BMC Bioinformatics

Software

iHAP – integrated haplotype analysis pipeline for characterizing the haplotype structure of genes

Chun Meng Song†[1], Boon Huat Yeo†[1], Erwin Tantoso†[1], Yuchen Yang†[2], Yun Ping Lim[1], Kuo-Bin Li[3] and Gunaretnam Rajagopal*[1]

Address: [1]Bioinformatics Institute, 30 Biopolis Street, #07-01, 138671, Singapore, [2]Institute of Molecular Cell and Biology, 61 Biopolis Drive (Proteos), 138673, Singapore and [3]Bioinformatics Center, National Yang-Ming University, Taipei, 112, Taiwan

Email: Chun Meng Song - alfreds@bii.a-star.edu.sg; Boon Huat Yeo - yeobh@bii.a-star.edu.sg; Erwin Tantoso - erwint@bii.a-star.edu.sg; Yuchen Yang - ycyang@imcb.a-star.edu.sg; Yun Ping Lim - yunping@bii.a-star.edu.sg; Kuo-Bin Li - kbli@ym.edu.tw; Gunaretnam Rajagopal* - guna@bii.a-star.edu.sg

* Corresponding author †Equal contributors

Published: 01 December 2006

BMC Bioinformatics 2006, **7**:525 doi:10.1186/1471-2105-7-525

Received: 13 June 2006
Accepted: 01 December 2006

This article is available from: http://www.biomedcentral.com/1471-2105/7/525

Abstract

Background: The advent of genotype data from large-scale efforts that catalog the genetic variants of different populations have given rise to new avenues for multifactorial disease association studies. Recent work shows that genotype data from the International HapMap Project have a high degree of transferability to the wider population. This implies that the design of genotyping studies on local populations may be facilitated through inferences drawn from information contained in HapMap populations.

Results: To facilitate analysis of HapMap data for characterizing the haplotype structure of genes or any chromosomal regions, we have developed an integrated web-based resource, iHAP. In addition to incorporating genotype and haplotype data from the International HapMap Project and gene information from the UCSC Genome Browser Database, iHAP also provides capabilities for inferring haplotype blocks and selecting tag SNPs that are representative of haplotype patterns. These include block partitioning algorithms, block definitions, tag SNP definitions, as well as SNPs to be "force included" as tags. Based on the parameters defined at the input stage, iHAP performs on-the-fly analysis and displays the result graphically as a webpage. To facilitate analysis, intermediate and final result files can be downloaded.

Conclusion: The iHAP resource, available at http://ihap.bii.a-star.edu.sg, provides a convenient yet flexible approach for the user community to analyze HapMap data and identify candidate targets for genotyping studies.

Background

The identification of Single Nucleotide Polymorphisms (SNPs) that contribute to complex diseases has made them the preferred choice for diagnostics and therapeutics studies. For instance, the methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism (dbSNP: rs1801133) has been reported to be associated with gastric cancer in the Chinese population [1]. To uncover novel markers that may be associated to a disease, genotyping studies are conducted to determine the genetic variations between diseased and healthy subjects, allowing for further functional characterization that could lead to therapeutic applications. While it may not be sufficient coverage just to genotype only these specific disease-related SNPs, it is costly to genotype all available SNPs from a large sample of individuals. As such, by genotyping only a subset (also known as the tag SNPs [2]), which may include the disease associated SNPs, the cost and effort involved in association studies can be effectively reduced with minimal compromise to the power of such studies.

In the absence of comprehensive genotype data from local populations, genotyping studies can be designed using data from the International HapMap Project [3]. Recent studies show that, despite differences in the fine details of linkage disequilibrium (LD) patterns between populations [4], tag SNPs selected from one HapMap population can be used to characterize other populations reasonably well [5-7]. These findings indicate that HapMap data is currently the most ideal freely-available dataset for tag SNP selection and association studies.

Currently, there exist several tools to facilitate haplotype block inference and tag SNP selection. For example, the International HapMap Project website [8] not only provides bulk download of genotype and frequency data from the International HapMap Project, but also interactive access to visualize the distribution of SNPs for any genomic region of interest. Haploview [9] is a standalone application that performs LD and haplotype block analysis on publicly available or user supplied genotype data. htSNPer1.0 [10] and HaploBlock [11] can also be used to analyze genotype data supplied by users. HaploBlockFinder [12] is a web-based tool that allows for the inference of haplotype blocks and tag SNP selection from genotype data uploaded by users. The Genome Variation Server (GVS) [13] and PupaSuite [14] are other tools with several online analysis utilities for accessing human genotype data. More recently, TAMAL [15] was developed adopting a pre-processing strategy to facilitate the selection of potential genotyping targets.

