

VariantSEQR™ System Process for the Discovery of Human Gene Variants

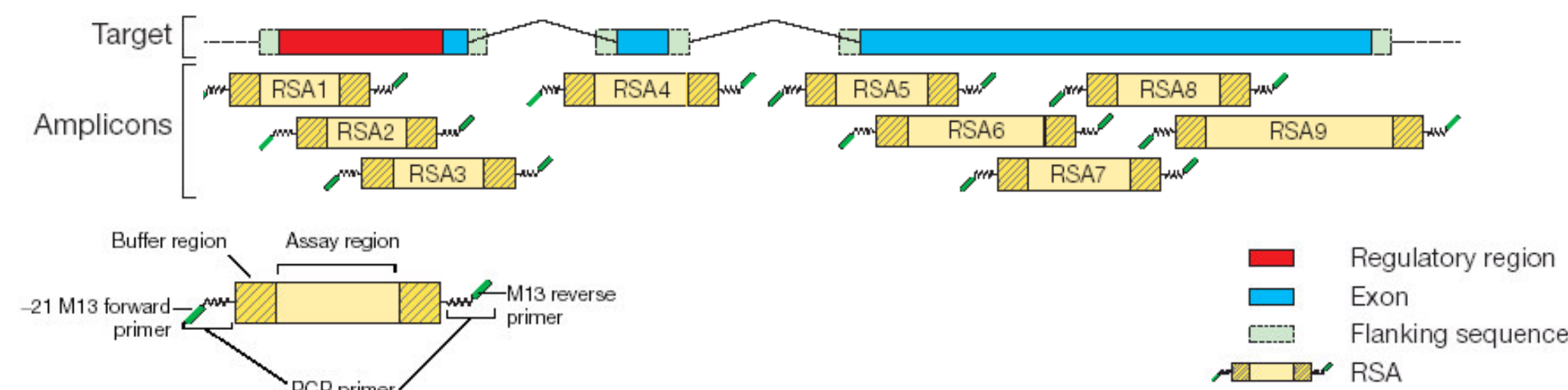


Diane Bond, Peter Ma, Rixun Fang, Primo Baybayan, Soumya Nidtha, Lilley Leong, Xiaohui Chen, Mary Ann Rydland¹, Xiaoying Lin¹, Karen Poulter, Jon Sorenson and Cheryl Heiner
Applied Biosystems, Foster City, CA 94404 and 1. Celera Genomics, 45 West Gude Drive, Rockville, MD 20850

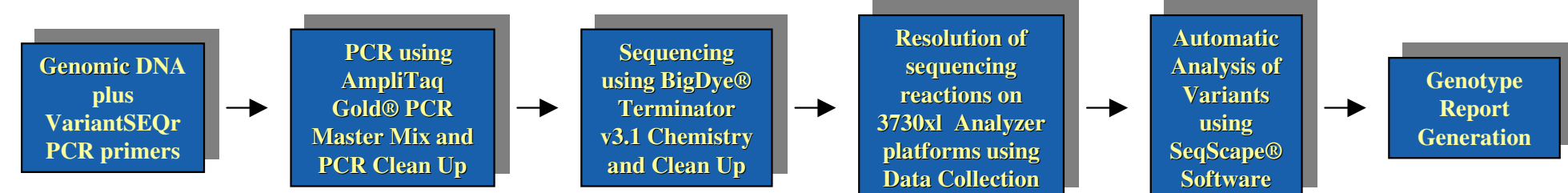
Abstract

With the completion of sequencing human and other genomes, scientists are now focusing on the discovery of variations among individuals and populations. The process of resequencing or mutational profiling may be expensive and tedious, including considerable time and effort on designing and validating the primers. Recently, we have developed an effective way to streamline the whole resequencing process, from primer design to the analysis and reporting of variants. The performance of hundreds of thousands of PCR primers has been analyzed, using millions of sequence trace files. This information was used to select primer sequences with a very high probability of generating high quality sequencing information. These primers contain flanking universal M13 forward and reverse primer sequences to streamline set-up of the sequencing reactions. Currently we have available more than 100,000 primer sequences covering over 3,000 genes in our VariantSEQR™ system product line, and other resequencing sets are under development. We will present typical results, and also show how sequencing results are affected by various difficult cases.

