SOP³: a web-based tool for selection of oligonucleotide primers for single nucleotide polymorphism analysis by Pyrosequencing[®]

Angela M. Alexander¹, Christopher Pecoraro¹, Alexis Styche¹, William A. Rudert¹, Panayiotis V. Benos², Steven Ringquist¹, and Massimo Trucco¹

BioTechniques 38:87-94 (January 2005)

SOP³ is a web-based software tool for designing oligonucleotide primers for use in the analysis of single nucleotide polymorphisms (SNPs). Accessible via the Internet, the application is optimized for developing the PCR and sequencing primers that are necessary for Pyrosequencing[®]. The application accepts as input gene name, SNP reference sequence number, or chromosomal nucleotide location. Output can be parsed by gene name, SNP reference number, heterozygosity value, location, chromosome, or function. The location of an individual polymorphism, such as an intron, exon, or 5' or 3' untranslated region is indicated, as are whether nucleotide changes in an exon are associated with a change in an amino acid sequence. SOP³ presents for each entry a set of forward and biotinylated reverse PCR primers as well as a sequencing primer for use during the analysis of SNPs by Pyrosequencing. Theoretical pyrograms for each allele are calculated and presented graphically. The method has been tested in the development of Pyrosequencing assays for determining SNPs and for deletion/insertion polymorphisms in the human genome. Of the SOP³-designed primer sets that were tested, a large majority of the primer sets have successfully produced PCR products and Pyrosequencing data.

INTRODUCTION

Pyrosequencing[®] has been developed to allow for accurate sequencing of short stretches of DNA, initially for the analysis of expressed sequence tags (ESTs; References 1–3). Sequencing of lengths between 50 and 150 nucleotides has been reported during HLA genotyping as well as during the analysis of PCR-amplified cloned DNA (4–7). The method is suited for use during the analysis of single nucleotide polymorphisms (SNPs) and the detection of genomic insertion and deletion polymorphisms because of its ability to provide high-quality sequencing for short stretches of DNA and to accurately resolve heterozygous nucleotides by enabling out-of-phase sequencing (8). Pyrosequencing is performed by the addition of dNTPs individually, in a predefined dispensation order, so that the nascent nucleotide chain is extended one nucleotide residue per dispensation event.

The detection of nucleotide sequence is performed by way of a chain of enzymatic reactions involving the activities of DNA polymerase, apyrase, ATP sulfurylase, as well as luciferase (1,5). The incorporation of a particular nucleotide is displayed graphically in the form of a pyrogram of nucleotide dispensation event versus the intensity of emitted light. The Pyrosequencing reaction is quantitative in that increased light intensity is produced upon incorporation of multiple nucleotides.

The design of oligonucleotide primer sets for use during PCR and Pyrosequencing, a necessary step in the process of Pyrosequence-based mapping of genetic disease loci, while not individually cumbersome, is time-consuming to develop and test when large numbers of SNP-containing sites are to be examined. Publicly and commercially available software applications have been developed for PCR primer design, but these have not taken into account the added constraints that

are essential when optimizing Pyrosequencing reagents (9,10). Moreover, applications currently available for developing Pyrosequencing primers do not also allow for the design of primers for PCR. Thus, none of the existing software applications available via the Internet are capable of designing the complete set of primers required for Pyrosequence-based typing (PSBT) during SNP-scanning projects while also allowing for the simultaneous development of assays for multiple SNPs. The development of Pyrosequencing assays can be accomplished by exploitation of the sequence databases derived from the human and model organism genome projects and can be combined with primer design algorithms when large batches of SNPs are to be analyzed. Assay development relies on the availability of genome databases containing the identity and location of relevant SNPs for input into primer design software applications. Currently, this can be achieved only by

¹University of Pittsburgh School of Medicine and ²University of Pittsburgh, Pittsburgh, PA, USA

transporting the data between several software applications, each of which is specialized for use at a different stage of the process. Combining these steps into a single software application can improve the efficiency of the experimental design, increasing the level of data integrity by avoiding the use of different software applications to accomplish each of these goals independently (11).

