

DNA Resequencing and Variant Identification of the CCL24 Gene



Denise M. Schmidt, Jennifer D. Tan, George A. Fry, Primo A. Baybayan, Quynh C. Doan, Anjali A. Pradhan, Sourmya S. Nidtha, and Thomas J. McElroy. Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA, USA, 94404

ABSTRACT

In the course of designing more than 200,000 primer pairs and generating more than 18 million sequence reads, bioinformatics specialists and R&D scientists who worked on the Applera Genome Initiative gained valuable experience in primer design, as well as insight into the genomic complexities and technical challenges associated with them. The 3130x and 3130xl Genetic Analyzers, when combined with the Applied Biosystems VariantSeq™ Resequencing System and SeqScape® Software v2.5 for seamless data analysis, provide the most robust and efficient resequencing system for customers who require low to medium throughput. The components of this integrated system simplify the rate-limiting steps found in current resequencing protocols, and offer a complete, cost-effective solution for laboratories performing either large or small resequencing studies, thus, revolutionizing DNA sequencing as a tool for the detection of human gene mutations.

A HIGH PERFORMANCE SYSTEM FOR SEQUENCING APPLICATIONS

The 3130 Series Genetic Analyzers are fully automated, high-performance, fluorescence-based, capillary electrophoresis systems that can analyze multiple samples simultaneously. Samples can be resequenced in 35 minutes when using the 16-capillary 3130xl system or the 4-capillary 3130 system with 3130 POP-7™ Polymer and the 36-cm capillary array. After the sequence data is collected, the KB™ Basecaller will automatically process the data and provide a length of read (LOR) greater than 500 base pairs (bp). Furthermore, the throughput of the 3130xl system, using the UltraSeq36_POP7 run module, can efficiently sequence up to 41 runs (656 samples) in a 24-hour period, generating high quality, high resolution data with minimal hands-on time.

The 3130 series system is designed for ease-of-use to maximize laboratory productivity while reducing the overall cost per sample. Now more than ever, researchers have the flexibility to choose one configuration for all their resequencing needs. The Automated Polymer Delivery System in the 3130 series system allows automatic polymer loading, which minimizes hands-on time and maintenance while maximizing performance. The system enables the use of 3130 POP-7 polymer, not only for the 36-cm capillary array, but also for the 50-cm and 80-cm capillary arrays. The run configurations, specific for POP-7 polymer, incorporate a higher temperature through the detection cell heater, which yields peaks of greater resolution, less run-to-run variability, and faster electrophoresis times than any other capillary electrophoresis system available on the market today.

KEY FEATURES OF THE 3130 SERIES SYSTEMS

Automated Polymer Delivery System	<ul style="list-style-type: none"> Ease to use Easy to maintain
New 3130 POP-7™ Polymer	<ul style="list-style-type: none"> One polymer & one array for all applications
Detection Cell Heater and 3130 POP-7 Polymer	<ul style="list-style-type: none"> Faster turn around times Longer Reads Better peak resolution



3130 POP-7™ Polymer Sequencing Run Modules for Variant Identification

The table below lists Data Collection Software v3.0 run modules that were used for the variant identification experiment.

