Original Research Paper

Ten Bioinformatics Tools for Single Nucleotide Polymorphisms Detection

[1]**Mahin Ghorbani and** [2]**Hamed Karimi**

[1]*Department of Biotechnology, Fergusson College, F.C. Road, Pune, Maharashtra 411004, India*
[2]*Department of Computer Science, Payam Noor University of Farokh-Shahr, Farokh-Shar, Chaharmah va Bakhtiari, Iran*

Corresponding author:
Mahin Ghorbani
Department of Biotechnology,
Fergusson College, F.C. Road,
Pune, Maharashtra 411004, India
Email: ghorbani_mahin@yahoo.ca

Abstract: Single Nucleotide Polymorphisms (SNPs) as unique genetic variations are widely used as biomarkers for drug designing and development for individualized therapy. SNP genotyping techniques such as SBE with MS detection, SBE with fluorescence intensity detection, The FP-TDI, Pyrosequencing and The TaqMan 5' nuclease assay are low-speed, time consuming and very expensive. To overcome such disadvantages, bioinformatics tools are emerged as alternative methods for detection of SNPs in different populations for different diseases such as cancer, diabetes, heart diseases, Down syndrome and other genetic related disorders to help in rapid and low-cost and high accuracy detection methods. In this review we have listed 10 bioinformatics tools for SNP detection to introduce them for the researchers who are interested in SNP detection required studies in order to help them in saving time and cost and enhancement of accuracy of their results. The bioinformatics tools are selected in this review are poly phred, HGBASE, Human chromosome 21 cSNP database, Dog Genome SNP Database, JSNP database, Polybays, novoSNP, HaploSNPer, PupaSNP and QualitySNPng, which have their own specific performance for detection of SNPs. Researcher can choice appropriate programs on the basis of their research's need and application. We hope this review can be helpful and integrates scientific information with bioinformatics tools in drug discovery processes and other studies involved SNP detection and lead to short time-low cost and high accuracy biomarker detection in drug designing and development process.

Keywords: SNP Detection, Single Nucleotide Polymorphism, Biomarker, Bioinformatics Tools, Drug Designing

Introduction

Single nucleotide polymorphism are defined as unique genetic variation within members of the same species which present in a single base change in a DNA sequence (Voisey and Morris, 2008). Their ability to address the problem of why there are different responses to the same drug among individuals of a population, made them to act as genetic marker in detection of some genetic associated disorders such as cancer, diabetes, Asthma, heart diseases, in genome-wide association studies of large sample population (Rai *et al*., 2010; Sripichai and Fucharoen, 2007). Based on the types of SNP, the degree of their significance in contribution to diseases are classified in three classes*, first harmless or

Innocent variation: They only show change in phenotype, second *harmful or Dangerous variation*: They are directly associated with genetic disorders like cancer, diabetes, heart diseases and hemophilia and third is *Latent (Hidden) variation*: They are not dangerous on their own but are susceptible to become dangerous under certain condition and found in coding and regulatory regions e.g., Susceptibility to lung cancer (Giacomini *et al*., 2007). According to their location within DNA sequence they are classified in 6 groups: Type I: Coding, synonymous, non conservative, Type II: Coding, synonymous, conservative, Type III: Coding, synonymous, Type IV: Non-coding. 5'-UTR, Type V: Non-coding. 3'-UTR and Type VI: Other non-coding (Cargill *et al*., 1999).

Bioinformatics Tools as Alternative Methods for SNP Detection

Detection of SNPs as molecular markers plays an important role in drug designing and development and genome-wide association studies (Giacomini *et al*., 2007). Laboratory genotyping methods of SNPs such as SBE with MS detection, SBE with fluorescence intensity detection, The FP-TDI, Pyrosequencing and the TaqMan 5' nuclease assay are time consuming and low speed processes and demand very expensive instruments (Chen and Sullivan, 2003). Thus automated tools are generated as alternative method to overcome such disadvantages and to cover weakness of genotyping methods (Chen and Sullivan, 2003; Batley *et al*., 2007). Although several automated tools for detection of SNP are developed over last decay but still some researchers are not aware of generation and availability and use of such tools (Batley *et al*., 2007). The aim of this review is that to introduce different bioinformatics tools for SNP detection to researches who are interested in using such tools to utilize their benefits in cost and time and accuracy (Chen and Sullivan, 2003).

Bioinformatics Tools for SNP Detection

Poly phred: (http://droog.gs.washington.edu/polyphred/) (Direct link to access the program). It is an integrated program to identify heterozygous locations for SNP. The program together with three other programs is used to compare florescence based sequence across DNA traces obtained from different individuals to detect heterozygous sites for SNPs. The contributed programs are as *Phred* (is responsible for detection of potential heterozygous sites for SNP using base call and peak characterization)*, Phrad* (provides assemblies or sequence alignment) *and Consed* (responsible for editing, scanning, interpretation and visualization of results) (Sachidanandam *et al*., 2001). For instance, the project of "*Automating the identification of DNA variations using quality-based fluorescence re-sequencing*: *Analysis of the human mitochondrial genome*" used Poly phred program in which the program detected 378 variants containing 29 novel variants and 2 heteroplasmic sites from 12 sequenced genome from human mitochondrial genome as source (Marth *et al*., 1999).

