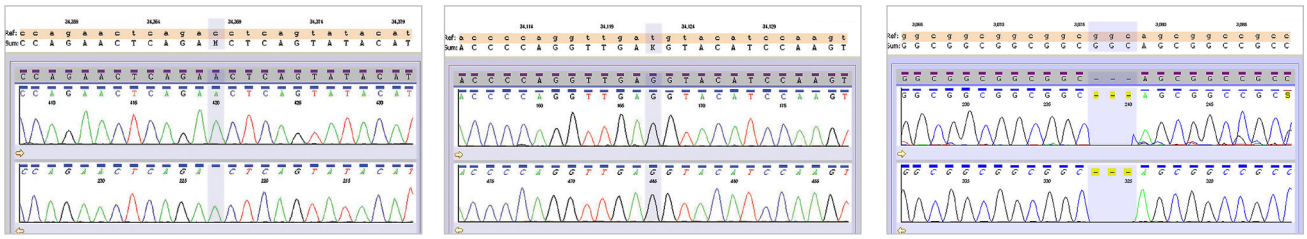
DNA Resequencing Analysis of X-linked Variants Using the Applied Biosystems 3500xL Genetic Analyzer



A. Non-recurrent **B.** Recurrent **C.** Indel

Figure 1. Non-synonymous SNPs and indels identified from 19 blinded probands suspected of MRX. **A.** A non-recurrent missense variant was identified in the gene *ZNF41* c.374T>G (Exon 4a amplicon) p.I125R. **B.** Recurrent missense variants were identified in the gene *ZNF41* c.945T>G (Exon 4b amplicon) p.D315E. **C.** One non-recurrent in-frame indel was identified in the gene *ARX* c.333-335delGGC (Exon 2a amplicon) p.A112del.

Introduction

Mental retardation affects a larger number of males in the population than females, and this has been attributed, in part, to the increased proportion of mutations in genes on the X-chromosome. The number of causative mutations leading to X-linked mental retardation (XLMR) is unknown, and will remain so, until all of the genes on the X chromosome have been fully investigated. While this research is ongoing, screening for XLMR remains a challenge. In this Application Note, analysis of a non-syndromic X-linked Mental Retardation (MRX) Resequencing Panel will be used to demonstrate the resequencing application on 3500 Series Genetic Analyzers and highlight several advantages to using this capillary electrophoresis platform.

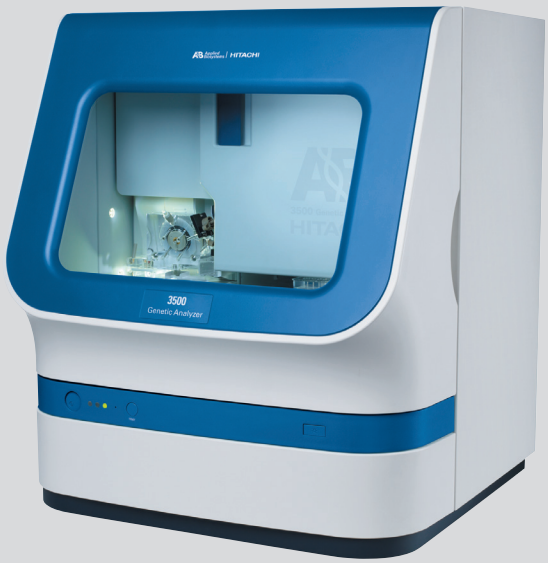
Nonsyndromic X-linked Mental Retardation Resequencing Panel