Primers designed for a typical Resequencing Set (RSS). The target region of the Resequencing Amplicons (RSA, shown in yellow) have been designed to provide complete coverage of promoter regions (red), intron/exon junctions (green/blue boundaries) and all exons (blue). The regions flanking the target regions (represented in green) are part of the amplicons but may not always have data of the desired high quality. Each PCR primer is tailed with priming sites for either the M13 universal forward or reverse sequencing primer to simplify sequencing reaction set-up and standardize sequencing reaction conditions.



Workflow of the VariantSEQR™ Resequencing System. The system has been designed to fit the workflow of a typical resequencing laboratory. Optimized protocols are provided for PCR amplification, PCR clean-up, sequencing reactions, and sequencing reaction clean-up. PCR primers are validated by a combination of laboratory and computational systems to the generation of genotype reports. New software allows integration of the 3100 and 3730 Data Collection software with SeqScape® Software. At the end of each run, sequence files are automatically basecalled and can be exported to SeqScape® Software for automatic trimming, alignment and assembly against a reference sequence. Using SeqScape® software, results can be easily reviewed and genotype reports generated.



PCR reaction:
gDNA* 1.0 µl
AmpliTaq Gold® PCR Mater Mix 5.0 µl
Water 0.4 µl
50% glycerol 1.6 µl
RSA primer mix (1.2 uM total) 2.0 µl
Total 10.0 µl
96C 5min, (94°C 30 sec, 60°C 45 sec, 72°C 45 sec) x 40 cycles
72°C 10min, 4C hold
* For human genes: 10ng/µl and

Clean-up PCR reaction: Exo-SAP digestion
PCR product 10 µl
ExoSAP-IT® (USB, PN 78200) 2 µl
Total 12 µl
37°C 30 min, 80°C 15 min, 4°C hold

Sequencing reaction:
Digested PCR product 2 µl
BigDye® Terminator v3.1 4 µl
M13 forward or reverse primer (3.2 uM) 1 µl
Water 3 µl
Total 10 µl
96°C 1 min, (96°C 10 sec, 50°C 2 sec, 60°C 4 min) x 25 cycles,
4°C hold

Clean-up sequencing reaction: 96-well ethanol precipitation
1. Mix 2.5 µl of 125 mM EDTA to each 10 µl sequencing reaction,
2. Add 30 µl of 100% ethanol and mix by inverting four times,
3. Incubate the plate a room temperature for 15 min,
4. Spin at 4,000 rpm for 30 min then invert spin at 500 rpm for 1 min,
5. Add 30 µl of 70% ethanol,
6. Spin at 4,000 rpm for 10 min, invert spin at 500 rpm for 1 min,
7. SpeedyVac dry and resuspend in 10 µl of Hi-Di™ Fromamide.

