A ReSequencing primer set for 3,000 genes implicated in cancer genetics

Robert Nutter, Leslie Johnston-Dow, Jon Sorenson, H. Carey Gire, Stephen Glanowski, Carl Fosler, Jackie Hoglund, Mary Ann Rydland, Pei-Hong Shen, Ken Glasser, Craig Forbes, Indresh Singh, Vikram Sathineni, Lini Wu, Bryant Small, Patt Dunn, Nathan Edwards, Chia-Chien Chang, Jingwei Ni, Ben Jones, Kerry Woodford, Paolo Vatta, Raisa Loboda, Gary Wang, Brian Murphy, Pamela St. John, Steve Ferreira, Andrew Parker, Evangeline Gonzalez, Carol Kosman, Quynh Doan, Lin Zuo Pham, Oliver Bell **Applied Biosystems, Foster City, CA and Rockville, MD USA**

The Applied Biosystems 3730xl and 3730 DNA Analyzers includes patented technology licensed from Hitachi, Ltd. as part of a strategic partnership between Applied Biosystems and Hitachi, Ltd., as well as patented technology of Applied Biosystems.

For Research Use Only. Not for use in diagnostic procedures.

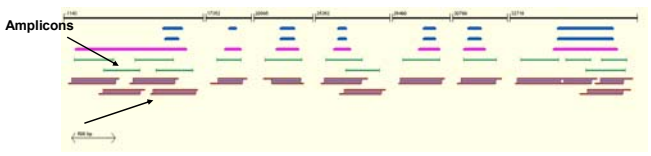
Abstract

Presentation Number: P085

The completion of a reference sequence for the human genome and improvements in high-throughput sequencing technology, including the Applied Biosystems 3730xl DNA analyzer and the BigDye® Terminators v3.1 Cycle Sequencing Kit, have motivated the development of easier solutions for quickly resequencing human genes. We report here work towards the development of a complete and validated resequencing workflow for high-throughput resequencing of the promoter regions, exon regions, and flanking intronic regions for 3,000 genes implicated in cancer. This workflow includes pre-designed primer sequences for amplicons covering these regions, protocols for PCR amplification and cycle sequencing, and software analysis tools specifically tailored to resequencing. The resequencing workflow takes advantage of the latest capillary electrophoresis technology for DNA sequencing. With this development, sequencing remains the most accurate method for rapid SNP discovery and comprehensive SNP screening.

Applied Biosystems Resequencing Primer Sets:

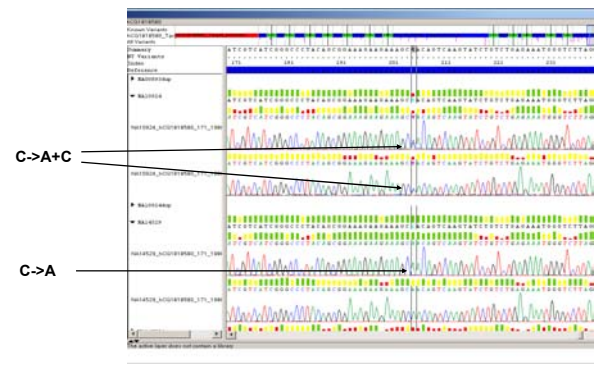
MLF1 example



F and R Sequence and Region with Double Coverage

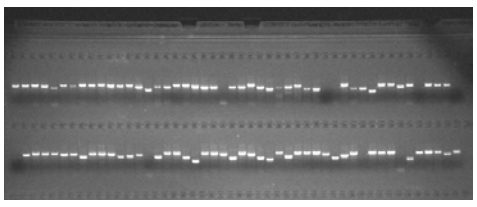
- Primer sets designed using proprietary Applied Biosystems primer design pipeline
  - MLF1 is shown above as an example
    - Amplicon coverage is shown in green, Sequenced coverage is shown by the triple red and blue lines
- Full gene coverage for 3000 Cancer genes including:
  - all verified transcripts for each gene -> 'supertranscript'
  - all exons (coding and non-coding) plus 5' UTRs
  - 50bp intron flanking each exon
  - 'Regulatory region' = 1kb 5' of exon 1 start
  - Annotation of regions where there is < 100% coverage
  - Sequence consensus validation with QVs>30
- Validated protocol
  - Genomic DNA QC ->PCR ->Sequencing -> Analysis
- Software for results generation and analysis
  - Data collection protocol for 3100 and 3730
  - SeqScape® v2.1 software and Gene-Specific Project templates