The prevalence of XLMR is estimated to affect approximately 1 in 1,000 males, with MRX thought to account for two thirds of these cases. XLMR can be subdivided into two groups: syndromic X-linked mental retardation (MRXS), characterized by external features and/or neuromuscular and metabolic disorders; and MRX, where intellectual disability and the X-linked mode of inheritance is the only consistent indication. To date, there have been approximately 90 X-linked genes implicated in XLMR, with the majority of these genes associated with MRXS. A subset of genes on the X chromosome have been associated with MRX where the only discernible feature is intellectual disability. Significant overlap between these two sets of genes indicates that syndromal and non-syndromal cases can both result from alterations in many of the XLMR genes. The identification of males with MRX is more straightforward when a positive family history exists. Unfortunately, many other males with unconfirmed MRX exist, and in the absence of an X-linked pedigree, identification of their underlying etiology is much more difficult. It is estimated that MRX accounts for roughly two thirds of the total number of XLMR cases, adding to the difficulty in definitively identifying affected individuals.

To screen for MRX, the Molecular Diagnostic Laboratory at the Greenwood Genetic Center (Greenwood, South Carolina) has developed a panel of sequencing primers for the interrogation of nine genes. These nine genes were most commonly associated with MRX (according to the literature) at the time the panel was designed. The panel provides sequence coverage for 95 amplicons encoding the exons and intron junctions for the following genes: acyl-CoA synthetase long-chain family member 4 (*ACSL4*); Aristaless Related Homeobox (*ARX*), FTSJ Homolog 1 (*FTSJ1*), GDP dissociation inhibitor 1 (*GDI1*), interleukin 1 receptor accessory

The 3500 Series Genetic Analyzers

The 3500 Series Genetic Analyzers are automated, fluorescence-based, capillary electrophoresis systems. Samples are sequenced in less than 40 minutes on the 24-capillary 3500xL system or on the 8-capillary 3500 system, using 3500 POP-7™ Polymer and the 50 cm capillary array. For convenience, the 3500 Series instruments also include a sequencing module that incorporates the BigDye® XTerminator™ Purification Kit workflow. After the sequence data are collected, the KB™ Basecaller embedded in the 3500 Data Collection software automatically processes the sequenced samples and provides read length (based on Q20) greater than 850 base pairs (bp), with average base Quality Values greater than 20 (QV20). Furthermore the 3500xL system, using the RapidSeq50_ POP7 run module, can efficiently sequence up to 840 samples in a 23-hour period, generating high-quality, high-resolution data with minimal hands-on time.



- Fully automated from the moment you place an 8-tube strip, 96-well, or 384-well sample plate on the instrument and start the run. The instrument provides continuous, unattended operation for every phase of the process, including polymer-loading, sample injection, separation, detection, and primary data analysis. The 3500 Series Systems feature simplified, easy-to-install consumables. The Anode and Cathode Buffer Containers are supplied as ready-to-use 1X Genetic Analysis Buffer formulations.
- RFID (radio frequency identification) tags on buffers, polymer, and arrays enables automated electronic tracking of lot number, usage, and expiration date information within the Data Collection Software.
- Easy-to-use wizards for instrument operation and maintenance ensure predictable, hassle-free performance. The majority of applications can be analyzed on a single configuration of POP-7™ polymer with a 50 cm capillary array.
- Integrated primary analysis and QC software, includes improved Data Collection software with an intuitive workflow from plate setup to primary analysis that performs base calls and applies quality control flags to alert the user to failed or low-quality samples.

protein-like 1 (*IL1RAPL1*), Jumonji/ARID domain-containing protein 1C (*JARID1C*), oligophrenin 1 (*OPHN1*), Polyglutamine Binding Protein 1 (*PQBP1*), and Zinc Finger Protein 41 (*ZNF41*).

For this study, extracted gDNA from 19 blinded probands suspected of having MRX (obtained from the Greenwood Genetic Center) was directly sequenced on the Applied Biosystems 3500xL Genetic Analyzer using the Greenwood MRX resequencing panel (the M13-tailed primers used in the study were synthesized based on sequence data provided by the Greenwood Genetic Center). Data for the 19 samples was collected using the 3500xL Genetic Analyzer in combination with the associated 3500 system consumables and data analysis software. When compared to data obtained at the Greenwood Genetic Center (using a 3730 Genetic Analyzer) with the same 19 samples, the results were found to be concordant. Variant detection in the nine genes across the 19 samples will be discussed below.