Human Genic Bi-Allelic Sequence (HGBASE): (http://hgbase.interactiva.de/) (Direct link to access the program). It is a database containing polymorphisms are located in human intra-genic sequence (from promoter to end of transcription) and used by users for study of SNPs located in intra-genic sequences and are associated with genetic diseases (Sarkar *et al*., 1998).

Human chromosome 21 cSNP database: (http://csnp.unige.ch/) (Direct link to access the program). It is web based program used for detection of SNPs are located in Chromosome 21 of human genome. As chromosome 21 is associated with several genetic disorders like down syndrome so identification of several cSNP are associated with such disorders can be detected using this Database. Result of completion of sequencing chromosome 21 disclosed 225 genes. Thus providing detection of potential Single Nucleotide Polymorphism within cDNA (cSNP) for appropriate projects (Deutsch *et al*., 2001).

Dog Genome SNP Database (DoGSD): (http://dogsd.big.ac.cn/) (Direct link to the program). It is Dog Genome SNP Database which provides information about already identified SNPs in dog/wolf related genetic diseases. Dog and wolf share the same diseases with human due to the same ancestor. Dog/wolf's diseases which are common to human or transfer from them to human have significant role in human's disease studies so providing database for Dog's SNPs plays crucial role in detection, similar SNPs in human genome (Bai *et al*., 2014).

JSNP database: (http://snp.ims.u-tokyo.ac.jp/) (Direct link address to the program). It provides database SNPs for different genetic related disorders in the Japanese population. This database searching program facilitates web based-fast search service for potential SNPs involved in different diseases among Japanese population, also it provides other information related searched SNP like it's location in chromosome, Its laboratory where it was detected, Its type and frequency, detail of electrophoresis of individual SNP can be found in database (Hirakawa *et al*., 2002).

Polybays: (http://clavius.bc.edu/~marth/PolyBayes/pages/main.html) (Direct link address to the program). The PolyBayes software is used to detect SNPs in redundant DNA sequences. By three steps, Creation of multiple sequence alignment, Identification of paralog and SNP detection (Marth *et al*., 1999).

novoSNP: (http://www.molgen.ua.ac.be/bioinfo/novosnp/) (Direct link address to the program). novoSNP is a program used for detection of SNP in SNP reconstructing/resequencing based researches. It's input is a reference sequence and a number of sequencing trace files and output will come in the form of a list of possible variations with a quality score (Weckx *et al*., 2005).

HaploSNPer: (http://www.bioinformatics.nl/tools/haplosnper/) (Direct link address to the program). It is another applicable SNP detection tool which produce three outputs: First, Showing the parameter's setting, second, listing information of clusters, haplotypes and statistic information of SNPs and third is to display each cluster with it's haplotypes, SNP and it's sequence alignment (Tang *et al*., 2008).

PupaSNP: (http://pupasnp.bioinfo.cnio.es/) (Direct link address to the program). It is another web snp detection tool which resequence SNPs that show

influence on conserved regions used by cellular machinery for gene processing such as intro/exon boundaries, predicted Transcription Factor Sites (TFBS) and amino acid changes of proteins (Conde *et al*., 2004).

QualitySNPng:
(http://www.bioinformatics.nl/QualitySNPng/) (Direct link address to the program). It is another software tool for detection of SNPs, designed in such a way that uses haplotype based scheme, it uses DNA sequences derived from next generation sequencing data. It performs on diploid as well as polyploidy species. Complete sequenced reference genome is not required by the tool (Nijveen *et al*., 2013).

Discussion

SNP genotyping techniques such as SBE with MS detection, SBE with fluorescence intensity detection, The FP-TDI, Pyrosequencing and The TaqMan 5' nuclease assay have shown some weakness in cost, throughput and time and high demand of instruments, technical support and experiences. To overcome such disadvantages, bioinformatics tools have occupied a place in biomedical research as alternative methods which provide an economic, high speed, high accuracy detection method for SNPs. Bioinformatics tools for SNP detection have been developed last decay and improving rapidly. Generation of these tools assume challenge for researchers who are not directly involved in program generation but requiring to use the tools for their researches. In this article we reviewed ten bioinformatics tools such as poly phred, HGBASE, Human chromosome 21 cSNP database, Dog Genome SNP Database, JSNP database, Polybays, novoSNP, HaploSNPer, PupaSNP and QualitySNPng involved in Single nucleotide polymorphism detection, to introduce them as alternative methods for SNP genotyping which are time consuming, expensive, require technical support and experimental experiences. By providing information about the name, direct address and performance of the tools, we hope to provide essential information about them to be a starting point to use them according research's appropriates in their projects to utilize their advantages.