Automatic Detection of Base Changes in Different Specimens



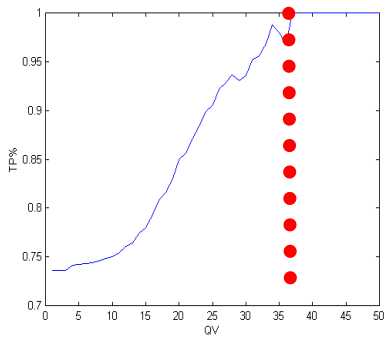
SeqScape® software will automatically shows both the position and identity of a base change. When mutations occur they will lead to either heterozygous alleles or homozygous mutant alleles. In a given project, some specimens may be heterozygous and others may be homozygous for each mutation. The SeqScape® software shows the mutation type for each specimen and allows the comparison of a given genotype across all specimens in a SeqScape ® project. Again, individual electropherograms from different biological specimens can be viewed to allow verification of basecalling. In this example, one of the specimens has a M mixed base (C+A) at the indicated position and the other specimen has a homozygous C->A mutation.

Agarose Gel of PCR Products



2% Agarose Gel of 96 PCR products generated using primers designed for resequencing. PCR products were generated according to our standard protocol. Excess primers and dNTPs were digested by treatment with Exonuclease III and Shrimp Alkaline Phosphatase. Ten percent of the PCR product from each reaction was run on the gel, stained with ethidium bromide and photographed. The picture shows nearly all of the primers were amplified and there is excellent uniformity in the amount of PCR product.

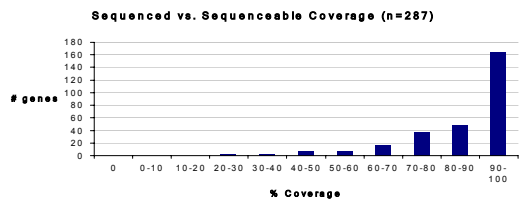
Heterozygote detection in SeqScape® v2.1 software



The key to any correct heterozygote basecalling and assignment of accurate quality values lies in the fundamental accuracy of basecalling. Above is shown an analysis of the correlation between Quality Values and accuracy of heterozygous basecalling. A dataset of sequences from an HLA locus with a known number and identity of heterozygous positions was analyzed with SeqScape® v1.1 software. Quality values were assigned to each base and correlated to the confidence that the mixed bases were called correctly. Quality values above 30 resulted in a 95% or greater confidence the mixed base was identified and called correctly. The quality value for the assembled sequences in our pilot experiment was >37. This graph shows the confidence that mixed bases were correctly identified in this experiment would be 97% or greater (red dotted line).

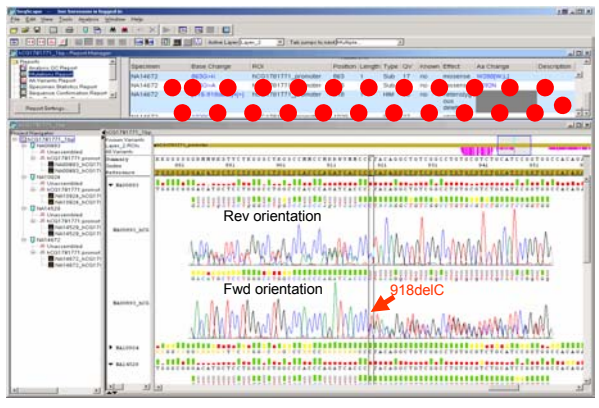
Sequenced vs. Sequenceable:

actual coverage vs. theoretical design for 332 gene pilot



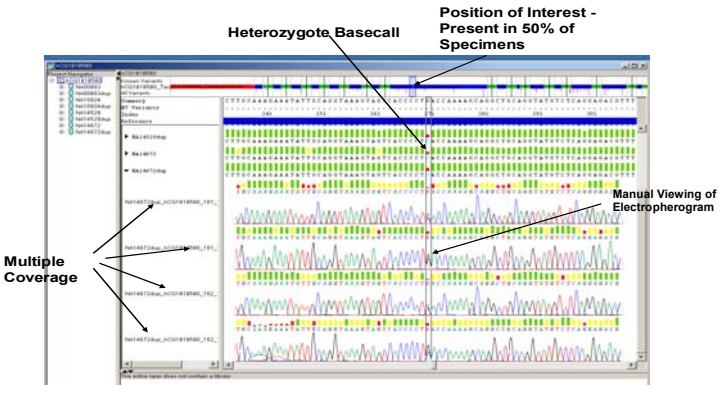
- The 332 gene pilot
  - Average of 18 amplicons per gene (largest=102 amplicons)
  - Average amplicon size of 539 bases
- Theoretical coverage for the design of the 332 gene pilot study was 91%
  - This has recently been improved to 96% for the 3000 cancer gene set.
- 89% of designed amplicons are sequenceable
  - Failures generally due to CpG islands, high GC content, homopolymer regions, low complexity repeats
  - This information has been folded into design of 3000 cancer gene set
- 74% of genes met expectation (80% coverage of sequenceable amplicons)

Detection of Heterozygous Indel Mutations (HIMs) with SeqScape® v2.1 software



Shown above is a heterozygous indel mutation (HIM) that was seen in the BAK (BCL2L7, CDN1) gene using the Resequencing Primer set for BAK. The mutation is correctly identified in SeqScape® v2.1 software and reported using the recommended HUGO nomenclature for HIMs. Applications using Applied Biosystems technology have been designed to provide high quality data assuring the high accuracy basecalling and heterozygote detection. Correct assignment of HIMs is made possible by the use of latest generation DNA sequencing technologies such as the AB 3730 Genetic Analyzer and the new AB KB basecaller coupled with BigDye® Terminator v3.1 chemistry. This systems approach provides high quality data assuring the accurate heterozygote detection required by any reliable HIM detection algorithm.

Heterozygote Detection



This data is from one of the 332 genes in our proof-of-principle experiment. This figure shows the Project View in SeqScape® software and shows the entire gene target, potential regions of interest and the number of specimens where the position of interest was found. This particular example is covered by two separate amplicons and provides independent confirmation of basecalling. A heterozygous base S (C+G) is shown in this example and the presence of this mixture can be verified by manual viewing of individual electropherograms. Due to the improved data quality resulting from the Applied Biosystems BigDye® Terminator v3.1 sequencing chemistry, more accurate basecalling of heterozygote positions is achieved.

Conclusions

- Resequencing primer sets have been designed using expert-reviewed datasets.
- Primers have been designed for all exons of all known transcripts plus a 1kb 'regulatory region' upstream of the first exon for 3000 genes implicated in cancer biology.
- Data from initial 332 genes shows 90% success in sequencing with ave. QVs =37.
- Analysis software designed to allow user to review, edit and generate reports on gene data.
- Enhanced heterozygote detection achieved due to optimized system performance.
- Complete protocol for genomic DNA QC, PCR, PCR clean-up, Sequencing, Analysis.
- Entire system supported by Applied Biosystems worldwide.
- 3000 cancer genes will be available for sale mid-2003.

Document #: 127MI15-01 http://docs.appliedbiosystems.com/search.taf

© 2002. Applied Biosystems. All rights reserved. AB (Design) Applera, PUP-5, PUP-6 and PUP-37 are trademarks and Applied Biosystems and BigDye are registered trademarks of Applera Corporation or its subsidiaries in the US and certain other countries.