The 3500 and 3500xL Genetic Analyzers, when combined with Applied Biosystems AmpliTaq Gold® PCR Master Mixes, BigDye® Terminator Cycle Sequencing Kits, BigDye® XTerminator™ Purification Kit, and Variant Reporter™ Software v1.1, provide a highly robust and efficient resequencing system for scientists with throughput requirements of up to 840 samples per day. Together, the components of this integrated system provide a complete workflow that simplifies the steps of resequencing protocols, and offer a complete, efficient solution for laboratories, while delivering highly accurate and reproducible results. Furthermore, by using Applied Biosystems Fast Resequencing Workflow, scientists can identify human gene mutations in a fraction of the time than was previously required with conventional cycle sequencing workflow analysis. The MRX Resequencing Panel coupled with the Fast Resequencing Workflow (see Figure 3) highlights the advantages of using capillary

TABLE 1. SUMMARY OF CODING VARIANT RESULTS OBSERVED FOR 19 SPECIMENS (LABELED A–S) USING THE GREENWOOD MRX RESEQUENCING PANEL AND ANALYZED ON THE 3500XL GENETIC ANALYZER.

Non-recurrent				Recurrent			
Variation Type	**Specimen ID**		**Variation Type Total**	**Specimen ID**		**Variation Type Total**	**Total Variants**
Nonsense	None	None Observed	0	None	None Observed	0	**0**
Missense	E	*ZNF41* (Exon 4a) c.374T>G p.I125R	3	D, N	*ZNF41* (Exon 4b) c.945T>G p.D315E	2	**5**
	C	*ZNF41* (Exon4b) c.1107C>A p.D397E		E, I, N	*OPHN1* (Exon2) c.115G>A p.V39I		
	P	*JARID1C* (Exon 26) c.4637G>A p.R1546Q					
Silent	P	*IL1RAPL1* (Exon 1) c.36C>T	4	A,B,E,G,J, L,Q	*GDI1* (Exon 2-3) c.219T>C	2	**6**
	I	*JARID1C* (Exon 13-14) c.1764G>A					
	Q	*JARID1C* (Exon 13-14) c.1794C>G		D,N	*ZNF41* (Exon 4a) c.786G>A		
	Q	*JARID1C* (Exon 13-14) c.1884G>A					
Inframe Indels	A	*ARX* (Exon 2a) c.333-335delGGC (A112del)	1	None	None Observed	0	**1**
Out-of-Frame Indels	None	None Observed	0	None	None Observed	0	**0**
Splice	None	None Observed	0	None	None Observed	0	**0**
Total Non-recurrent			**8**	**Total Recurrent**		**4**	**12**

electrophoresis for DNA resequencing and variant identification across a large number of samples and genes.

The Fast Resequencing Workflow

Resequencing (also known as comparative, direct, medical, or PCR sequencing) is commonly used in medical research for the identification of causal or associated disease variants in specific genes or multigene disorders, such as MRX. Resequencing using CE-based platforms is commonly employed in these cases because it allows direct detection of DNA variants, including single nucleotide polymorphisms (SNPs) as well as small insertions and deletions.

The resequencing protocol employed for the analysis of the 19 samples was the Applied Biosystems Fast Resequencing Workflow (described in the Application Note, *Advances in Fast PCR Contribute to a Fast Resequencing Workflow*, 104AP04-02). The protocol was followed as described in the application note, with the following exception: the BigDye® Terminator Cycle Sequencing Kit v3.1 was used instead of the BigDye® Terminator v1.1 Kit. All other steps in the workflow were adapted to the 3500 system.

Results and Discussion

Sequencing data collected on the Applied Biosystems 3500xL Genetic Analyzer were analyzed for missense variants (or non-synonymous SNPs) using Applied Biosystems Variant Reporter™ Software (see sidebar). Coding region variants (by gene) that were identified in the 19 samples using the Greenwood MRX resequencing panel are summarized in Table 1.

