

SEQUENCE POLYMORPHISM IDENTIFICATION WITHIN THE *TP53* GENE OF CANCER SAMPLES: A comparison of traditional Sanger- and next-generation sequencing as tools for supporting drug development.

Tom Burr, Richard Dixon, Simon Mayes, and Cliff Murray

Source BioScience, Nottingham, UK

Summary

- Source BioScience undertook a comparison of the sequence polymorphisms identified within the human *TP53* gene from cancer samples using traditional Sanger sequencing approaches versus next-generation sequencing data derived from our Illumina Genome Analyzer (GA) platform.
- Using various human cancer samples, 3 exonic single nucleotide polymorphisms were identified in the samples by Sanger sequencing.
- Sequence data from the Illumina genome analyzer validated these 3 single nucleotide polymorphisms and identified a further 42 putative mutations.

Introduction

The p53 protein is a transcription factor encoded by the *TP53* gene found on chromosome 17. It is an important regulator of the cell cycle and functions as a tumour suppressor important in preventing cancer. Acquired mutations associated with the *TP53* gene are found in approximately 50% of cancers. This study reports on the effectiveness of two different DNA sequencing approaches for the detection of polymorphisms (SNPs and acquired mutations) within the same cancer sample. Exonic PCR products from tissue samples were subjected to traditional Sanger DNA sequencing and also to next-generation sequencing (NGS) using the Illumina Genome Analyzer. Detection of polymorphisms within the NGS data sets utilised the Maq algorithm.

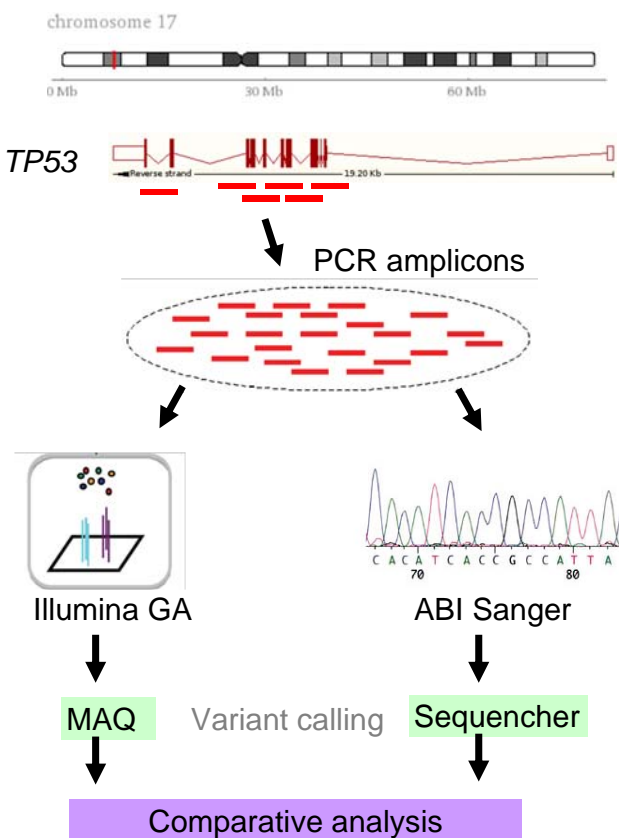


Figure 1: An illustration of the experimental design.

Results

- A summary of the Maq alignment of the Illumina sequence reads against the *TP53* reference genome sequence (Exons 2-11) is shown in **Figure 2**. A minimum coverage depth of 15,000-20,000 reads was obtained across all the exons that were sequenced (exons 2-11).
- All of the mutations detected by Sanger sequencing were also detected by the Illumina Genome Analyzer (GA) data.
- A further 42 putative exonic single nucleotide polymorphisms were detected by the Maq algorithm in the Illumina sequence data from these samples.

Methods

Six genomic intervals were amplified using DNA from human cancer samples and PCR reactions. The six regions containing exons 2 to 11 of the *TP53* gene were mixed in equimolar amounts and concatenated prior to library preparation to obtain even coverage. Preparation of the libraries and sequencing were performed using the Illumina sequencing platform (Genome Analyzer, Illumina Inc) following the manufacturer's instructions. Image analysis and base calling were performed using Firecrest and Bustard algorithms from the Illumina analysis pipeline software (v1.3). The purity filtered sequence reads (38bp) were aligned to the human genome sequence using the Maq program (v0.7.1). Nucleotide variants were calculated using the Maq tools 'cns2snp' and 'SNPfilter' with default parameters.

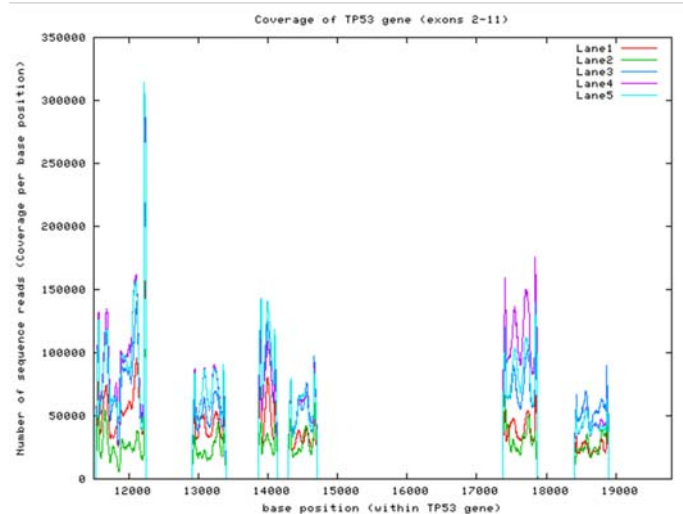


Figure 2: Sequence coverage of the *TP53* exons with purity filtered Illumina short reads. The 6 amplicons sequenced contained exons 2-4, 5-6, 7, 8, 9-10 and exon 11.

Conclusions

Our results show that the NGS data validated the polymorphisms detected by the Sanger sequencing approach in the same samples. However, NGS data analysis also identified a large number of additional polymorphisms undetected by the Sanger approach in the same samples. This enhanced sensitivity of NGS suggests that use of this technology can dramatically increase knowledge of the somatic genetics of the *TP53* gene. This has implications for the identification of potential therapeutic targets and patient populations that may or may not respond to current or future therapies.