Development of a Resequencing Workflow for Variant Analysis in the *MLH1* and *MSH2* Genes

Applied Biosystems® 3130*xl* and 3500xL Series Genetic Analyzers

THE FUNCTIONAL UNIT OF ONCOLOGY AND MOLECULAR GENETICS, LILLE CHRU, FRANCE

Founded in 1993 within the Department of Biochemistry and Molecular Biology, The Functional Unit of Oncology and Molecular Genetics, Centre Hospitalier Regional Universitaire De Lille (CHRU) in Lille, France, was initially involved in the diagnosis of multiple endocrine neoplasia and similar syndromes. Clinical diagnosis protocols of colorectal cancer susceptibilities, including Lynch Syndrome, were established by the laboratory in 1997. The activities of the unit have extended to the genomic analysis of solid tumors and rare diseases of metabolism and human development. The laboratory of Dr. Marie-Pierre Buisine uses automated capillary electrophoresis (CE) to perform genetic and genomic analysis of samples, with an emphasis on *MLH1* and *MSH2* mutations and their role in colorectal cancer susceptibility.

Introduction

The *MLH1* and *MSH2* genes belong to the mismatch repair gene family and play an essential role in DNA repair. Variants in *MLH1* and *MSH2* are highly heterogeneous and contribute to genome instability, and for this reason are associated with increased risks of cancer [1,2].

Sanger sequencing by capillary electrophoresis is considered the industry standard method for sequence analysis. Dr. Marie-Pierre Buisine and colleagues have developed a resequencing workflow for the analysis of *MLH1* and *MSH2* mutations using the Applied Biosystems® 3130*xl* Genetic Analyzer. A concordance sequencing analysis study was also performed at Applied Biosystems using the Applied Biosystems® 3500xL Genetic Analyzer.

MLH1 and *MSH2* Mutations in Colorectal Cancer

Lynch syndrome or hereditary nonpolyposis colorectal cancer (HNPCC) is an inherited syndrome that predisposes individuals to colon cancer and other cancers. The overwhelming majority of HNPCC cases are attributed to mutations in *MLH1* and *MSH2* genes [2,3].

MLH1 and *MSH2* mismatch repair genes are large (19 exons, 2,271 bp of coding sequence within a 57,360 bp region for *MLH1* and 16 exons, 2,805 bp of coding sequence within a 80,099 bp region for *MSH2*), and more than 200 unique variants have been characterized in each gene. The mutations associated with HNPCC are not confined to any particular gene region and comprise either nucleotide substitutions (missense, nonsense, or splicing errors) or insertions/deletions, as well as large rearrangements (deletions/duplications). When the MLH1 and MSH2 proteins are absent or ineffective, the number of mistakes that are left unrepaired during cell division increases substantially, contributing to carcinogenesis.

The size of the *MLH1* and *MSH2* genes and the sheer number of sequence variations requiring confirmation across the large genetic region makes analysis by Sanger sequencing an excellent choice for these investigations. More information on the benefits of capillary electrophoresis for mutation analysis can be found in the highlight box to the right.

Automated Capillary Electrophoresis for DNA Sequencing

DNA sequencing via automated capillary electrophoresis is a highly referenced and robust technique, which is widely acknowledged as the industry standard method for sequence variation detection. In the last few years, improvements in the sequencing workflow have reduced the time-to-results from days to just a few hours, and the reduced cost-per-sequence has helped establish CE sequencing as a technique that can deliver accurate analysis without the need for prescreening steps.

CE sequencing is an attractive option for mutation detection where the locations of mutations are unknown (unlike techniques such as SNP analysis that validate known or expected sites of variation in a known sequence). Also, CE instrumentation offers assay flexibility—the same instrument can be used for both sequencing and for microsatellite analysis and can be employed for assays involving anywhere from a few to several hundred samples in a single day.

This "Your Innovative Research" article describes work done by researchers using Applied Biosystems® products. Note that the protocols and data described here are those of the contributing laboratory and may include modifications to the recommended Applied Biosystems® protocol. Optimization may be required for best results in your laboratory setting. Please be advised that Applied Biosystems only supports the performance of AmpliTaq Gold® DNA Polymerase, GeneAmp® PCR System 9700, BigDye® Terminator v1.1 Kit, BigDye XTerminator® Purification Kit, and the 3130xl and 3500xL DNA Analyzers when Applied Biosystems® protocols, reagent formulations, and kit storage recommendations are followed.