Of the 19 samples analyzed, a total of twelve coding variants were identified including five missense, six silent, and one in-frame indel. The missense variants were subsequently assessed using PolyPhen software (*Poly*morphism *Phen*otyping) to determine the effect the amino acid variant would putatively have on the phenotype of the protein. PolyPhen (http://genetics.bwh.harvard.edu/pph/) is a web-based tool that uses structural, phylogenetic, alignment, and annotation information to assign a score and estimate the putative impact of an amino acid substitution on the structure/function of a protein. The amino acid substitution's effect on protein function is given one of four predictions: benign, possibly damaging, probably damaging, or unknown.

Variant Reporter™ Software for Powerful Sequence Analysis

Variant Reporter™ software enables researchers to create sequencing projects for mutation detection and analysis, variant discovery and validation (including SNPs, heterozygous insertion/deletion mutations (HIMs), insertions, and deletions), and sequence confirmation. This analysis tool integrates improved algorithms for variant detection that are trained to discover accurate sequence variations and report review status for traceability. Variant Reporter™ software can be used to create easy-to-read Quality Control reports that contain trace files and data annotation. A guided workflow facilitates ease of use, and users can view, edit, print, and export sequence data. The software supports complex disease research strategies that interrogate multiple genes and can analyze projects containing in excess of 4000 trace files.

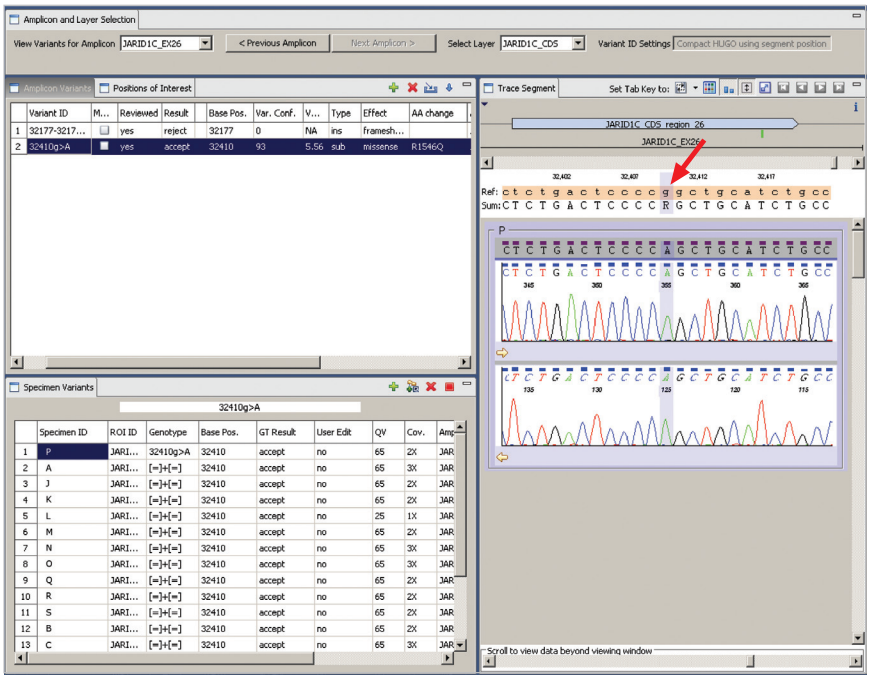


Figure 2. A missense variant in Exon 26 of the *JARID1C* gene (c.4637G>A) resulting in R1546Q substitution was identified in specimen P (indicated by the red arrow).