As a web-based tool, the iHAP resource complements the capabilities of existing tools for analyzing haplotype structures and selecting tag SNPs using data from HapMap. While most tools deploy limited algorithms for selecting tag SNPs, iHAP provides a wider selection of algorithms and parameter settings for both haplotype block partitioning as well as tag SNP definitions. This is achieved through the use of HapBlock [16] as iHAP's backend haplotype analysis tool. The HapMap Project website and Haploview select tag SNPs using Tagger [17], which is based on the pairwise LD r^2 [2] method. In addition to r^2 methods, iHAP incorporates tag SNP definitions including common haplotypes [18], haplotype diversity [19], haplotype entropy [20] and haplotype determination coefficient [21]. This provides users with fully customizable options for on-the-fly analysis as opposed to pre-processed results provided by TAMAL. iHAP also highlights SNPs found in coding regions so potentially significant SNPs may be "force included" as tag SNPs at users' discretion, a feature unavailable in GVS and TAMAL. Being integrated with our local repositories of genotype and gene data from HapMap and the UCSC Genome Browser Database [22] respectively, iHAP relieves users of the hassle of having to locate and download genotype data as is the case with Haploview. Furthermore, iHAP generates result pages that graphically depict the haplotype structures, including blocks, haplotype patterns and tag SNPs, alongside the exons and introns of genes found within the chromosomal region. Alternative sets of inferred tag SNPs are also presented with the respective scores. The key differences between iHAP and other similar tools are highlighted in Table 1.

Implementation

The iHAP resource was written in the PHP 5.1.4 scripting language with the GD library of image functions. Using a backend MySQL 4.1.14 relational database, this resource is currently deployed on a Solaris environment with Apache HTTP Server 2.0.58 running on a Sun Fire V240 Server. An overall schematic architecture of iHAP is shown in Figure 1.

Table 1: Comparison of iHAP vs other similar tools

	iHAP	GVS	TAMAL	PupaSuite	htSNPer1.0
Integration with HapMap data?	√	√	√	√	X
Integration of gene information?	√	√	√	√	X
Accept user's genotype data?	X	√	X	√	√
Web-based?	√	√	√	√	X
Force tag SNP selection?	√	X	X	X	X
Variety of block partitioning algorithms?	√	X	X	X	√
Variety of tag SNP definitions?	√	X	√	X	√
Graphical display of blocks and genes along chromosome?	√	X	X	X	X
Real-time analysis?	√	√	X	√	√