Conclusion

Applying alternative tools of bioinformatics to genotyping techniques for detection of Single Nucleotide Polymorphisms leads in saving time and cost for researchers who are not interested in genotyping methods in their research concerned with SNPs. Thus to become familiar with such programs, will help them in finding appropriate program for their research like drug designing to provide individualized therapy. But using these tools for detection of SNPs still require sufficient knowledge and background of principle behind of each programs.

References

Bai, B., W.M. Zhao, B.X. Tang, Y.Q. Wang and L. Wang *et al*., 2014. DoGSD: The dog and wolf genome SNP database. Nucl. Acids Res. DOI: 10.1093/nar/gku1174

Batley, J., E. Jewell and D. Edwards, 2007. Automated discovery of single nucleotide polymorphism and simple sequence repeat molecular genetic markers. Meth. Mol. Biol., 406: 473-94. PMID: 18287708

Cargill, M., D. Altshuler, J. Ireland, P. Sklar and K. Ardlie *et al*., 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nat. Genet., 22: 231-8. DOI: 10.1038/10290

Chen, X. and P.F. Sullivan, 2003. Single nucleotide polymorphism genotyping: Biochemistry, protocol, cost and throughput. Pharmacogenom. J., 3: 77-96. DOI: 10.1038/sj.tpj.6500167

Conde, L., J.M. Vaquerizas, J. Santoyo, F. Al-Shahrour and S. Ruiz-Llorente *et al*., 2004. PupaSNP Finder: A web tool for finding SNPs with putative effect at transcriptional level. Nucleic Acids Res. PMID: 15215388

Deutsch, S., C. Iseli, P. Bucher, S.E. Antonarakis and H.S. Scott, 2001. A cSNP map and database for human chromosome 21. Genome Res., 11: 300-307. PMID: 11157793

Giacomini, K.M., C.M. Brett, R.B. Altman, N.L. Benowitz and M.E. Dolan *et al*., 2007. The pharmacogenetics research network: From SNP discovery to clinical drug response. Clin. Pharmacol. Ther., 81: 328-45. DOI: 10.1038/sj.clpt.6100087

Hirakawa, M., T. Tanaka, Y. Hashimoto, M. Kuroda and T. Takagi *et al*., 2002. JSNP: A database of common gene variations in the Japanese population. Nucleic Acids Res., 30: 158-62. PMID: 11752280

Marth, G.T., I. Korf, M.D. Yandell, R.T. Yeh and Z. Gu *et al*., 1999. A general approach to single-nucleotide polymorphism discovery. Nat. Genet., 23: 452-456. DOI: 10.1038/70570

Nijveen, H., M. van Kaauwen, D.G. Esselink, B. Hoegen and B. Vosman, 2013. QualitySNPng: A user-friendly SNP detection and visualization tool. Nucleic Acids Res., 41: W587-90. DOI: 10.1093/nar/gkt333

Rai, A.J., J. Yee and M. Fleisher, 2010. Biomarkers in the era of personalized medicine-a multiplexed SNP assay using capillary electrophoresis for assessing drug metabolism capacity. Scand. J. Clin. Lab. Invest., 242: 15-8. DOI: 10.3109/00365513.2010.493355

Sachidanandam, R., D. Weissman, S.C. Schmidt, J.M. Kakol and L.D. Stein *et al.*, 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature, 409: 928-933. DOI: 10.1038/35057149

Sarkar, C., F.R. Ortigão, U. Gyllensten and A.J. Brookes, 1998. Human genetic bi-allelic sequences, HGBASE, a database of intra-genic polymorphisms. Mem. Inst. Oswaldo Cruz., 93: 693-4. DOI: 10.1590/S0074-02761998000500023

Sripichai, O. and S. Fucharoen, 2007. Genetic polymorphisms and implications for human diseases. J. Med. Assoc. Thai., 90: 394-8. PMID: 17375650

Tang, J., J.A. Leunissen, R.E. Voorrips, C.G. van der Linden and B. Vosman, 2008. HaploSNPer: A web-based allele and SNP detection tool. BMC Genet., 9: 23-23. DOI: 10.1186/1471-2156-9-23

Voisey, J. and C.P. Morris, 2008. SNP technologies for drug discovery: A current review. Curr. Drug. Discov. Technol., 5: 230-5. DOI: 10.2174/157016308785739811

Weckx, S., J. Del-Favero, R. Rademakers, L. Claes and M. Cruts *et al.*, 2005. novoSNP, a novel computational tool for sequence variation discovery. Genome Res., 15: 436-42. DOI: 10.1101/gr.2754005