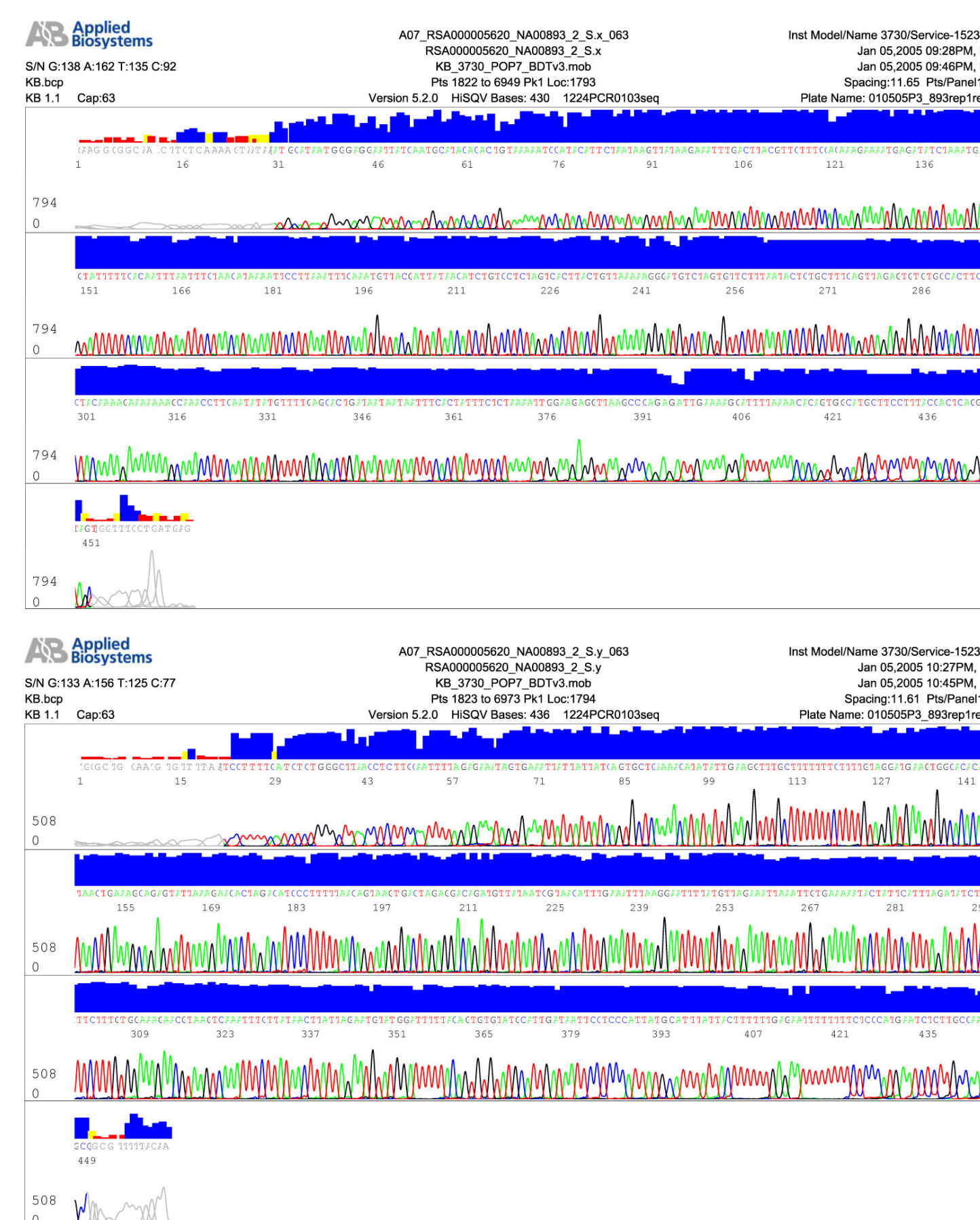
3730xl DNA Analyzer electrophoresis

Item	Setting
Polymer	POP-7™ polymer
Running buffer	Applied Biosystems 3730/3730xl Running Buffer with EDTA
Array	36 cm
Run file	RapidSeq_POP7_1
Mobility file	KB_3730_POP7_BDTv3.mob
Basecaller	KB.bcp