Materials and Methods

This *MLH1* and *MSH2* mutation sequencing protocol was developed as part of a collaborative research effort between Applied Biosystems and the laboratory of Marie-Pierre Buisine at the Functional Unit of Oncology and Molecular Genetics of Lille CHRU, France. Dr. Buisine's team from the Molecular Oncology Laboratory collected blood samples from 15 individuals identified as possibly having Lynch syndrome and analyzed them in an anonymous and blinded study.

Amplification Reactions and Conditions

Genomic DNA was isolated from 350 µL of blood from each of the 15 samples using the EZ1® DNA Blood Kit (QIAGEN) following the manufacturer's recommendations. Amplification primers were designed to produce one PCR amplicon for each exon (a total of 35 PCR primer pairs were designed). Amplification products were designed to cover the coding sequences and flanking intronic regions (see Table 1 for *MLH1* and *MSH2* primer sequences). Amplifications were performed using 50 ng of the extracted DNA in a 25 µL reaction with AmpliTaq Gold® DNA Polymerase on an Applied Biosystems® GeneAmp® PCR System 9700.

The *MLH1* and *MSH2* sequence structure did not permit a single optimal thermal profile for all amplification primers (see Tables 2 and 3 for the amplification and thermal cycling protocols). For both the *MLH1* and *MSH2* genes, two different PCR conditions and annealing conditions were used (Tables 2 and 3). In complex DNA regions (containing stretches of polyA) for both *MLH1* and *MSH2* genes, sequence-specific primers were used in place of the standard amplification primers (see Table 4 for *MLH1* and *MSH2* gene-specific primer sequences).

Purification of PCR Products and Cycle Sequencing Conditions

Twenty-five microliters of PCR product was purified with exonuclease (USB) and SAP (Promega), following standard protocols. Cycle sequencing was performed using BigDye® Terminator Cycle Sequencing Kit v1.1 with modifications to the manufacturer's recommended protocol (see Table 5).

Sequencing reactions were purified using the Applied Biosystems® BigDye XTerminator® Purification Kit according to the manufacturer's protocol.

Capillary Electrophoresis and Analysis

Purified sequencing reactions were analyzed independently at Lille CHRU on the Applied Biosystems® 3130*xl* Genetic Analyzer using the BigDye XTerminator® Ultraseq 36_POP7 Module. A concordance analysis of the samples was performed at Applied Biosystems using the 3500xL Genetic Analyzer (see highlight box below) using the ShortReadSeq50_POP7xl Module. The sequence results for each sample were analyzed independently at the two laboratories using SeqScape® Software (Figure 1) to verify the sequencing results and to identify the putative mutations for each sample.

Results

Using the optimized *MLH1* and *MSH2* gene sequencing protocol presented here, Dr. Buisine's team observed the following mutations in the 15 human blood samples:

- Mutation in *MSH2*, exon 6: c.1022C>T, p.Leu341Pro produces an abnormal protein (Figure 2)
- Mutation in *MLH1*, exon 15: c.1731G>A produces a splicing alteration, resulting in a premature stop codon and production of truncated MLH1 protein: p.Ser556ArgfsX14

Moreover, the mutations observed by the collaborators were identical to those identified by Applied Biosystems scientists when the samples were analyzed on the 3500xL Genetic Analyzer.

Conclusion

The comparison of sequencing data between the 3130*xl* and 3500xL Series Genetic Analyzers resulted in 100% concordance of mutation detection using Sanger sequencing technology. Dr. Buisine's team reported the following advantages when using this method for *MLH1* and *MSH2* mutation studies:

- The optimized workflow demonstrated in this study is easy to implement in research laboratories, and enables reliable and reproducible results
- BigDye XTerminator® Purification Kit used for purification of sequencing reactions eases automation and streamlines the workflow
- The workflow was optimized to use one annealing temperature for the majority of the *MLH1* and *MSH2* amplicons and a second annealing temperature was optimized for amplicons in difficult-to-amplify regions; the entire sequencing effort used only two thermal profiles which significantly reduced sequencing time for both genes

3500 Series Genetic Analyzers: Built on a Legacy of Innovation

The 3500 Series is setting a new standard in capillary electrophoresis—integrating a number of platform improvements designed for the optical and thermal subsystems, and pioneering an innovative consumables system approach. Working together, these elements provide the highest level of performance from a genetic analyzer to date.

