SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci

Yong Yong SHI[1,2], Lin HE[2,3,*]

[1]*Bio-X Life Science Research Center, Shanghai Jiao Tong University, Shanghai, China;*
[2]*Institute for Nutritional Sciences, Shanghai Institute of Biological Sciences, Chinese Academy of Sciences, Shanghai, China;*
[3]*NHGG, Bio-X Center, Shanghai Jiao Tong University, Shanghai, China*

ABSTRACT

In multiloci-based genetic association studies of complex diseases, a powerful and high efficient tool for analyses of linkage disequilibrium (LD) between markers, haplotype distributions and many chi-square/p values with a large number of samples has been sought for long. In order to achieve the goal of obtaining meaningful results directly from raw data, we developed a robust and user-friendly software platform with a series of tools for analysis in association study with high efficiency. The platform has been well evaluated by several sets of real data.

Keywords: software, linkage disequilibrium, haplotype analysis, genetic association study.

INTRODUCTION

At present, Gene based association study is a widely used method to study the genetic etiology of complex diseases. Since there are huge numbers of candidate genes, susceptible polymorphism sites and repetition studies in different samples, a user-friendly software platform with high efficient algorithms is urgently needed for researchers. To meet this need, therefore, we recently developed such a platform with our suggested name "SHEsis", which was placed on the website http://www.nhgg.org/analysis/ (Fig. 1).

METHODS

Linkage disequilibrium (LD): We calculated Lewontin's D' (|D'|) [1] and r^2 between each pair of genetic markers.

Haplotype analysis: a Full-Precise-Iteration (FPI) algorithm was used in haplotype reconstruction and frequency estimation inner randomly chosen samples. If there are two SNP loci, the FPI algorithm is based on the equation N(11)=2N(11/11)+N(12/11)+N(11/12)+P[(11/22)|(XX)]

*N(XX). Here, N(11) means the number of "11" haplotype; N(12/11) means the number of samples who carry "12" haplotype on one chromosome and "11" on another, etc; N(XX) is the number of samples carry both "1/2" genotype at the two loci, which is ambiguous for haplotype recognition; $P[(11/22)|(XX)] = \dfrac{N(11/22)}{N(11/22)+N(12/21)}$. As N(11/22) is linked with N(11) and N(22), we could find out the answer by iteration.

Case control study: Monte carlo simulation test, normal chi-square test and odds ratio test were all implemented for alleles and genotypes on single locus and multi-loci haplotypes.

RESULTS AND DISCUSSION

This platform supports all kinds of currently-used genetic markers of chromosomes and has capacity of analyzing the correctly formatted input parameters in seconds. It is quite user-friendly and easy-handling.

In haplotype analysis, we created an FPI algorithm, which could reconstruct ambiguous haplotypes and estimate haplotype frequencies in the given random sample set. The more samples are used in analysis, the more accuracy it will show. The platform can estimate haplotype frequency individually in controls and in samples to give the results of both single haplotype and a global data

*Correspondence: Lin HE, Shanghai Jiao Tong University, Bio-X Center, Hao Ran Building, 1954 Hua Shan Road, Shanghai 200030, China; or Institute for Nutritional Sciences, SIBS, Chinese Academy of Sciences, 294 Taiyuan Road, Shanghai 200031, China.
Tel & fax: 86-21-62822491; E-mail: helin@nhgg.org*

Fig. 1 The interface of our web-based analysis platform.

automatically. Haplotypes with low frequencies could be collected together to give one result in the case-control analysis.

In order to detect the reliability of our platform, we evaluated it with several published scientific papers[2-6] and obtained exactly the same result but very robust rate. Compared with the tools used in cited papers[2-6], our platform shows the advantage of easier handling and the capability of analyzing complicated data. Therefore, it could become a powerful and useful tool for studies of complex diseases in the near future. The SHEsis software platform is now available for free at the above website address. And we also will go on updating its power and capacity.

ACKNOWLEDGMENTS

REFERENCES

1 Lewontin R. On measures of gametic disequilibrium. Genetics 1988; **120:**849-52.
2 Chen WY, Shi YY, Zheng YL, et al. Case-control study and transmission disequilibrium test provide consistent evidence for association between schizophrenia and genetic variation in the 22q11 gene ZDHHC8. Hum Mol Genet 2004; **13:**2991-5.
3 Shi Y, Zhao X, Yu L, et al. Genetic structure adds power to detect schizophrenia susceptibility at SLIT3 in the Chinese Han population. Genome Res 2004; **14:**1345-9.
4 Zhao X, Shi Y, Tang J, et al. A case control and family based association study of the neuregulin1 gene and schizophrenia. J Med Genet 2004; **41:**31-4.
5 Guo S, Shi Y, Zhao X, et al. No genetic association between polymorphisms in the AMPA receptor subunit *GluR4* gene (GRIA4) and schizophrenia in the Chinese population. Neurosci Lett 2004; **369:**168-72.
6 Yang MS, Yu L, Guo TW, et al. Evidence for association between single nucleotide polymorphisms in T complex protein 1 gene and schizophrenia in the Chinese Han population. J Med Genet 2004; **41:**e63.