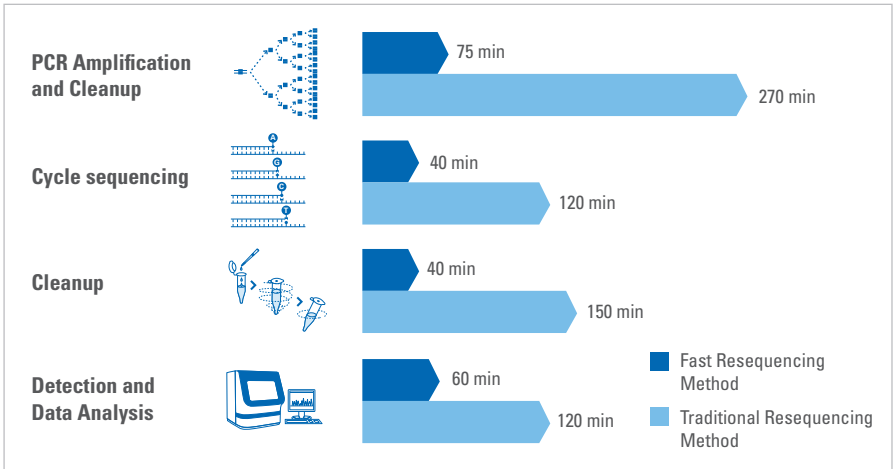


Figure 3. Protocol time from amplification to results: comparing fast and traditional resequencing methods.

One non-recurrent in-frame indel was identified in the *ARX* gene, which encodes a transcription factor required for normal brain development. In specimen A, the p.A112del variant (Exon 2a amplicon, c.333-335delGGC) was identified in the first poly-alanine tract of *ARX*. The *ARX* gene contains four GCG trinucleotide repeat regions encoding the amino acid alanine, expansion of the first poly-alanine tract from 10 residues to 17 has been found in association with infantile epileptic encephalopathy type 1 (EIEE1). In addition to causative mutations in EIEE1 and MRX, mutations in the *ARX* gene are responsible for: lissencephaly X-linked type 2 (LISX2), myoclonic epilepsy X-linked with intellectual disability and spasticity (XMEIDS), and Partington syndrome (PRTS). The p.A112del variant seems unlikely to impact the function of *ARX*, since expansion of the poly-alanine tract has been associated with altering protein function.

One non-recurrent missense variant was identified in the gene *JARID1C*. The *JARID1C* gene encodes a histone demethylase that specifically demethylates the lysine-4 residue of histone H3, and is involved in the transcriptional repression of neuronal genes. In specimen P, the p.R1546Q variant (Exon 26 amplicon, c.4637G>A) was identified (Figure 2). The amino acid change was given a PolyPhen score of 1.283; this variant is predicted to be a benign change with minimal effect on protein function.

Two non-recurrent missense variants and one recurrent missense variant were identified in the gene *ZNF41*, which encodes a Krueppel-type, zinc-finger domain containing protein with a putative role in transcriptional regulation. In specimen E, the p.I125R variant (Exon 4a amplicon, c.374T>G) was identified. This polymorphism is a previously identified polymorphism (dbSNP: rs17147624) in a variant sequence region (141–176) of the gene that is missing in isoform 7 and isoform 8 of *ZNF41*

The Laboratory of Dr. Michael Friez at the Greenwood Genetic Center

In 2008, Dr. Friez was named Director of the Diagnostic Laboratories at GGC. His primary duties as Director are to supervise the diverse genetic testing activities the laboratory has developed since its inception. Protocols for analyzing a variety of conditions related to mental retardation, muscular dystrophies, skeletal dysplasias, and other adult-onset genetic conditions are utilized on a daily basis. Recognizing the importance of basic research in his work, Dr. Friez is also involved in assisting several of the scientific research projects that are based in the nearby JC Self Research Institute.

(Shoichet et al.). The physicochemical property change from a medium-size, hydrophobic amino acid (isoleucine) to large, basic amino acid (arginine) was given a PolyPhen score of 1.043; this variant is predicted to have a benign effect on protein function.