The 3500 platform can run a wide variety of applications—including de novo sequencing and resequencing (mutational profiling)—as well as microsatellite analysis, MLPA™, AFLP®, LOH, MLST, and SNP validation or screening. The majority of applications can be run on a single polymer and capillary array, and the 3500 Series Data Collection Software integrates seamlessly with several downstream Applied Biosystems® software packages to provide comprehensive analysis of genetic data:

- Variant Reporter® Software—designed for mutation detection and analysis, SNP discovery and validation, and sequence confirmation
- Sequencing Analysis Software with KB™ Basecaller—designed to analyze, display, edit, save, and print sequencing data
- SeqScape® Software—used for mutation detection and library-based allele identification
- GeneMapper® Software—an ideal tool for genotyping, allele calling, fragment sizing, and SNP analysis

For more information on the 3500 Series Genetic Analyzers, visit us online at **www.appliedbiosystems.com/3500**.

The advanced Applied Biosystems® 3500 and 3500xL Series Genetic Analyzers offer the following benefits:

- Easy instrument setup
- Ready-to-use consumables—just load and run
- Simplified user interface of the 3500 Data Collection Software—easy display of consumable and array usage information, quick-start functionality, and system maintenance reminders
- Performance-check functionality, for chemistry checks and verification of instrument performance (as required by laboratory operating procedures)
- Security, audit trail, and electronic signature features to help process-controlled laboratories comply with 21CFR Part 11 requirements

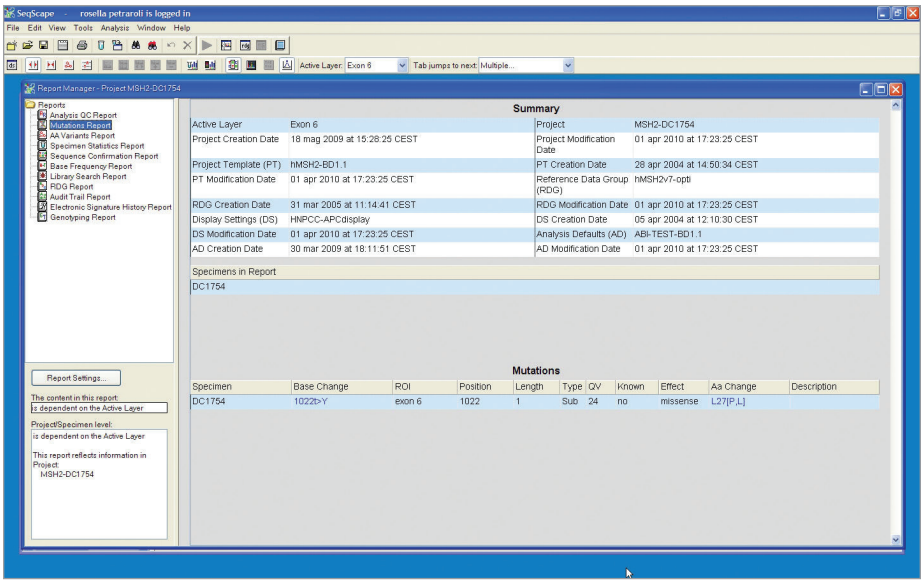

Figure 1. Data Summary View of SeqScape® Software. The software detects all variations between the reference sequence and the subject sample sequence, including deletions, insertions, mismatches, heterozygous bases, and heterozygous insertions/deletions. The variations between the consensus and the reference sequence are reported as mutations in the Mutations Report (pictured here).