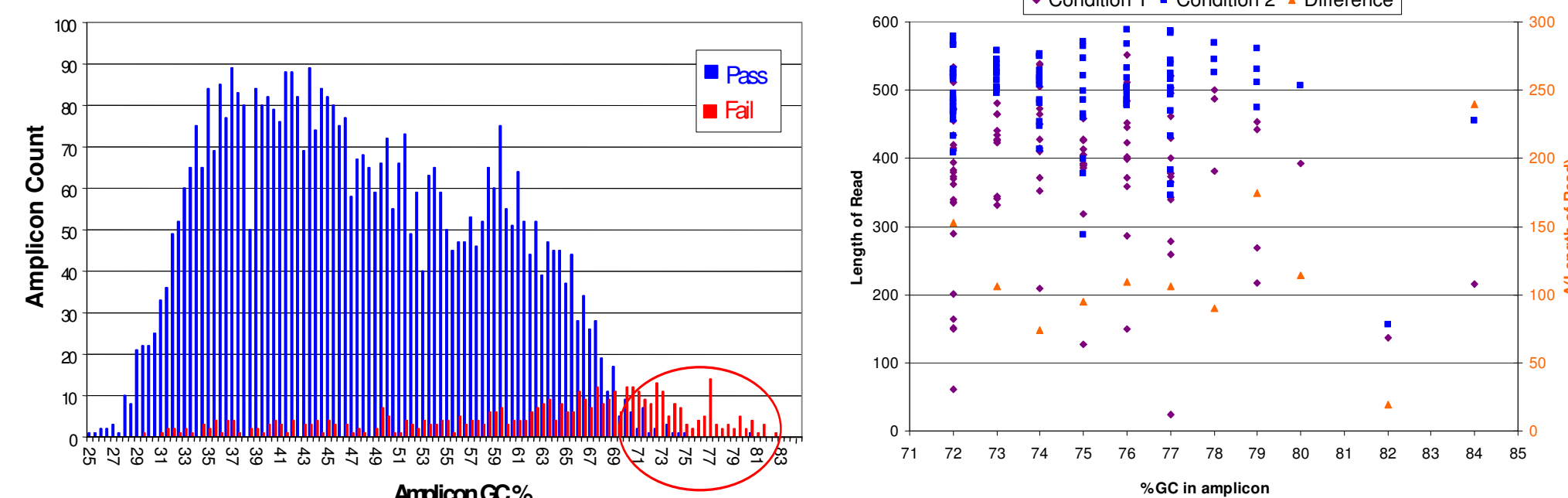
Primer design

Analysis of large datasets from several resequencing projects has allowed us to characterize features of sequences that are most likely to impact the ability to generate high quality data in both directions. Based on the results of this analysis, an algorithm was developed to predict primer performance.

Typical Results We present the typical sequencing result of an amplicon in the forward (top) and reverse (bottom) direction.

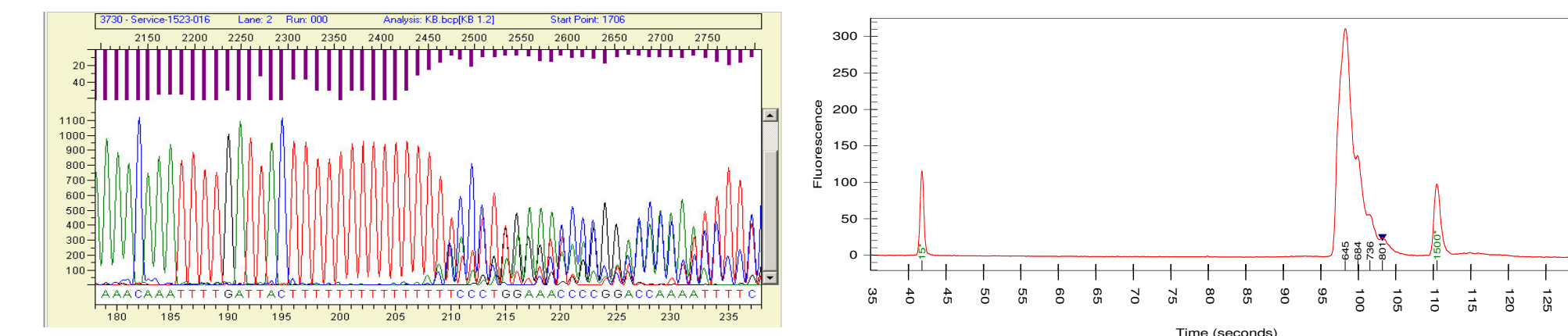


Effect of GC Content in PCR and Sequencing. Success in PCR and sequencing is related to GC content, when an amplicon is > 60% GC. Using standard PCR conditions, for amplicons with > 72% GC, successful sequence generation is rare (below, left). We are experimenting with new PCR conditions to improve the PCR success rate and sequencing length of read. A set of amplicons with GC content ranging from 72% to 84% were evaluated by using 20 gDNAs (below, right). These amplicons all failed under standard PCR conditions.