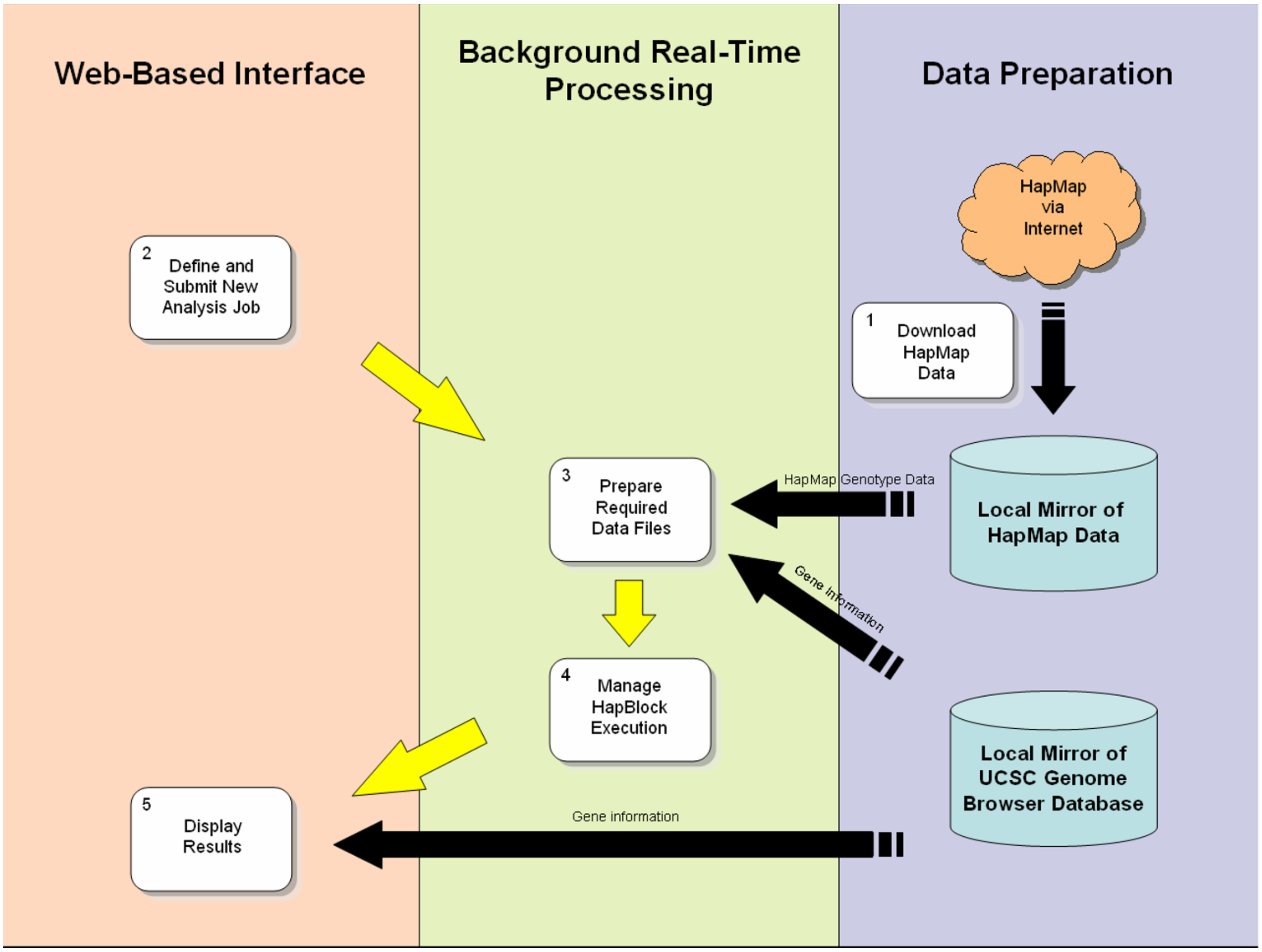

Web-Based Interface
Background Real-Time Processing
Data Preparation
HapMap via Internet
1 Download HapMap Data
2 Define and Submit New Analysis Job
3 Prepare Required Data Files
HapMap Genotype Data
Local Mirror of HapMap Data
Gene information
4 Manage HapBlock Execution
Local Mirror of UCSC Genome Browser Database
5 Display Results
Gene information

Figure 1
Overall schematic of iHAP. The iHAP resource may be conceptualized as having three components. The first involves batch-based data preparation while the second and the third are for real time analyses. Users submit jobs to iHAP via the web-based interface and each job is then processed in the background. Upon completion, results are returned to the users via the web-based interface.

Choice of backend haplotype analysis tool

Apart from HapBlock, other tools including HaploBlock and HaploBlockFinder were also considered and evaluated for suitability as iHAP's backend haplotype analysis tool. Eventually, HapBlock was preferred over these alternatives because it offers a wider selection of haplotype block definitions and tag SNP selection algorithms. HapBlock is also capable of accommodating the option for "forcing" specific SNPs to be selected as tags, which is helpful if one wants to include prior information into the analysis.

Local data repositories

Essential to the execution of the iHAP resource are two data repositories. The first is a database created for storing HapMap data (HapMap Public Release #21, Jul 2006, on NCBI Build 35 assembly). This database adopts a schema that was designed to support efficient queries for genotype and haplotype data for any population and genomic regions of interest. Genotype data of the four HapMap human populations, namely CEPH (Utah Residents with Northern and Western European Ancestry) (CEU), Han Chinese in Beijing, China (CHB), Japanese in Tokyo, Japan (JPT), Yoruba in Ibadan, Nigeria (YRI), was subsequently downloaded from the International HapMap Project website [8] and populated into the local database.