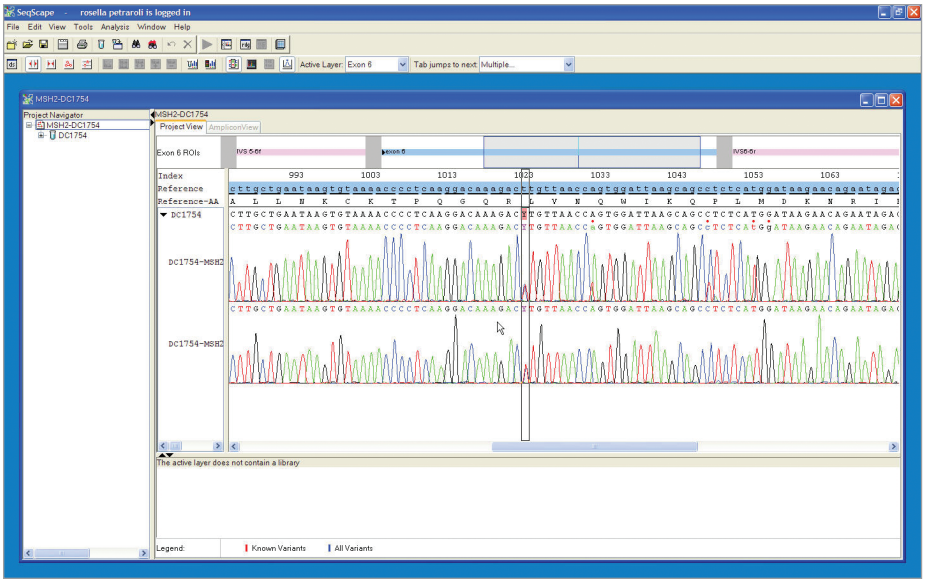

Figure 2. *MSH2* Electropherograms Showing Sequence Variation. Through automated sequence analysis, superior basecalling, and sophisticated data visualization tools, SeqScape® Software v2.7 facilitates definitive mutation analysis, as exemplified in this study by mutation of *MSH2*, exon 6: c.1022C>T, p.Leu341Pro (abnormal protein). The region showing the sequence variation is boxed in black.

Acknowledgement

We acknowledge Laure Bagait, Applied Biosystems Field Application Scientist, Molecular Biology, for her contribution to this collaboration. We also thank the members of Dr. Buisine's team for sharing their data: Marie-Pierre Buisine, PD, PhD, Mr. Tonio Lovecchio, Mme. Marie-Claire Dieu, Mr. Denis Boidin, and Mr. Michel Crépin.

References

1. Peltomäki P (2001) DNA mismatch repair and cancer. *Mutat Res* 488:77–85.
2. Mitchell RJ, Farrington SM, Dunlop MG et al. (2002) Mismatch repair genes hMLH1 and hMSH2 and colorectal cancer: a HuGE review. *Am J Epidemiol* 156:885–902.
3. Hitchins MP, Ward RL (2009) Constitutional (germline) *MLH1* epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. *J Med Genet* 46:793–802.

Table 1. *MLH1* and *MSH2* Primer Sequences.