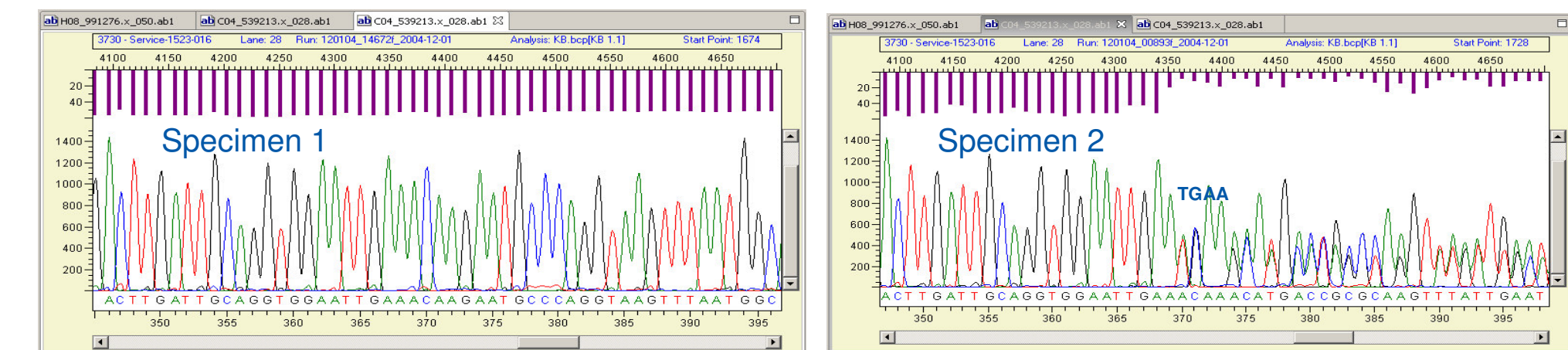


Difficult Cases

Effect of homopolymers on sequence quality This figure shows how a poly-T stretch in the amplicon interferes with the high quality sequence over the entire region (left). Further analysis of the amplicon using Agilent 2100 Bioanalyzer shows shoulders, indicating slippage during PCR generates multiple PCR products, as shown in the right panel. Retrospective analysis of data from large resequencing projects has allowed us to determine that homopolymer regions of less than 10 nucleotides are not likely to cause data quality problems.



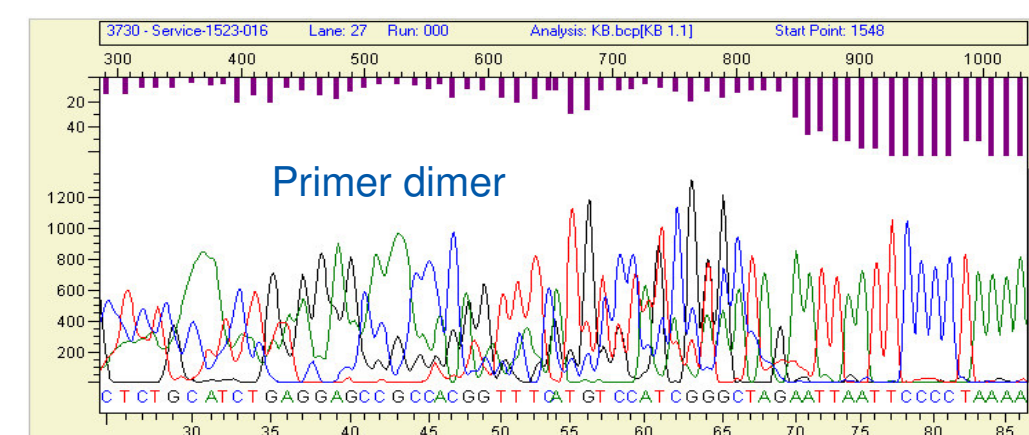
Detection of heterozygous indel mutations with VariantSEQR™ system Only sequence based resequencing systems have the ability to detect indels of any size. Indels produce clean sequence to the point where the deletion or insertion occurs, then exhibit double sequences after that. Specimen 2, right panel, has a heterozygous insertion of four bases (TGAA) at base 370.



