even superior to those obtained from microsatellite markers. The maximum LOD scores of pairwise analysis using microsatellite and SNP markers, were 1.7 and 1.5, respectively. In multipoint analysis, maximum parametric LOD score of 1.8, and nonparametric allele sharing LOD and NPL scores of 1.8 and 2.4, respectively, were obtained using both microsatellite and SNP markers.

We tested the effect of LD setting on the number of markers and LOD scores of parametric multipoint analysis employing Genome-Wide Human SNP array 6.0. Approximately 70000 SNP markers are placed on chromosome 1 of SNP array 6.0. Of these, about 31000 were selected with parameter settings of 100–500 bp interval, call rate = 1,



locus1	locus2	locus3	locus4	locus5
rs7573815	rs6712279	rs12200008	rs9394287	rs870816
rs10173062	rs13418640	rs9392980	rs707969	rs6569657
rs12713425	rs6730966	rs672278	rs7746988	rs1932105
rs10184932	rs6739734	rs9378528	rs9470007	rs6939344
rs13415061	rs2461773	rs10484751	rs3798343	rs9492480
rs17538127	rs2461754	rs17143504	rs9658125	rs4895867
rs13029229	rs7582039	rs2817084	rs4713859	rs4897380
rs6545832	rs2437901	rs2078480	rs7757405	rs2876097
rs1432297	rs6737259	rs9406122	rs11962317	rs6927351
rs842631	rs16863235	rs1934788	rs9296155	rs7759336
rs10203477	rs34489473	rs10484750	rs2894403	rs2876096
rs3732179	rs10210906	rs4959483	rs9470055	rs2064873
rs12479056	rs6723304	rs17143665	rs1413896	rs12198460
rs10208137	rs12693068	rs9392217	rs4713897	rs7758496
rs4643526	rs10179721	rs995838	rs7757037	rs6923783
rs1864268	rs10497442	rs1543731	rs9368881	rs9492538
rs2564114	rs6742019	rs915355	rs2766551	rs6928320
rs1177234	rs4392224	rs2327064	rs4713923	rs6909986
rs3108630	rs11694981	rs6927589	rs2817064	rs724324
rs1177279	rs9287986	rs11964691	rs9462128	rs4141939
rs1177264	rs4344900	rs4960424	rs6923015	rs11154572
rs1177288	rs4072816	rs6923437		rs1398816
rs2600671	rs13394930	rs6942177		rs2127022
rs1059541	rs4599077	rs230253		
rs2290324	rs7578330	rs330109		
rs2060824	rs6734627	rs4960438		
rs12622458	rs1446825	rs1796697		
rs778157	rs1348843	rs162281		
rs778143	rs11685409	rs850185		
rs17008940	rs765578	rs2984100		
rs3821222	rs6756201	rs7753781		
rs2305155	rs1446832	rs9393054		
rs766448	rs2168	rs9328500		
rs3732171	rs2044491	rs7739573		
rs778752	rs2044490			
	rs11681401			
	rs6738457			
	rs4972795			
	rs1446660			

Figure 6

Effect of LD between markers on multipoint parametric heterogeneity LOD scores on chromosome 2 and 6. Multipoint analysis by Allegro (intervals of 100–500 bp, confidence score < 0.02, and HWE > 0.05) were conducted with strict LD settings (D' = r² = 0.2) or without settings. Results of chromosome 2 and 6 were shown. DNAs obtained from four affected sibling pairs were analyzed by Genome-Wide Human SNP array 6.0. SNP IDs of five loci were extracted. Colored SNP IDs are those eliminated in analysis with LD settings.

confidence score < 0.02, and HWE > 0.05. SNP markers were eliminated proportionately with decreasing D' and r^2 and about 28000 SNP markers were retained when D' = 0.2 and r^2 = 0.2, indicating that there are many neighboring markers that are in LD from each other (Figure 5). When multipoint parametric linkage analysis of four pedigrees including two affected siblings without parent genotypes was conducted without setting a LD threshold, the multipoint HLOD (heterogeneity LOD) scores showed inflation compared with those obtained at the setting of D' < 0.2, r^2 < 0.2 (Figure 6). Inflation was severer at the loci employing many markers in LD (loci 2, 3 and 5) than at the locus where no or only few LD markers were found (locus 1 and 4), suggesting this inflation was mainly due to the LD of markers. Given that our result was obtained from only four families with two affected siblings, markers in LD can have serious effects on the calculation of LOD scores when a large number of families are simultaneously analyzed, as sometimes LOD scores can inflate markedly as simulated in a previous study [10].