The other resource is the local mirror of the UCSC Genome Browser Database. To ensure that the SNP positions obtained from HapMap are consistent with the gene locations from the UCSC Genome Browser Database, the hg17 assembly is used. The human reference sequence in this assembly is based on NCBI build 35. This resource is used for determining the chromosomal locations of genes, including the positions of their respective introns and exons.

Job execution and management

Based on the settings supplied by users, iHAP generates the necessary input files in the format required by HapBlock and triggers its execution as a background job. Depending on the nature of individual haplotype analysis jobs, the execution time could vary from seconds to hours. In addition, the storage requirements for each job also vary according to the availability of genotype data for the selected chromosomal region. Therefore, it is necessary to optimize job scheduling to present users with a logical and coherent interface without compromising server performance.

To address this issue, a job manager module was devised. This module not only initiates each HapBlock execution as a background job, but also monitors the execution process through periodic polls. With this module, the progress of each job can be tracked so users may be updated with the current status of their jobs via the web-based interface. An email alert mechanism is also in place to inform users upon completion of their analysis jobs. To keep storage requirements in check, a script that automatically cleans up redundant files belonging to old jobs is also executed periodically.

Result display

As individual jobs are completed, the job manager module extracts information pertaining to haplotype blocks and tag SNPs to an intermediate format. Alternative sets of results are collated along with their respective scores while exon and intron information of genes found within the chromosomal region of interest is obtained from the local mirror of the UCSC Genome Browser Database. Such information is then combined along with additional details such as SNP names and locations in the dynamically generated image that illustrates the haplotype structure graphically. Intermediate files relating to individual jobs are finally archived in ZIP files which can be downloaded conveniently.

Results and discussion

Based on the submitted gene name, the iHAP resource determines the chromosomal region of interest using the UCSC Genome Browser Database. The setup of the analysis job is then defined according to parameters such as HapMap population, allele frequency threshold, block definitions, tag SNP definitions, permutation test settings, as well as SNPs to be "force included" as tags. Snippets of help for each parameter are ergonomically positioned to facilitate the configuration of each job.

The necessary files required by HapBlock are then generated by iHAP. These include the parameter, genotype or haplotype data, SNP names, SNP position lookup, and "forced tag SNP" files. iHAP then invokes the execution of HapBlock as a background job and monitors its progress through periodic polls so as to keep users updated on their job progression. Upon completion of the analysis, results are converted to dynamically generated images for display as a webpage.

The results page first provides a summary of the settings used for the analysis. A graphical representation of the genomic region is displayed with the locations of genes and their respective intronic and exonic regions illustrated as grey boxes, and inferred blocks as yellow rectangles. SNP locations are marked as blue vertical lines with those selected as tags augmented with red triangles. The next section depicts the structure of each block, including the dbSNP identifiers and the haplotype patterns along with their respective frequencies. The scores of the displayed and alternative tag SNP sets for each block are tabulated according to various criteria in the following section. Finally, an archive (ZIP) file containing all the files generated in the analysis can be downloaded via a link on the results page. A typical analysis flow is exemplified in Figure 2.

Conclusion

The iHAP application provides a one-stop resource for inferring haplotype blocks and selecting tag SNPs from HapMap data. Apart from providing a wider selection of algorithms and integrating genotype data with gene information, iHAP also offers greater flexibilities by allowing users to "force include" specific SNPs as tags. Additionally, iHAP displays the results obtained graphically for intuitive interpretation and includes alternative sets of tag SNPs attained. In essence, iHAP is a practical tool that can be used to analyze HapMap data for the selection of candidate targets in genotyping studies.

Availability and requirements

Project name: iHAP (integrated haplotype analysis pipeline)