Amplicon	Primer Sequence		Amplicon	Primer Sequence	
	Forward	Reverse		Forward	Reverse
MLH1-1	f 5'- AGGCACTGAGGTGATTGGC -3'	r 5'- TCGTAGCCCTTAAGTGAGC -3'	MSH2-1T	f 5'- TCGCGCATTTTCTTCAAC -3'	r 5'- ACTCTCTGAGGCGGGAAA -3'
MLH1-2	f 5'- AATATGTACATTAGAGTAGTTG -3'	r 5'- CAGAGAAAGGTCCTGACTC -3'	MSH2-2T	f 5'- CATGAAGTCCAGCTAATACAGTGC -3'	r 5'- AAATTCTTCACATTTTTATTTTTCTACTC -3'
MLH1-3T	f 5'- CTGGATTAAATCAAGAAAATGGG -3'	r 5'- TGACAGACAATGTCATCACAGG -3'	MSH2-3.1	f 5'- GCTTATAAAATTTTAAAGTATGTTC -3'	r 5'- CCTTTCCTAGGCCTGGAATC -3'
MLH1-4	f 5'- AACCTTTCCCTTTGGTGAGG -3'	r 5'- GATTACTCTGAGACCTAGGC -3'	MSH2-4T	f 5'- GCACATTGTATAAACATTTAATGTAGG -3'	r 5'- GAGATAAATATGACAGAAATATCCTTCTA -3'
MLH1-5	f 5'- GATTTTCTCTTTTCCCCTTGGG -3'	r 5'- CAAACAAAGCTTCAACAATTTAC -3'	MSH2-5	f 5'- AAGGAAAATGAGGGACTTCAG -3'	r 5'- CCAATCAACATTTTTAACCC -3'
MLH1-6	f 5'- GGGTTTTATTTTCAAGTACTTCTATG -3'	r 5'- GCTCAGCAACTGTTCAATGTATGAGC -3'	MSH2-6	f 5'- GTTTTCACTAATGAGCTTGCC -3'	r 5'- CAATAAGTGGTATAATCATGTGGG -3'
MLH1-7T	f 5'- CTAGTGTGTGTTTTTGGCAACTC -3'	r 5'- ACATCATAACCTTATCTCCACCAG -3'	MSH2-7	f 5'- TGAGACTTACGTGCTTAGTTG -3'	r 5'- GTATATATTGTATGAGTTGAAGG -3'
MLH1-8	f 5'- CTCAGCCATGAGACAATAAATCC -3'	r 5'- GGTTCCCAAATAATGTGATGG -3'	MSH2-8	f 5'- ATGATTTGTATTCTGTAAAATGAGATC -3'	r 5'- GGCCTTTGCTTTTTAAAAATAAC -3'
MLH1-9	f 5'- AGGACCTCAAATGGACCAAG -3'	r 5'- GTGAGTGGATTCCCCATGTG -3'	MSH2-9T	f 5'- CCCATTATTTATAGGATTTTGTCAC -3'	r 5'- AGAATTATTCCAACCTCCAATG -3'
MLH1-10	f 5'- CATGACTTTGTGTGAATGTACACC -3'	r 5'- GAGGAGAGCCTGATAGAACATCTG -3'	MSH2-10T	f 5'- AGTCAGAAACTAACATTCATAAGGG -3'	r 5'- CATCATGTTAGAGCATTTAGGGA -3'
MLH1-11	f 5'- CTCCCCCTCCCACTATCTAAGG -3'	r 5'- AAAATCTGGGCTCTCACGTCTG -3'	MSH2-11	f 5'- CACATTGCTTCTAGTACAC -3'	r 5'- CCAGGTGACATTCAGAAC -3'
MLH1-12.1	f 5'- GTACTGCTCCATTTGGGGAC -3'	r 5'- ACAGGACCATTCCAGCACCA-3'	MSH2-12	f 5'- ATTCAGTATTCCTGTGTAC -3'	r 5'- CGTTACCCCCACAAAGC -3'
MLH1-13	f 5'- TGCAACCCACAAAATTTGGC -3'	r 5'- CTTTCTCCATTTCCAAAACC -3'	MSH2-13	f 5'- TGTAAACCTACGCGATTAAT -3'	r 5'- GGACAGAGACATACATTTC -3'
MLH1-14	f 5'- TGGTGTCTCTAGTTCTGG -3'	r 5'- CATTGTTGTAGTAGCTCTGC -3'	MSH2-14T	f 5'- ATGTTACCACATTTTATGTGATGG -3'	r 5'- AAGGGTAGTAAGTTTCCCATTACC -3'
MLH1-15T	f 5'- GTATCTCAAGCATGAATTCAGCT -3'	r 5'- CTACTATTTTCAGAAACGATCAGTTG -3'	MSH2-15	f 5'- CTCTTCTCATGCTGTCCC -3'	r 5'- ATAGAGAAGCTAAGTTAAAC -3'
MLH1-16	f 5'- CATTTGGATGCTCCGTTAAAGC -3'	r 5'- ACCCGGGTGGAAATTTTATTTG -3'	MSH2-16	f 5'- ATTACTCATGGGACATTCAC -3'	r 5'- TACCTTCATTCCATTACTGG -3'
MLH1-17	f 5'- GGAAAGGCACTGGAGAAATGGG -3'	r 5'- CCCTCCAGCACACATGCATGTACCG -3'			
MLH1-18	f 5'- TAAGTAGTCTGTGATCTCCG -3'	r 5'- AGCTGACTGTTGTGTTTGAAC -3'			
MLH1-19	f 5'- GACACCAGTGTATGTTGG -3'	r 5'- GAGAAAGAAGAACACATCCC -3'			

Table 2. *MLH1* PCR Conditions.