Table 1. 3130 POP-7 Polymer Sequencing Run Modules

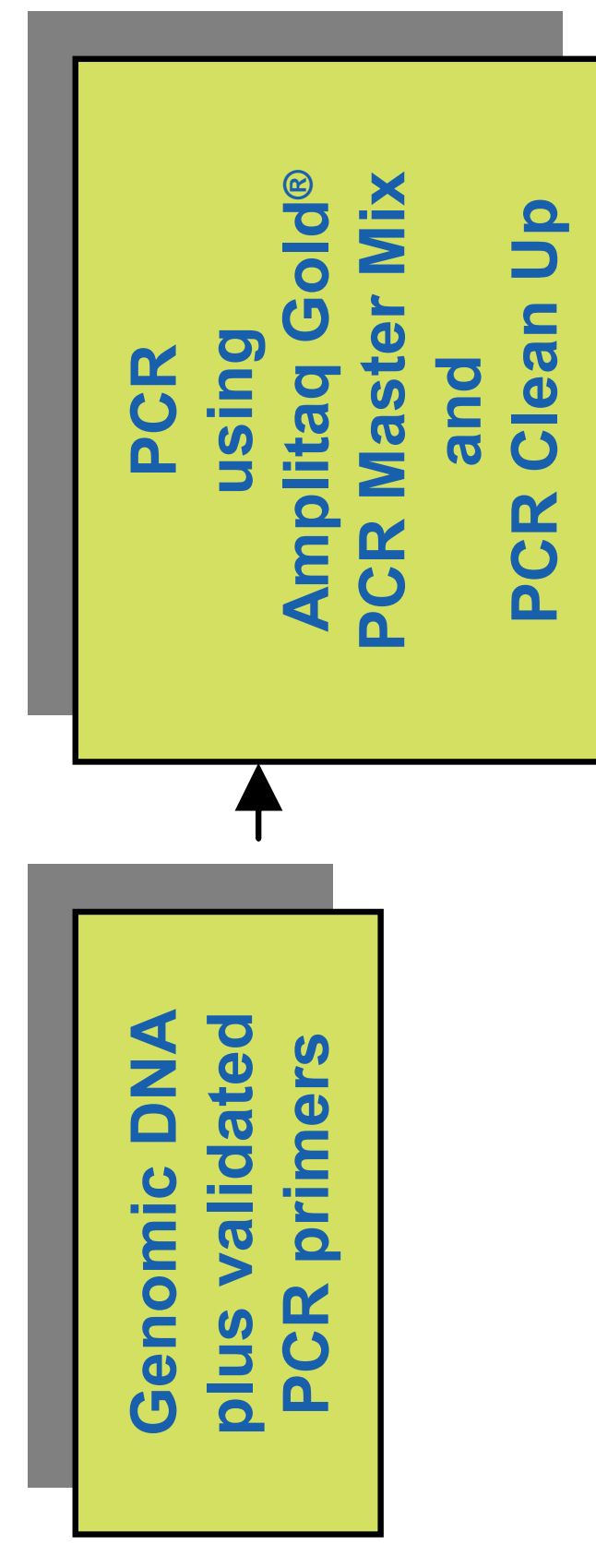
Sequencing Run Modules	Array Length (cm)	Samples in 24-hr		KB™ Basecaller QV ₂₀ LOR**
		3130 system	3130xl system	
UltraSeq36_POP7	36	164	656	500
RapidSeq36_POP7	36	96	384	600

*Sequencing Analysis Software v5.2 provides a metric Length Of Read (LOR), defined as the usable range of high-quality or high-accuracy bases determined by Quality Values (QV) generated by the KB Basecaller v1.2. The LOR is determined using a sliding window of 20 bases, which has an average QV greater than 20.