The runtime for preparing lkin files is less than 10 minutes (usually from about 10 second to a few minutes), and the runtime of whole genome linkage analysis of a pedigree performed using general personal computer was about 4 hours for pairwise analysis, when using all of approximately 1 million markers on Genome-Wide Human SNP array 6.0. For multipoint analysis less than 1 hour was required even in the case of a family including consanguineous loops when intermarker distances were set to be varied from 300 bp to 100 kbp. These results show that extremely dense markers that are now mainly utilized for the genome wide association study (GWAS) can also be utilized for high-throughput linkage analysis.

Conclusion

We have developed the SNP HiTLink, system for executing parametric/nonparametric linkage analysis using SNP data. This is the first and unique system that directly accepts recent 100 K, 500 K and 1 M markers of Affymetrix SNP CHP files and prepares very flexible marker-selecting implementations for linkage analysis, although some convenient pipelines that pass the SNP data to a linkage analysis program [18,19] or tools for visualization and removal of LD [20,21] have been developed thus far. The results using this system were comparable or even superior to those obtained using microsatellite markers, convincing us the advantage of using SNP data obtained by DNA microarray for linkage analysis. The number of SNP data located on a single chip is continuing to increase owing to recent developed technologies and demands for dense markers for GWAS. On the other hand, we should be carefully concerned about typing error data when using such dense SNP data for multipoint linkage analysis. Quite flexible marker-selecting implementations on SNP HiTLink will be advantageous from this point of view. Although SNP HiTLink only accepts Affymetrix SNP Chip files, improvements that support multiple platforms for SNP typing such as Illumina are required in the future. Furthermore, more user-friendly interface where analyses can be processed simply (for instance, through integrated single GUI) rather than transporting files from Windows to Unix OS, will be desirable. This system can be widely applied for linkage analysis using microarray-based SNP data, with which one can expect high-throughput and reliable linkage analysis.

Authors' contributions

YF, HA and EN dealt with the computational aspects in development of the system, and YF carried out analyses of the data. YN, YT performed SNP genotyping, and AM, RK performed microsatellite genotyping. HD, JG contributed to general planning and interpretation. ST provided overall guidance for this project. All authors read and approved the final manuscript.

Additional material

Additional file 1

SNP HiTLink. The SNP HiTLink main program.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-121-S1.zip]

Additional file 2

run_linkage.pl. The run_linkage.pl program for executing linkage analysis program in Unix OS.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-121-S2.zip]

Additional file 3

sample_data. Sample file set including a map file, pedin.dat and pedin.pre file.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-121-S3.zip]

Additional file 4

SNP HiTLink manual. The manual for SNP HiTLink.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-10-121-S4.pdf]