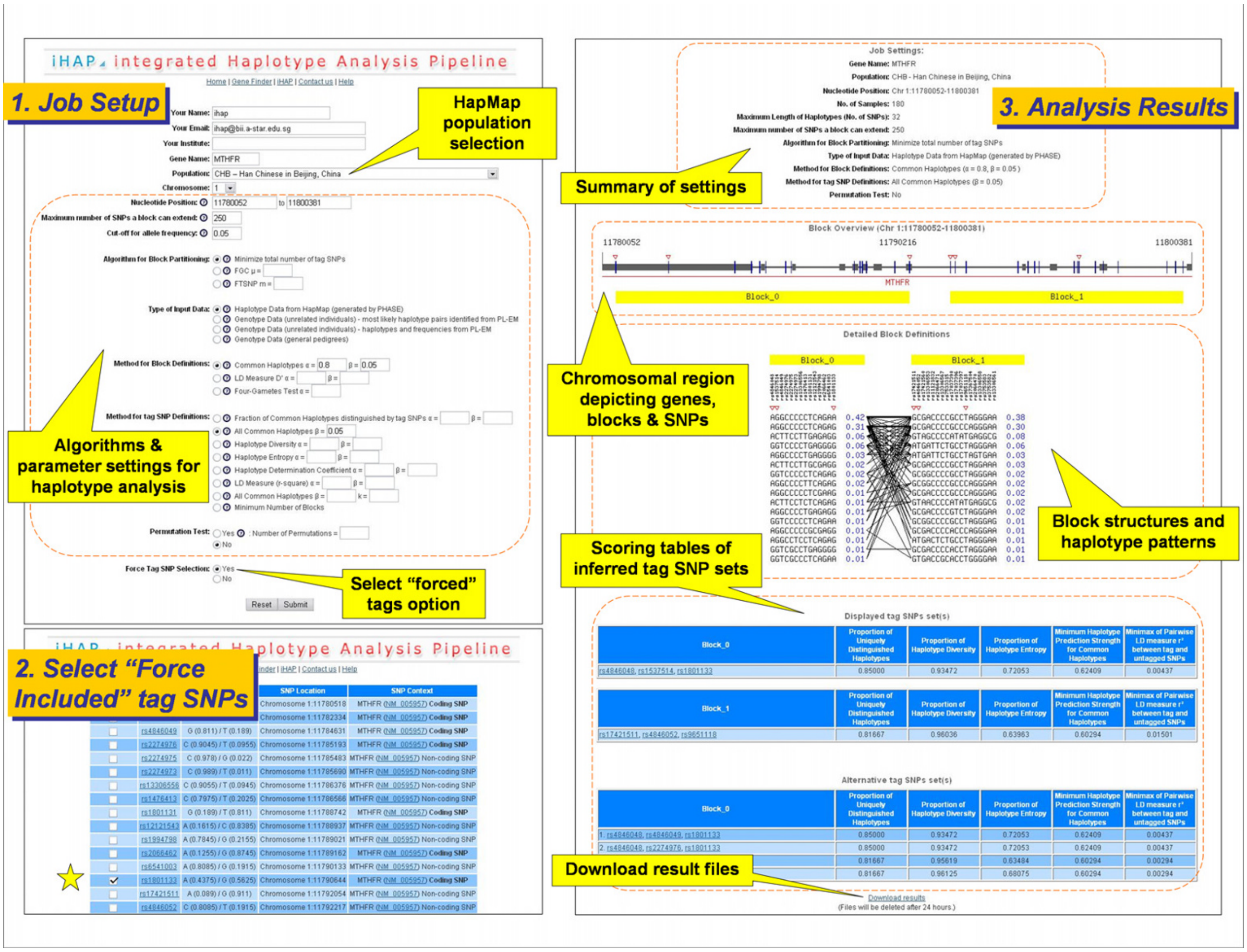
Project home page: http://ihap.bii.a-star.edu.sg

Operating system: Solaris (or any other OS that supports Apache, MySQL and PHP)

Programming language: PHP (with GD library)

Other requirements: MySQL, Apache HTTP Server

License: none



1. Job Setup
HapMap population selection
Algorithms & parameter settings for haplotype analysis
Select “forced” tags option
2. Select “Force Included” tag SNPs
3. Analysis Results
Summary of settings
Chromosomal region depicting genes, blocks & SNPs
Block structures and haplotype patterns
Scoring tables of inferred tag SNP sets
Download result files
iHAP integrated Haplotype Analysis Pipeline
Home | Gene Finder | iHAP | Contact us | Help
Reset
Submit
11780052
11790216
11800381
MTHFR
Block_0
Block_1
Download results

Your Name: ihap
Your Email: ihap@bii.a-star.edu.sg
Your Institute:
Gene Name: MTHFR
Population: CHB – Han Chinese in Beijing, China
Chromosome: 1
Nucleotide Position: 11780052 to 11800381
Maximum number of SNPs a block can extend: 250
Cut-off for allele frequency: 0.05