In specimen C, the p.D397E variant (Exon 4b amplicon, c.1107C>A) was identified. This polymorphism is in a region (425–447) of the *ZNF41* gene that encodes a degenerate type 5 zinc-finger domain (C2H2). Both residues (aspartic acid and glutamic acid) are medium-size, acidic amino acids and the change was given a PolyPhen score of 0.488. This variant is predicted to be a benign change to protein function.

In specimens D and N, the p.D315E variant (Exon 4b amplicon, c.945T>G) was identified. This polymorphism is a previously identified polymorphism (dbSNP: rs2498170) in a region of the *ZNF41* gene that encodes a degenerate type 2 zinc-finger domain (C2H2) (Shoichet et al.). Both residues (aspartic acid and glutamic acid) are medium size, acidic amino acids, and the change was given a PolyPhen score of 0.191. This variant is predicted to be a benign change, with no effect on protein function.

A recurrent variant, p.V39I variant (Exon 2 amplicon, c.115G>A), was identified in specimens E, I, and N in *OPHN1*. The *OPHN1* gene is expressed in the brain and encodes a rhoGAP protein. The change was given a PolyPhen score of 0.025; this variant is predicted to be a benign change to protein function.

No truncating mutations were identified, and the gene with the most non-synonymous SNPs was *ZNF41*.

Conclusions

When used to analyze 19 blinded samples from the Greenwood Genetic Center, the 3500xL Genetic Analyzer generated resequencing data that corroborated the mutation results observed by Drs. Friez and Bartel (Greenwood Genetic Center, personal communication). These results show that the 3500xL Series Genetic Analyzers have great utility for complex resequencing strategies and should prove invaluable in future analyses of this type, including investigations involving more diverse sample types and different genetic disease conditions.

The standard-setting 3500 Series Genetic Analyzers are specifically designed to support the demanding performance needs of life science researchers in basic research laboratories as well as validated and regulated environments. The advanced capabilities of 3500 Series Genetic Analyzers such as new thermal control systems, enhanced optical detection, and new consumables designs provide an easy-to-use platform for the detection and primary analysis of resequencing samples. When coupled to the optimized Fast Resequencing Workflow and Variant Reporter™ v1.1 secondary analysis, the 3500 Series Genetic Analyzers provide a rapid, reliable, and easy-to-use platform for mutation detection across multiple genes and large numbers of samples.

Acknowledgements

Applied Biosystems would like to acknowledge the generous contribution of gDNA samples and primer sequences from Dr. Mike Friez and Dr. Frank Bartel, Molecular Diagnostic Laboratory, Greenwood Genetic Center, SC, which were used to generate the data shown in this Application Note.

References

Shoichet S.A., Hoffmann K., Menzel C., Trautmann U., Moser B., Hoeltzenbein M., Echenne B., Partington M., Van Bokhoven H., Moraine C., Fryns J.-P., Chelly J., Rott H.-D., Ropers H.-H., Kalscheuer V.M.

"Mutations in the ZNF41 gene are associated with cognitive deficits: identification of a new candidate for X-linked mental retardation."

Am. J. Hum. Genet. 73:1341–1354 (2003).

ORDERING INFORMATION

3500 Series System Packages		3500	3500xL
Package Name	**Description**	**P/N**	**P/N**
3500 Series Genetic Analyzer for Resequencing & Fragment Analysis	3500 Series System with Data Collection Software, Sequencing Analysis, Variant Reporter™, and GeneMapper® Software packages. System package also includes DNA Sequencing and Fragment analysis reagent kits for system qualification.	4440462	4440463
3500 Series Genetic Analyzer for Resequencing & Fragment Analysis With SAE	3500 Series System with Data Collection Software (includes additional functionality for Security, Audit Trail, and Electronic Signature (SAE) capabilities), Sequencing Analysis, Variant Reporter™, and GeneMapper® Software packages. System package also includes DNA Sequencing and Fragment analysis reagent kits for system qualification.	4440464	4440465
3500 Series Genetic Analyzer for Resequencing	3500 Series System with Data Collection Software, Sequencing Analysis, and Variant Reporter™ Software packages. System package also includes DNA Sequencing reagent kits for system qualification.	4440466	4440467
3500 Series Genetic Analyzer for Fragment Analysis	3500 Series System with Data Collection Software and GeneMapper® Software. System package also includes DNA Fragment Analysis reagent kits for system qualification.	4440468	4440469
3500 Series Genetic Analyzer for Sequence Typing & Fragment Analysis	3500 Series System with Data Collection Software, Sequencing Analysis, SeqScape®, and GeneMapper® Software packages. System package also includes DNA Sequencing and Fragment analysis reagent kits for system qualification.	4440470	4440471