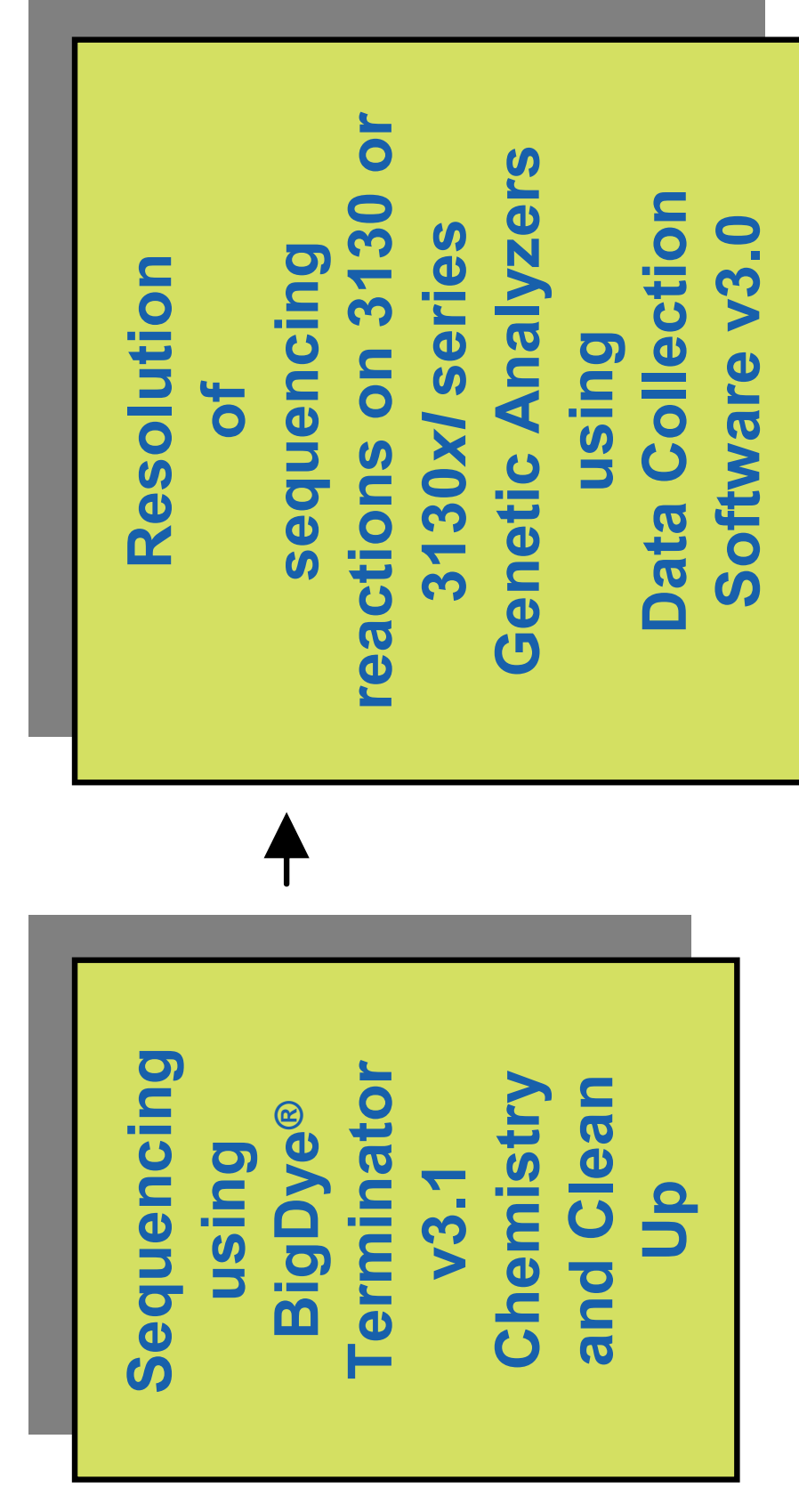
**98.5% basecalling accuracy, less than 2% N's