Algorithm for Block Partitioning: ● Minimize total number of tag SNPs ○ FGC μ = ○ FTSNP m =

Type of Input Data: ● Haplotype Data from HapMap (generated by PHASE) ○ Genotype Data (unrelated individuals) - most likely haplotype pairs identified from PL-EM ○ Genotype Data (unrelated individuals) - haplotypes and frequencies from PL-EM ○ Genotype Data (general pedigrees)

Method for Block Definitions: ● Common Haplotypes α = 0.8 β = 0.05 ○ LD Measure D' α = β = ○ Four-Gametes Test α =

Method for tag SNP Definitions: ○ Fraction of Common Haplotypes distinguished by tag SNPs α = β = ● All Common Haplotypes β = 0.05 ○ Haplotype Diversity α = β = ○ Haplotype Entropy α = β = ○ Haplotype Determination Coefficient α = β = ○ LD Measure (r-square) α = β = ○ All Common Haplotypes β = k = ○ Minimum Number of Blocks

Permutation Test: ○ Yes : Number of Permutations = ● No

Force Tag SNP Selection: ● Yes ○ No

			SNP Location	SNP Context
			Chromosome 1:11780518	MTHFR (NM_005957) Coding SNP
			Chromosome 1:11782334	MTHFR (NM_005957) Coding SNP
□	rs4846049	G (0.811) / T (0.189)	Chromosome 1:11784631	MTHFR (NM_005957) Coding SNP
□	rs2274976	C (0.9045) / T (0.0955)	Chromosome 1:11785193	MTHFR (NM_005957) Coding SNP
□	rs2274975	C (0.978) / G (0.022)	Chromosome 1:11785483	MTHFR (NM_005957) Non-coding SNP
□	rs2274973	C (0.989) / T (0.011)	Chromosome 1:11785690	MTHFR (NM_005957) Non-coding SNP
□	rs13306556	C (0.9055) / T (0.0945)	Chromosome 1:11786376	MTHFR (NM_005957) Non-coding SNP
□	rs1476413	C (0.7975) / T (0.2025)	Chromosome 1:11786566	MTHFR (NM_005957) Non-coding SNP
□	rs1801131	G (0.189) / T (0.811)	Chromosome 1:11788742	MTHFR (NM_005957) Coding SNP
□	rs12121543	A (0.1615) / C (0.8385)	Chromosome 1:11788937	MTHFR (NM_005957) Non-coding SNP
□	rs1994798	A (0.7845) / G (0.2155)	Chromosome 1:11789021	MTHFR (NM_005957) Non-coding SNP
□	rs2066462	A (0.1255) / G (0.8745)	Chromosome 1:11789162	MTHFR (NM_005957) Coding SNP
□	rs6541003	A (0.8085) / G (0.1915)	Chromosome 1:11790133	MTHFR (NM_005957) Non-coding SNP
☑	rs1801133	A (0.4375) / G (0.5625)	Chromosome 1:11790644	MTHFR (NM_005957) Coding SNP
□	rs17421511	A (0.089) / G (0.911)	Chromosome 1:11792054	MTHFR (NM_005957) Non-coding SNP
□	rs4846052	C (0.8085) / T (0.1915)	Chromosome 1:11792217	MTHFR (NM_005957) Non-coding SNP

Job Settings:
Gene Name: MTHFR
Population: CHB - Han Chinese in Beijing, China
Nucleotide Position: Chr 1:11780052-11800381
No. of Samples: 180
Maximum Length of Haplotypes (No. of SNPs): 32
Maximum number of SNPs a block can extend: 250
Algorithm for Block Partitioning: Minimize total number of tag SNPs
Type of Input Data: Haplotype Data from HapMap (generated by PHASE)
Method for Block Definitions: Common Haplotypes (α = 0.8, β = 0.05)
Method for tag SNP Definitions: All Common Haplotypes (β = 0.05)
Permutation Test: No

Block Overview (Chr 1:11780052-11800381)

