ISSN:0974-7230

Journal of Computer Science & Systems Biology

The International Open Access Journal of Systems Biology and Applied Computer Science

Available online at: OMICS Publishing Group (www.omicsonline.com)

Digital Object Identifier: http://dx.doi.org/10.4172/jcsb.1000073

Computer Science
Systems Biology

Kumar et al. J Comput Sci Syst Biol 2011, 4:2
http://dx.doi.org/10.4172/jcsb.1000073

Research Article Open Access

Computational Analysis of SNPs in 10 kb Region of Human Chromosome 1

Gupta Manish Kumar[1]*, Nutan Prakash[2], Chaturvedi Pragya[3] and Misra Krishna[3]

[1]Department of Bioinformatics, University Institute of Engineering and Technology, Chhatrapati Shahu Ji Maharaj University, Kanpur -208024, India
[2]Department of Biotechnology, Shree M & N Virani Science College, Rajkot, Gujarat
[3]Center of Biomedical Magnetic Resonance, SGPGIMS, Luck now-India

Abstract

As a result of human genome project a large burst of genomic data comes. Researchers are trying to correlate this sequenced data to find out variations, which will help to study the effect of variations on disease progression. Single nucleotide polymorphism is one of the genetic markers which are most widely used in genetic association studies of a population. SNPs are DNA sequence variations that occur when a single nucleotide (A, T, C, or G) in the genome sequence is altered. SNPs found within a coding sequence are of particular interest to researchers because they are more likely to alter the biological function of a protein. Occasionally, SNPs can cause disease and can be used to search and isolate diseased gene. The SNPs found in this region and its linkage disequilibrium analysis to find out the effect of SNPs found and there correlation. However it is much easier, cheaper and faster than in vitro analysis, computational analysis will provide an insight to probable disease causing SNPs having some functional value which can be assayed in vitro. Present computational analysis is to find out SNPs in the chromosome1.

Keywords: SNPs; LD; LD analysis; ht-SNPs

Introduction

With the advent of human genome project there is large extent of sequenced data in the public databases. These data must be used in order to characterize these sequences and to find out functional aspect of DNA sequences rather than merely store it in a database as characters. To find out the functional aspect the first one is to characterize sequence then find out the sequence variations. The variations in the DNA sequences of genomes of different population may pose an effect in its functional prospects. The most common DNA variations are SNPs (Single nucleotide polymorphisms) which are single base variation in different genomes. These are found to have major roles in drug disposition, pharmacogenomics and in disease susceptibility. Most of the SNPs found are "silent" i.e. they do not cause any functional differences. But many others are found to be responsible for differential response. The most common method of detecting SNPs is Genotyping the DNA sequences. However it is not possible to conduct population study reproducibly because it is difficult to conduct such in vitro experiments again and again in a population. Bioinformatics provide different computational tools and methods to find out probable positions of SNPs which will further help the in vitro experiments to be done. Linkage disequilibrium analysis can also be done to find out association between these SNPs which can help further in fine scale mapping, population studies and association studies.

Materials and Methods

A 10-kb region on human chromosome 1 was selected, further we have retrieved DNA sequences of a population from Popset database which is a database of DNA sequences used in population study. After retrieval the sequences, multiple sequence alignment is done to find out variations. SNPs are find out using a SNUFER software and linkage disequilibrium analysis is done to find out the effect of these SNPs on phenotype and there correlation (Supplementary files included).

Result and Discussion

SNPs found in the 10 kb region of human chromosome 1 has unknown function in the different populations. The distribution of SNPs in different population showed that Some SNPs are found to be common in all populations but some are specifically distributed in populations. Some of the population has more number of SNPs while others have less. Hence it may happen that most of the SNPs may not contribute toward any functional effect. Only few of them are responsible for the functional effect. LD analysis is done to find out association between these SNPs which showed that only few of the SNPs are associated and play a functional role.