System Workflow for DNA Resequencing and Variant Identification

1. Genomic DNA Template Preparation and PCR Amplification



2. DNA Sequencing



3. Data Analysis

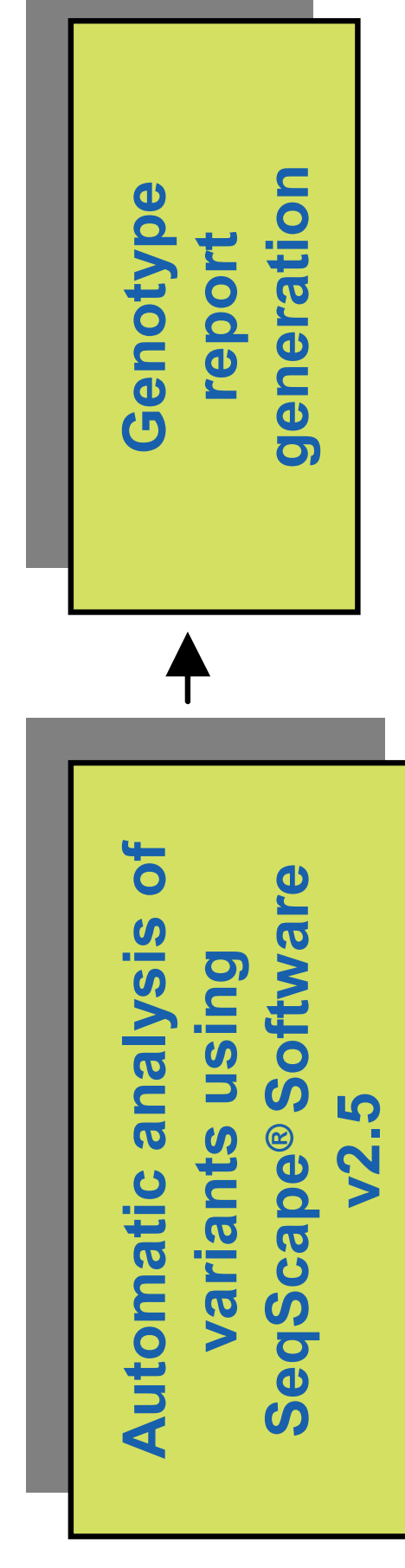
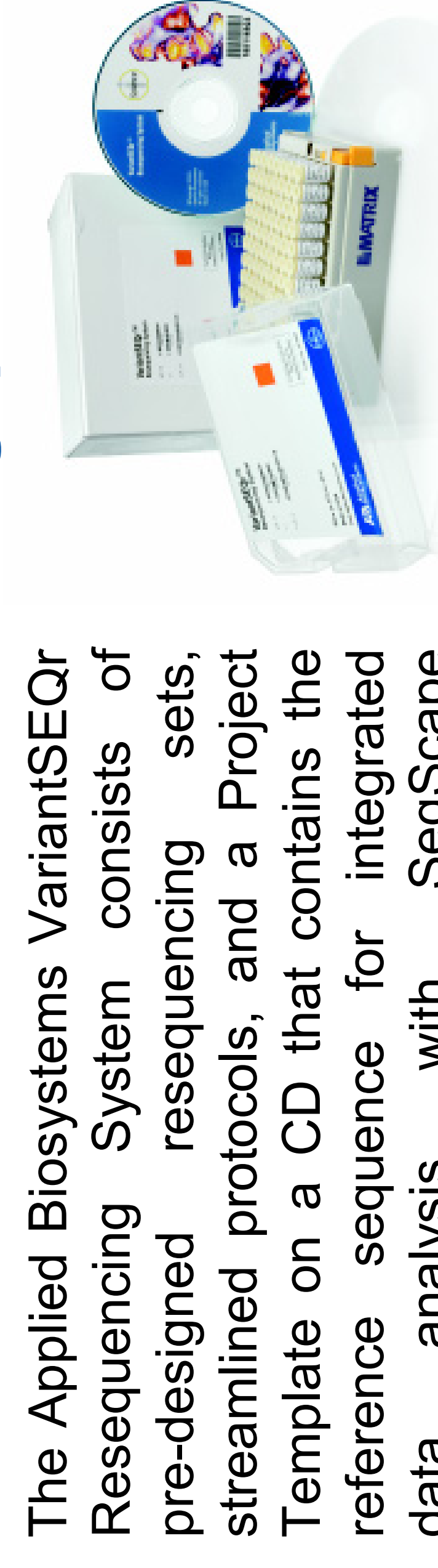


Figure 1. The system has been designed to fit the workflow of a typical resequencing laboratory. An optimized protocol provides complete integration of each step from PCR amplification using PCR primers validated by a combination of laboratory and computational systems, to the generation of genotype reports. An important new feature is the integration of the 3130 series system and Data Collection Software v3.0 with SeqScape software. At the end of each run, sequence files are automatically basecalled by SeqScape software where they are automatically trimmed, aligned, and assembled against a reference sequence. Results can then be easily reviewed and reports generated.

The VariantSeq™ Resequencing System for Low to Medium Throughput



The Applied Biosystems VariantSeq Resequencing System consists of pre-designed resequencing sets, streamlined protocols, and a Project Template on a CD that contains the reference sequence for integrated data analysis with SeqScape Software v2.5.