Detailed Block Definitions

Block_0		Block_1	
AGGCCCCCTCAGAA	0.42	GCGACCCCGCCTAGGGAA	0.38
AGGCCCCCTCAGAG	0.31	GCGACCCCGCCCAGGGAA	0.30
ACTTCCTTGAGAGG	0.06	GTAGCCCCATATGAGGCG	0.08
GGTCCCCTGAGGGG	0.06	ATGATTCTGCCTAGGGAA	0.06
AGGCCCCTGAGGGG	0.03	ATGATTCTGCCTAGTGAA	0.03
ACTTCCTTGCGAGG	0.02	GCGACCCCGCCTAGGAAA	0.03
GGTCCCCCTCAGAG	0.02	GCGGCCCCGCCTAGGGAA	0.02
AGGCCCCTTCAGAG	0.02	GCGGCCCCGCCCAGGGAA	0.02
AGGCCCCCTCGAAG	0.01	GCGACCCCGCCCAGGGAG	0.02
ACTTCCTCTCAGAG	0.01	GTAACCCCATATGAGGCG	0.02
AGGCCCCTGAGAGG	0.01	GCGACCCCGTCTAGGGAA	0.02
GGTCCCCCTCAGAA	0.01	GCGGCCCCGCCTAGGGAG	0.01
AGGCCCCCGCGAGG	0.01	GCGACCCCACCCAGGGAA	0.01
AGGCCTCCTCAGAG	0.01	ATGACTCTGCCTAGGGAA	0.01
GGTCGCCTGAGGGG	0.01	GCGACCCCACCTAGGGAA	0.01
GGTCGCCCTCAGAA	0.01	GTGACCGCACCTGGGGAA	0.01

Displayed tag SNPs set(s)

Block_0	Proportion of Uniquely Distinguished Haplotypes	Proportion of Haplotype Diversity	Proportion of Haplotype Entropy	Minimum Haplotype Prediction Strength for Common Haplotypes	Minimax of Pairwise LD measure r² between tag and untagged SNPs
rs4846048, rs1537514, rs1801133	0.85000	0.93472	0.72053	0.62409	0.00437

Block_1	Proportion of Uniquely Distinguished Haplotypes	Proportion of Haplotype Diversity	Proportion of Haplotype Entropy	Minimum Haplotype Prediction Strength for Common Haplotypes	Minimax of Pairwise LD measure r² between tag and untagged SNPs
rs17421511, rs4846052, rs9651118	0.81667	0.96036	0.63963	0.60294	0.01501

Alternative tag SNPs set(s)

Block_0	Proportion of Uniquely Distinguished Haplotypes	Proportion of Haplotype Diversity	Proportion of Haplotype Entropy	Minimum Haplotype Prediction Strength for Common Haplotypes	Minimax of Pairwise LD measure r² between tag and untagged SNPs
1. rs4846048, rs4846049, rs1801133	0.85000	0.93472	0.72053	0.62409	0.00437
2. rs4846048, rs2274976, rs1801133	0.85000	0.93472	0.72053	0.62409	0.00437
	0.81667	0.95619	0.63484	0.60294	0.00294
	0.81667	0.96125	0.68075	0.60294	0.00294

(Files will be deleted after 24 hours.)

Figure 2
Typical workflow of iHAP. The iHAP resource was used to analyze the MTHFR gene with gastric cancer related SNP (rs1801133) "force included" as tag.

Any restrictions to use by non-academics: On request and citation

Authors' contributions

The iHAP resource was conceptualized and developed by CMS, BHY, ET and YY. While CMS and BHY prepared the data and wrote the programs, ET and YY evaluated suitable backend haplotype analysis tools. KBL and GR supervised the project and provided guidance while YPL facilitated the process. CMS, BHY, ET and YY drafted the manuscript and all authors have read and approved the final manuscript.

Acknowledgements

We are grateful to Professor Sun Fengzhu from the Department of Biological Sciences, University of Southern California for permitting us to use HapBlock as the backend haplotype analysis tool. We are also grateful to Mr Danny Chuon from the Web Services team of Bioinformatics Institute (BII) for his technical support in this work. This work is supported by the Biomedical Research Council of the Agency for Science, Technology and Research (A*STAR) of Singapore.