System Consumables and Reagents

Description	P/N
3500xL Capillary Array (50 cm)	4404689
3500 Capillary Array (50 cm)	4404685
3500 POP-7™ Polymer (960 samples)	4393714
3500 POP-7™ Polymer (384 samples)	4393708
Anode Buffer Container (ABC) 3500 Series	4393927
Cathode Buffer Container (CBC) 3500 Series	4408256
Septa Cathode Buffer Container 3500 Series	4410715
Conditioning Reagent 3500 Series	4393718
Hi-Di™ Formamide (5 mL) 1 bottle	4401457
BigDye® Terminator v3.1 Cycle Sequencing Kit (1,000 rxns)	4337456
BigDye® XTerminator™ Purification Kit (1,000 rxns)	4376487

Visit www.AppliedBiosystems.com/3500Series or contact your local Sales Representative for more information.

FOR RESEARCH USE ONLY. NOT FOR USE IN DIAGNOSTIC PROCEDURES.

For those who require IVD-marked devices, the 3500 Dx and 3500xL Dx Genetic Analyzers and system accessories meet the requirements of the In Vitro Diagnostic Medical Devices Directive (98/79/EC). The 3500 Dx and 3500xL Dx systems are for distribution and use in specific European countries only. For more information about the 3500 Dx Series Systems, contact your Applied Biosystems representative. The purchase price of this Instrument includes a grant of a limited, non-transferable license under U.S. patents and method claims of its foreign counterparts, and under U.S. patents and element claims of its foreign counterparts, to use this particular instrument for electrophoresis methods employing fluorescence as a means of detection. No other licenses or rights are hereby conveyed either expressly, by implication, or estoppel including, but not limited to, any claims to a composition. This instrument incorporates technology subject to one or more patents licensed from Hitachi, Ltd. as well as patents and patented technology owned by or under control of Applied Biosystems. This instrument is Authorized for use in DNA sequencing and fragments analysis only. This Authorization is included in the purchase price of the instrument and corresponds to the up-front fee component of a license under process claims of U.S. patents and under all process claims for DNA sequence and fragment analysis of U.S. patents now or hereafter owned or licensable by Applied Biosystems for which an Authorization is required, and under corresponding process claims in foreign counterparts of the foregoing for which an Authorization is required. The running royalty component of licenses may be purchased from Applied Biosystems or obtained by using Authorized reagents purchased from Authorized suppliers in accordance with the label rights accompanying such reagents. Purchase of this instrument does not itself convey to the purchaser a complete license or right to perform the above processes. This instrument is also licensed under U.S. patent and apparatus and system claims in foreign counterparts thereof. No rights are granted expressly, by implication, or by estoppel under composition claims or under other process or system claims owned licensable by Applied Biosystems. For more information regarding licenses, please contact the Director of Outlicensing at Applied Biosystems, 850 Lincoln Centre Drive, Foster City, California 94404, USA.

Printed in the USA. 08/2010 Publication 106AP29-01 O-082428 0509

Applied Biosystems

Headquarters
850 Lincoln Centre Drive | Foster City, CA 94404 USA
Phone 650.638.5800 | Toll Free 800.345.5224
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