This system is designed to work seamlessly with the Applied Biosystems 3130 Series Genetic Analyzers, reagents, and SeqScape software for the study of human genes and other target regions. The goal is to discover variants within genes that may correlate to disease development and response to drug treatment.

SeqScape® Software v2.5 for Sequencing Applications

Applied Biosystems SeqScape® Software v2.5 is expressly designed for mutation profiling with robust algorithms, enhanced display capabilities, and detailed results reports. With a single mouse click, the software feeds the sequencing files into the analysis pipeline and generates detailed reports containing mutation information. Essential to the analysis pipeline are many robust algorithms integrated to ensure automated processing and accurate results from raw sequence data to the final mutation report. Unique to this tool are basecalling and consensus-calling algorithms that provide quality values for each base pair, sample, and mutation to enable easy distinction between poor and high quality data. Enhanced display capabilities reduce data review time with features such as hyperlinks between reported results and the actual base pairs to demonstrate the mutation.

RESULTS

The 3130 series systems provides seamless integration between the instrument and analysis software, ensuring automatic sample loading, generation of sequencing data, basecalling, and alignment of sequence and reference data. An integral part of successful resequencing, SeqScape Software analyzes data from both small- and large-scale projects.

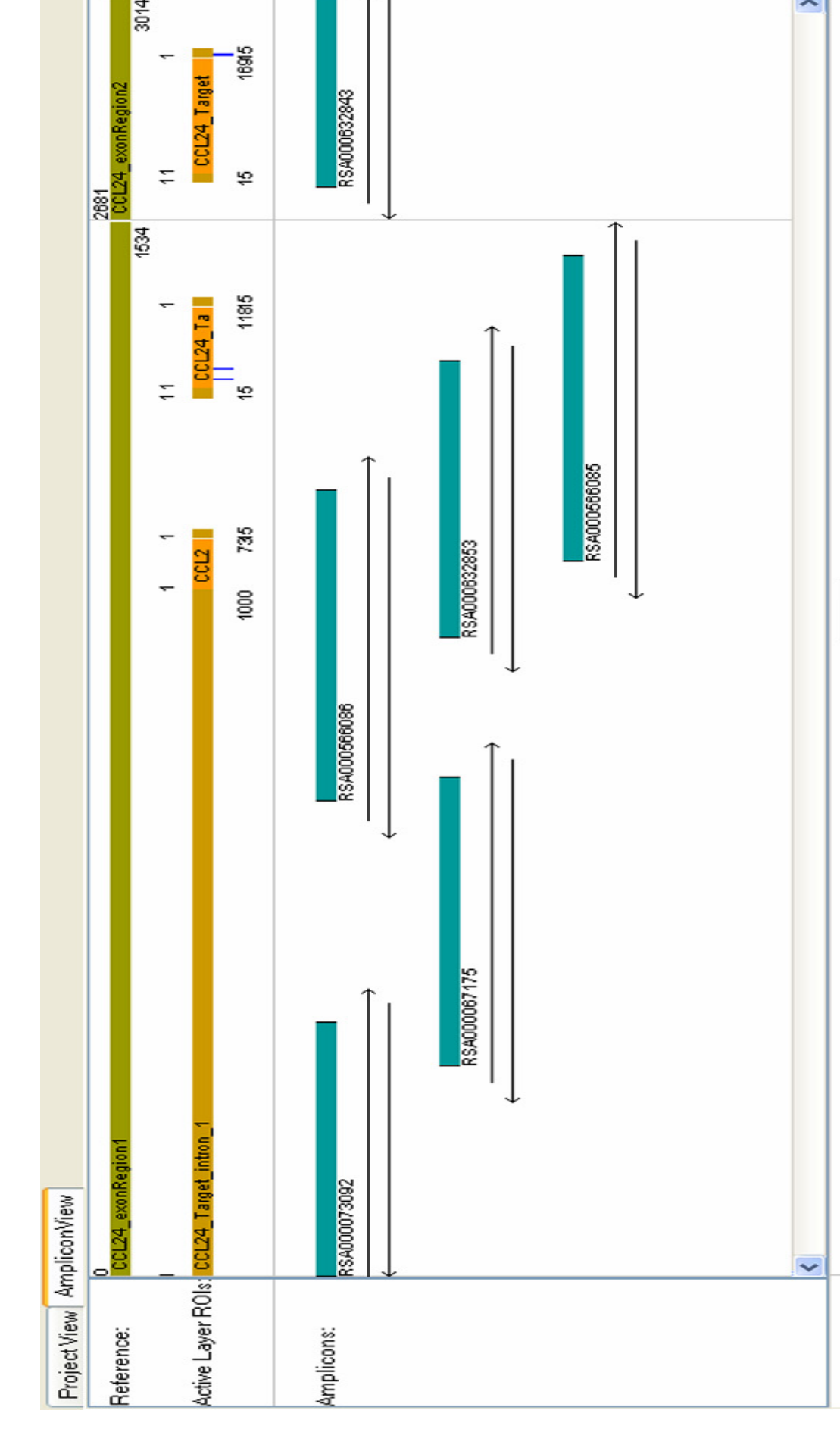


Figure 2. The Amplicon View is a new feature implemented in SeqScape® Software v2.5 in order to easily review the VariantSeq systems amplicons (RSA) coverage. Six amplicons are represented above spanning the target regions for the CCL24 gene. A teal bar is shown for each of the amplicons indicating complete coverage in the forward and reverse orientations. For this experiment, three unknown variants (blue bars) were detected in target exons and introns.

The process is completely automated providing fast and accurate data analysis. After the primers have generated sequences, the sequence files and project template (part of the resequencing primer product) are analyzed by SeqScape software (Figure 2).

Applied Biosystems provides project templates for each Resequencing Set (RSS). Each project template provides a reference sequence (Reference Data Group) to which all the specimens in a project are compared, settings that are used to analyze the data (Analysis Defaults) and settings that are used to display the data (Display Settings). By directly importing the RSS project template into the SeqScape Manager within the software, a project can be created. After the project is analyzed, the quality of the results can be reviewed and the variants can be examined (Figure 3).