Conclusion

SNPs are most common DNA sequence variations found in the human genome. Variations among the locus having SNPs found in all groups. Some of the SNPs locus is found common to all groups for example the position 310 which has found to have SNPs in all Groups. Some other SNPs are specific for a region which is showing that there is difference between the distributions of SNPs in different Geographical regions. It is also been concluded that some of the groups have large no of SNPs while some has only few. For example finish sequences are analyzed and only one SNP has been found in this group. However, about 150 SNPs are found in group having French sequences. It can be concluded from above results that there is difference between distribution of SNPs as well as number of SNPs. LD analysis has been done in order to find out association between the Different SNPs found in this Region in different population. Most of the results are not found to be significant because the parameters taken to analyze SNPs are not in the range of significant values (>0.5). However many of them are showing the values of D' and r^2 between 0.0-0.3 and only few SNPs are found to be associated with other which are tabulated as following:

***Corresponding author:** Gupta Manish Kumar, Department of Bioinformatics, University Institute of Engineering and Technology, Chhatrapati Shahu Ji Maharaj University, Kanpur -208024, India, E-mail: manishkumar.bi@gmail.com

Received March 22, 2011; **Accepted** April 12, 2011; **Published** April 14, 2011

Citation: Kumar GM, Prakash N, Pragya C, Krishna M (2011) Computational Analysis of SNPs in 10 kb Region of Human Chromosome 1. J Comput Sci Syst Biol 4: 033-034. doi:10.4172/jcsb.1000073

1. The pair wise LD D' measure is calculated to find out association between SNPs. D' plot of Nigerian sequences showed that single SNP whose value is in between 0.3-0.7 which is at 12th loci may be associate.
2. The pair wise LD measure of D' of South African sequences showed that most of the SNPs do not lie in the significant range. There is only single SNP showing association which lies at 12th position having value between 0.3-0.7.
3. D' measure of Swedish sequences showed that there are three SNPs having significant values found to be associated at 12th and 14th positions.
4. In the pair wise LD measure of r^2 in Swedish sequences show that there are two SNPs having significant values and SNP at 14th position is showing association with SNPs at 9th and 12th position.
5. In the D' measure of Kenya sequences three SNPs are shown to associated which are at 22nd and 19th position having values greater than 0.3.
6. r^2 measure of Kenya sequences showed that the SNPs at 19th and 22nd position are associated having significant values of r^2.

As the function of this region is not known the SNPs discovered and the association between them may be helpful to find out the effect of this gene on phenotypes. These SNPs may also contribute toward any genetic disease in future.

References

1. Giardina E, Pietrangeli I, Martone C, Asili P, Predazzi I, et al. (2007) In silico and in vitro comparative analysis to select, validate and test SNPs for human identification. BMC Genomics. 1471: 2164-8-457.
2. Mansur MA, Cardozo GP, Santos EV, Marins M (2008) SNUFER: A software for localization and presentation of single nucleotide polymorphisms using a Clustal multiple sequence alignment output file. Bioinformation 3: 63-64.
3. Fernandez F, Curtain RP, Colson NJ, Ovcaric M, Macmillan J (2007) Association analysis of chromosome 1 migraine candidate genes. BMC Medical Genetics 29: 8-57.
4. Yu N, Zhao Z, Fu YX, Sambuughin N, Ramsay M, et al. (2001) Global Patterns of Human DNA Sequence Variation in a 10-kb Region on Chromosome 1. Mol Biol Evol 18: 214-222.
5. Erichsen HC, Chanock SJ (2004) SNPs in cancer research and treatment. Br J Cancer 90: 747-751.
6. Beiswanger C, Coppock D, Mintzer J, Toj L, Greenberg J, et al. (2003) The International HapMap Project. Nature 426: 789-796.
7. Ding K, Zhou K, He F, Shen Y (2003) LDA–a java-based linkage disequilibrium Analyzer. Bioinformatics 19: 2147–2148.
8. Beiswanger C, Coppock D, Leach A (2007) A second generation human haplotype map of over 3.1million SNPs. Nature 449: 851-861.
9. Goldstein DB, Weale ME (2001) Linkage disequilibrium holds the key. Curr Biol 11: R576-R579.
10. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, et al. (2001) Linkage disequilibrium in the human genome. Nature 411: 199-204.
11. Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human Genome. Nat rev genet 3, 299-309.
12. Pritchard JK, Przeworski M (2001) Linkage Disequilibrium in Humans: Models and Data. Am J Hum Genet 69: 1-14.