References

1. Miao X, Xing D, Tan W, Qi J, Lu W, Lin D: **Susceptibility to gastric cardia adenocarcinoma and genetic polymorphisms in methylenetetrahydrofolate reductase in an at-risk Chinese population.** *Cancer Epidemiol Biomarkers Prev* 2002, **11(11):**1454-1458.
2. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA: **Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium.** *American journal of human genetics* 2004, **74(1):**106-120.
3. **A haplotype map of the human genome.** *Nature* 2005, **437(7063):**1299-1320.
4. Sawyer SL, Mukherjee N, Pakstis AJ, Feuk L, Kidd JR, Brookes AJ, Kidd KK: **Linkage disequilibrium patterns vary substantially among populations.** *Eur J Hum Genet* 2005, **13(5):**677-686.
5. Ahmadi KR, Weale ME, Xue ZY, Soranzo N, Yarnall DP, Briley JD, Maruyama Y, Kobayashi M, Wood NW, Spurr NK, Burns DK, Roses AD, Saunders AM, Goldstein DB: **A single-nucleotide polymor-**

phism tagging set for human drug metabolism and transport. *Nat Genet* 2005, **37(1):**84-89.

6. Bakker PIWD, Graham RR, Altshuler D, Henderson BE, Haiman CA: **Transferability of tag SNPs to capture common genetic variation in DNA repair genes across multiple populations: Hawaii.** Edited by: Russ B. Altman, A. Keith Dunker, Lawrence Hunter, Tiffany Murray, Klein TE. World Scientific Publishing Co. Pte. Ltd; 2006:478 -4486.
7. Magi R, Kaplinski L, Remm M: **The whole genome tagSNP selection and transferability among HapMap populations: Hawaii.** Edited by: Altman RB, Dunker AK, Hunter L, Murray T, Klein TE. World Scientific Publishing Co. Pte. Ltd.; 2006:535 -5543.
8. Thorisson GA, Smith AV, Krishnan L, Stein LD: **The International HapMap Project Web site.** *Genome research* 2005, **15(11):**1592-1593.
9. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics (Oxford, England)* 2005, **21(2):**263-265.
10. Ding K, Zhang J, Zhou K, Shen Y, Zhang X: **htSNPer1.0: software for haplotype block partition and htSNPs selection.** *BMC bioinformatics [electronic resource]* 2005, **6:**38.
11. Greenspan G, Geiger D: **High density linkage disequilibrium mapping using models of haplotype block variation.** *Bioinformatics (Oxford, England)* 2004, **20 Suppl 1:**I137-I144.
12. Zhang K, Jin L: **HaploBlockFinder: haplotype block analyses.** *Bioinformatics (Oxford, England)* 2003, **19(10):**1300-1301.
13. **Genome Variation Server** [http://gvs.gs.washington.edu/GVS/]
14. Conde L, Vaquerizas JM, Dopazo H, Arbiza L, Reumers J, Rousseau F, Schymkowitz J, Dopazo J: **PupaSuite: finding functional single nucleotide polymorphisms for large-scale genotyping purposes.** *Nucleic acids research* 2006, **34(Web Server issue):**W621-5.
15. Hemminger BM, Saelim B, Sullivan PF: **TAMAL: an integrated approach to choosing SNPs for genetic studies of human complex traits.** *Bioinformatics (Oxford, England)* 2006, **22(5):**626-627.
16. Zhang K, Qin Z, Chen T, Liu JS, Waterman MS, Sun F: **HapBlock: haplotype block partitioning and tag SNP selection software using a set of dynamic programming algorithms.** *Bioinformatics (Oxford, England)* 2005, **21(1):**131-134.
17. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D: **Efficiency and power in genetic association studies.** *Nature genetics* 2005, **37(11):**1217-1223.
18. Zhang K, Deng M, Chen T, Waterman MS, Sun F: **A dynamic programming algorithm for haplotype block partitioning.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99(11):**7335-7339.
19. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA: **Haplotype tagging for the identification of common disease genes.** *Nature genetics* 2001, **29(2):**233-237.
20. Nothnagel M, Furst R, Rohde K: **Entropy as a measure for linkage disequilibrium over multilocus haplotype blocks.** *Human heredity* 2002, **54(4):**186-198.
21. Stram DO, Haiman CA, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Pike MC: **Choosing haplotype-tagging SNPS based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study.** *Human heredity* 2003, **55(1):**27-36.
22. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, Roskin KM, Schwartz M, Sugnet CW, Thomas DJ, Weber RJ, Haussler D, Kent WJ: **The UCSC Genome Browser Database.** *Nucleic acids research* 2003, **31(1):**51-54.