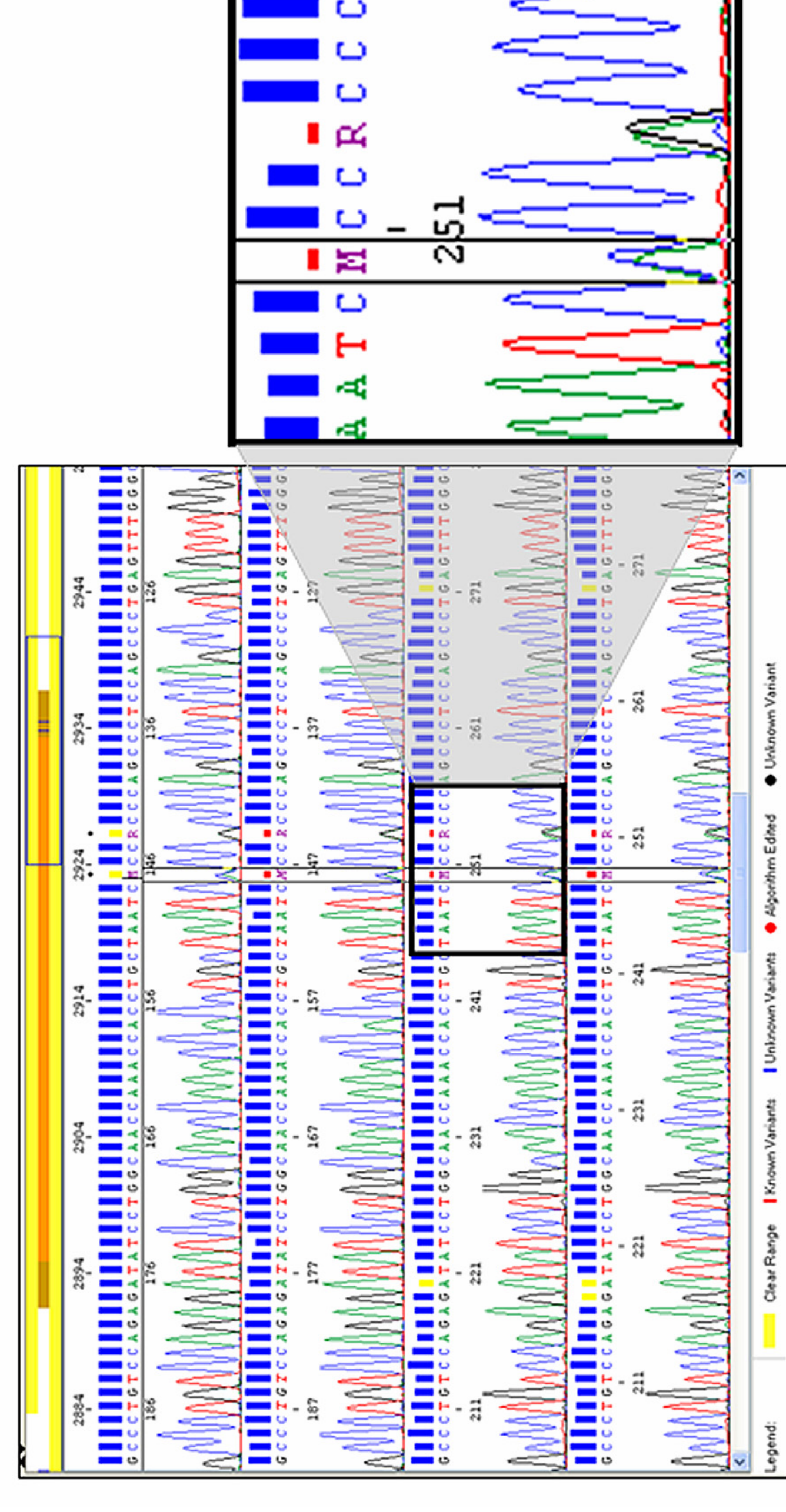


Figure 3. Heterozygote detection using SeqScape® Software v2.5. As noted in Figure 2, three variants were detected for this experiment using the CCL24 gene. Two of the variants are illustrated in the electropherogram above with black dots. The sequencing reactions used in this experiment were performed using Applied Biosystems BigDye® Terminator v3.1 Cycle Sequencing Kit following the VariantSeq Resequencing System protocol and subsequently purified using Centr-Sep™ spin columns. The 3130xl UltraSeq36_POP7 run module was used and the samples were analyzed using KB™ Basecaller v1.2 in SeqScape Software v2.5.

CONCLUSION

To address the challenges inherent in the use of DNA sequencing for mutation detection, Applied Biosystems developed an integrated system that comprises the new 3130 series Genetic Analyzers, the VariantSeq Resequencing System, and SeqScape software. Together, this system, along with the BigDye® Terminator chemistry kits, provides a complete, cost-effective solution for mutation detection.

© 2005 Applied Biosystems. All Rights Reserved. For Research Use Only. Not for use in diagnostic procedures.

VariantSeq Resequencing System is currently not optimized in the 3130 instrument series. We intend to begin optimization work and to support such application in the near future.

The Applied Biosystems 3130xl Genetic 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.

SeqScape® Software v2.5 - Notice to Purchaser: License Disclaimer. Purchase of this software product alone does not imply any license under any process, instrument or other apparatus, system, composition, reagent or kit rights under patent claims owned or otherwise controlled by Applied Biosystems, either expressly or by estoppel.

VariantSeq Resequencing System: The PCR process is covered by patents owned by Roche Molecular Systems, Inc. and F. Hoffmann-La Roche, Ltd. Applied Biosystems, BigDye, and SeqScape are registered trademarks and AB (Design), Applera, KB, iScience (design), POP-7, and VariantSeq are trademarks of Applied Biosystems or its subsidiaries in the US and/or certain other countries.

AmpliTaq Gold is a registered trademark of Roche Molecular Systems. All other trademarks are the sole property of their respective owners.