An Evaluation of Single Nucleotide Polymorphism Detection Programs

Rosalind Cutts **Ravi Mehta** **Suzanne Orton**

Philip East **Robert Gill**

Philippe Sanseau **Paul Matthews**
rjc33965@glaxowellcome.co.uk

Genome Informatics and Analysis Group, Genomics Unit, UK Discovery Genetics,
Glaxo Wellcome Medicines Centre,
Gunnels Wood Road, Stevenage SG1 2NY, United Kingdom

Keywords: SNP detection, automation

1 Introduction

Detection and mapping of single nucleotide polymorphisms (SNPs) is becoming of increasing importance for genetic analysis. It has been shown that a dense set of SNP markers can be used to identify disease susceptibility genes by association studies, and is believed that knowing the nature and location of SNPs will dramatically change genetic analysis. Several SNP detection algorithms have been developed by different groups, however there has been no benchmarking of these using sequence data where experimentally confirmed SNPs are present. Here we describe an evaluation of four SNP detection programs, three publicly available and one developed in-house, for accuracy in SNP detection. The results of the evaluation are presented to demonstrate the relative performance of these programs and their accuracy in SNP detection using in-house generated sequence data sets of genomic DNA.

2 Method and Results

Several SNP detection programs have been developed by various groups, and all are used to determine whether variations represent true polymorphisms or are due to base calling or sequencing errors. Here we describe an evaluation of 4 SNP detection programs based on ability to identify all true SNPs correctly whilst not calling false positives using in-house sequence data. Three of the programs are publicly available, TraceDiff [1], Polybayes [2] and Polyphred [3], and these were compared to the current in-house method of SNP detection called SAPI (Semi Automatic Polymorphism Identification). Each program was benchmarked using a number of sequence data sets produced in-house where experimentally confirmed SNPs had been identified using either Restriction Fragment Length Polymorphism analysis, Taqman assays (PE Biosystems) and/or oligo-ligation assays.

References

[1] Bonfield, J.K., Rada, C. and Staden, R., Automated detection of point mutations using florescent sequence trace subtraction, *Nucleic Acids Research*, 26:3404–3409, 1998 .

[2] Marth, G.T., Korf, I., Yandell, M.D., Yeh, R.T., Zhijie, G., Zakeri, H., Stitziel, N.O., Hillier, L., Kwok, P-Y. and Gish, W.R., A general approach to single-nucleotide polymorphism discovery, *Nature Genetics*, 23:452–456, 1999.

[3] Nickerson, D.A., Tobe, V.O. and Taylor, S.L., Polyphred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing, *Nucleic Acids Research*, 25:2745–2751, 1997.