PCR Reactions	
H_2O	13.35 µL
Buffer, 10X	2.5 µL
$MgCl_2$, 25 mM	2.5 µL
dNTPs, 2.5 mM each	2.5 µL
Primers (forward and reverse), 2.5 µM	3 µL
AmpliTaq Gold® DNA Polymerase	0.15 µL
DNA, 50 ng/µL	1 µL

Thermal Cycling Protocol	
	95°C, 7 min
35 cycles	94°C, 30 sec
	60°C or 55°C, 30 sec
	72°C, 1 min
	72°C, 7 min
	4°C, hold

Annealing Temperatures
60°C for amplicons 1, 5, 6, 8, 9, 10, 16, 17, 12
55°C for amplicons 2, 3, 4, 7, 11, 13, 14, 15, 18, 19

Table 5. Cycle Sequencing Conditions.

Cycle Sequencing Reactions	
H_2O	6 µL
PCR product (diluted 1:5)	1 µL
Primer (5 µM)	0.5 µL
BigDye® v1.1 Terminator Mix	1 µL
BigDye® Terminator Sequencing Buffer (5X)	1.5 µL

Thermal Cycling Protocol	
	96°C, 1 min
	96°C, 10 sec
25 cycles	50°C, 5 sec
	60°C, 2 min
	15°C, hold

Note: Researchers used the manufacturer's recommended protocol for thermal cycling. The above reaction mix is modified compared to manufacturer's recommendations (see footnote on page 1).

Table 3. *MSH2* PCR Conditions.

Exons 1, 2, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16		Exons 3, 7, 11	
H_2O	13.35 µL	H_2O	11.85 µL
Buffer, 10X	2.5 µL	Buffer, 10X	2.5 µL
$MgCl_2$, 25 mM	2.5 µL	$MgCl_2$, 25 mM	4.0 µL
dNTPs, 2.5 mM each	2.5 µL	dNTPs, 2.5 mM each	2.5 µL
Primers (forward and reverse), 2.5 µM	3.0 µL	Primers (forward and reverse), 2.5µM	3.0 µL
AmpliTaq Gold® DNA Polymerase	0.15 µL	AmpliTaq Gold® DNA Polymerase	0.15 µL
DNA, 50 ng/µL	1.0 µL	DNA, 50 ng/µL	1.0 µL

Thermal Cycling Protocol	
	95°C, 7 min
	94°C, 30 sec
35 cycles	60°C or 55°C, 30 sec
	72°C, 1 min
	72°C, 7 min
	4°C, hold

Annealing Temperatures
60°C for amplicons 3, 4, 7
55°C for amplicons 1, 2, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16

Table 4. Sequence-Specific Primers.

	Forward	Reverse
MLH1-12Af	f 5'- CCAGACTTTGCTACCAGGAC -3' Placed after T stretch	
MLH1-12Br		r 5'- CTTATCCTCTGTGACAATGGC -3' Internal primer to have two short sequences (as opposed to only one longer sequence)
MLH1-3	f 5'- AGAGATTGGGAAAATGAGTAAC -3' Shift primer to avoid noise in sequence	r 5'- ACAATGTCATCACAGGAGG -3'
MSH2-2eAF	f 5'- GAGCAAAGAATCTGCAGAGTG -3' T stretch	
MSH2-5Ar		r 5'- ACCTGAAAAAGGTTAAGGGC -3' A stretch
MSH2-4	f 5'- TTCTTATTCCTTTTCTCATAG -3' To avoid noise	r 5'- ATTGTAATTCACATTTATAATCC -3' To avoid noise
MSH2-8f	f 5'- GATTTGTATTCTGTAAAATGAGATC -3' Light shift to recover better quality	
MSH2-9	f 5'- CATTATTTATAGGATTTTGTCAC -3' Light shift to recover better quality	r 5'- GAATTATTCCAACCTCCAATG -3' Light shift to recover better quality

For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use.

© 2010 Life Technologies Corporation. All rights reserved. The trademarks mentioned herein are the property of Life Technologies Corporation or their respective owners. AmpliTaq Gold is a registered trademark of Roche Molecular Systems, Inc. EZ1 is a registered trademark of QIAGEN.
Printed in the USA. 06/2010. CO11799

applied biosystems™ by life technologies™

Headquarters
850 Lincoln Centre Drive | Foster City, CA 94404 USA
Phone 650.638.5800 | Toll Free 800.327.3002
www.appliedbiosystems.com

International Sales
For our office locations please call the division headquarters or refer to our Web site at
www.appliedbiosystems.com/about/offices.cfm