Secondary sequence

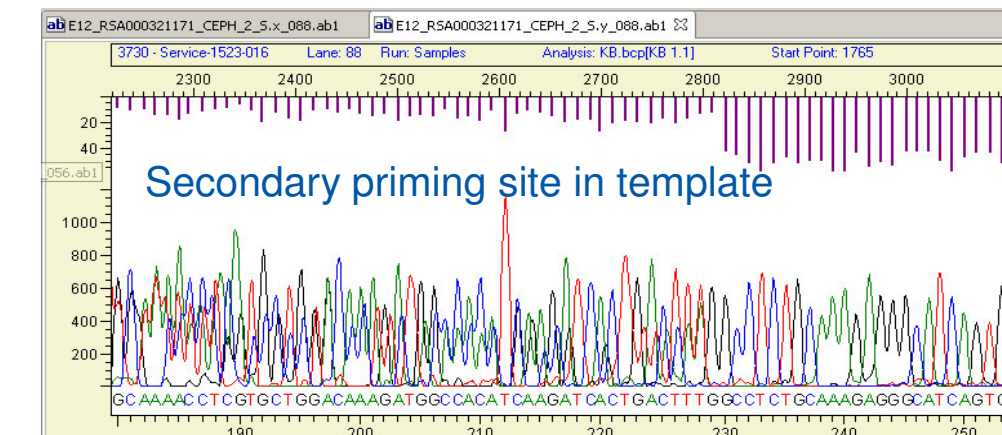
Primer dimer

The trace below shows two sequences for the first 70 bases. This is due to primer dimer, where one primer serves as template for the other primer during the PCR amplification. The sequence beyond the primer dimer region is clean.



Secondary priming site in template

Another cause of secondary sequence is a secondary priming site on the genomic DNA template. The second PCR product may be smaller (as shown below) or larger than the primary product. In this case, the primers need to be redesigned.



Analysis

For information on analysis, see poster by Anjali Pradhan: P86-M Mitochondrial DNA Data Analysis Using SeqScape® Software v2.5

Conclusions

The Applied Biosystems VariantSEQR™ Resequencing System provides full integration with existing AB CE platforms and reagents that enables scientists to focus on research instead of PCR primer design, validation and data analysis. The resequencing sets have been validated using a combination of laboratory investigation as well as the application of an expert-trained computational system. Not only does this eliminate the need for time consuming and expensive PCR primer validation, it also provides templates that will provide very high quality sequence data. Data analysis is greatly simplified by the use of SeqScape® Software. This analysis package uses gene content derived from the Celera Discovery System and public database and provided at no additional cost. This allows resequencing projects to be automatically basecalled, assembled and aligned against a reference sequence for review and report generation.

For Research Use Only. Not for use in diagnostic procedures. The PCR process is covered by patents owned by Roche Molecular Systems, Inc. and F. Hoffmann-La Roche Ltd. Applied Biosystems, SeqScape and BigDye are registered trademarks and AB(Design), VariantSEQR and Applera are trademarks of Applera Corporation or its subsidiaries in the U.S. and/or certain other countries. AmpliTaq Gold is a registered trademarks of Roche Molecular Systems, Inc. All other trademarks are the sole property of their respective owners.