Acknowledgements

The authors would like to thank Drs. Toshihisa Takagi and Tatsuhiko Tsunoda for their insightful comments and helpful suggestions for the system. This work was supported in part by KAKENHI (Grant-in-Aid for Scientific Research) on Priority Areas, Applied Genomics, the 21st Century COE Program, Integrated Database Project, Center for Integrated Brain Medical Science, and Scientific Research (A) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1. Evans DM, Cardon LR: **Guidelines for genotyping in genomewide linkage studies: single-nucleotide-polymorphism maps versus microsatellite maps.** *Am J Hum Genet* 2004, **75(4):**687-692.
2. Matise TC, Sachidanandam R, Clark AG, Kruglyak L, Wijsman E, Kakol J, Buyske S, Chui B, Cohen P, de Toma C, *et al.*: **A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set.** *Am J Hum Genet* 2003, **73(2):**271-284.
3. John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, Vasavda N, Mills T, Barton A, Hinks A, *et al.*: **Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites.** *Am J Hum Genet* 2004, **75(1):**54-64.
4. Cottingham RW Jr, Idury RM, Schaffer AA: **Faster sequential genetic linkage computations.** *Am J Hum Genet* 1993, **53(1):**252-263.
5. Lathrop GM, Lalouel JM, Julier C, Ott J: **Strategies for multilocus linkage analysis in humans.** *Proc Natl Acad Sci USA* 1984, **81(11):**3443-3446.
6. Montgomery GW, Campbell MJ, Dickson P, Herbert S, Siemering K, Ewen-White KR, Visscher PM, Martin NG: **Estimation of the rate of SNP genotyping errors from DNA extracted from different tissues.** *Twin Res Hum Genet* 2005, **8(4):**346-352.
7. Hong H, Su Z, Ge W, Shi L, Perkins R, Fang H, Xu J, Chen JJ, Han T, Kaput J, *et al.*: **Assessing batch effects of genotype calling algorithm BRLMM for the Affymetrix GeneChip Human Mapping 500 K array set using 270 HapMap samples.** *BMC Bioinformatics* 2008, **9(Suppl 9):**S17.
8. Saunders IW, Brohede J, Hannan GN: **Estimating genotyping error rates from Mendelian errors in SNP array genotypes and their impact on inference.** *Genomics* 2007, **90(3):**291-296.
9. Gordon D, Heath SC, Ott J: **True pedigree errors more frequent than apparent errors for single nucleotide polymorphisms.** *Hum Hered* 1999, **49(2):**65-70.
10. Huang Q, Shete S, Amos CI: **Ignoring linkage disequilibrium among tightly linked markers induces false-positive evidence of linkage for affected sib pair analysis.** *Am J Hum Genet* 2004, **75(6):**1106-1112.
11. Schaid DJ, McDonnell SK, Wang L, Cunningham JM, Thibodeau SN: **Caution on pedigree haplotype inference with software that assumes linkage equilibrium.** *Am J Hum Genet* 2002, **71(4):**992-995.
12. Fishelson M, Geiger D: **Exact genetic linkage computations for general pedigrees.** *Bioinformatics* 2002, **18(Suppl 1):**S189-198.
13. Abecasis GR, Cherny SS, Cookson WO, Cardon LR: **Merlin – rapid analysis of dense genetic maps using sparse gene flow trees.** *Nat Genet* 2002, **30(1):**97-101.
14. Gudbjartsson DF, Jonasson K, Frigge ML, Kong A: **Allegro, a new computer program for multipoint linkage analysis.** *Nat Genet* 2000, **25(1):**12-13.

15. Gudbjartsson DF, Thorvaldsson T, Kong A, Gunnarsson G, Ingolfsdottir A: **Allegro version 2.** *Nat Genet* 2005, **37(10):**1015-1016.
16. Affymetrix I: **BRLMM: an Improved Genotype Calling Method for the GeneChip® Human Mapping 500 K Array Set.** 2006 [http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf].
17. The International HapMap Consortium: **The International HapMap Project.** *Nature* 2003, **426(6968):**789-796.
18. Hoffmann K, Lindner TH: **easyLINKAGE-Plus – automated linkage analyses using large-scale SNP data.** *Bioinformatics* 2005, **21(17):**3565-3567.
19. Hiekkalinna T, Peltonen L: **New program: AUTOSCAN 1.0 automated use of linkage analysis programs.** *American Journal of Human Genetics* 1999, **65(4):**A254-A254.
20. Webb EL, Sellick GS, Houlston RS: **SNPLINK: multipoint linkage analysis of densely distributed SNP data incorporating automated linkage disequilibrium removal.** *Bioinformatics* 2005, **21(13):**3060-3061.
21. Gaunt TR, Rodriguez S, Zapata C, Day IN: **MIDAS: software for analysis and visualisation of interallelic disequilibrium between multiallelic markers.** *BMC Bioinformatics* 2006, **7:**227.