This paper presents software that can efficiently design PCR and Pyrosequencing primers for large numbers of SNPs. The method automates existing processes and takes into consideration restrictions that should be applied during Pyrosequencing (1,2,5–7,12). Moreover, the application aids in the sorting of SNPs for biologically interesting properties (e.g., the association of an SNP with a particular genetic structural element, whether it codes for an amino acid change in the resulting protein, and its heterozygosity within a population), which, in turn, can add to the usefulness of particular primer sets during physical mapping studies.

MATERIALS AND METHODS

Selection of Oligonucleotide Primers for PCR and Pyrosequencing (SOP³)

The SOP³ application consists of warehoused genomic sequence data from the human genome project downloaded from the University of California Santa Cruz (UCSC; Santa Cruz, CA, USA) Genome Browser and the National Center for Biotechnology Information (NCBI) human genome release, Build 34, of the finished human genome assembly (<http://hgdownload.cse.ucsc.edu/glodenPath/hg16/bigZips>). Written in preprocessor hypertext protocol (PHP; Zend Technologies Ltd., Ramat Gan, Israel), the SOP³ application and associated MySQL[®] database were developed on a Linux[®] SUSE Enterprise Server 8 for the AMD64 operating system with Apache[™] version 2.0.48 (Apache Software Foundation, Forest Hill, MD, USA) on a customized computational computer server (@Xi Computer, San Clemente, CA, USA) consisting of 2× AMD Opteron[™] 246 64-bit processors with 1024 kb Cache,

8192 MB random access memory (RAM), and equipped with 3× 250 GB drives. The application and associated warehoused databases were designed to run as an Internet-available web site.

Preparation of Genomic DNA

Samples of genomic DNA were obtained either from purchased human genomic DNA (Human Biological Data Interchange, Philadelphia, PA, USA) or whole blood extractions from healthy donor volunteers. Purification of genomic DNA was performed using a QIAamp[®] DNA Mini Kit (Qiagen, Valencia, CA, USA) as directed by the manufacturer. DNA yields were typically greater than 5 µg of genomic DNA per sample. The samples were stored frozen at -20°C and thawed immediately prior to use.

Polymerase Chain Reaction

Oligonucleotides and 5'-end biotinylated oligonucleotides, examples of which are listed in Supplementary Table S1 available at the *BioTechniques*⁷ web site at <http://www.BioTechniques.com/January2005/AlexanderSupplementary.html>, were purchased from Integrated DNA Technologies (Coralville, IA, USA). PCR amplification (13) was performed in 50 µL volumes containing *Taq* buffer, 2 mM MgCl₂, 0.2 µM each dNTP (buffers and nucleotides were purchased from Applied Biosystems, Foster City, CA, USA), 0.2 µM forward and biotinylated reverse primers, 1 U *Taq* polymerase, and 5 µL purified genomic DNA (roughly 10 ng DNA). Amplification included 96° incubation for 3 min, followed by 50 cycles at 96°, 55°, and 72°C, incubated for 30 s at each step. PCR cycling was followed by a final 5-min incubation at 72°C. The samples were then stored at -20° or 4°C prior to Pyrosequencing. Amplification was performed using the primer sets designed by the customized software application SOP³, a subset of which are listed in supplementary Table S1.

Pyrosequencing

The Pyrosequencing apparatus (Pyrosequencer PSQ[™] 96 MA) was

purchased from Pyrosequencing AB (Uppsala, Sweden). Pyrosequencing reactions were performed using reagents provided with the PSQ 96 Sample Preparation kit and the PSQ 96 SQA Reagent kit (Pyrosequencing AB). Briefly, samples of 20–40 µL amplified DNA from the PCR mixture were mixed with 4 µL streptavidin-coated beads (Amersham Biosciences, Piscataway, NJ, USA) and prepared for Pyrosequencing as recommended by the manufacturer. The appropriate Pyrosequencing primer was added to each well in a volume of 5 µL using a 3-µM stock solution. The samples were heated to 80°C for 2 min, and then allowed to cool for 5 min at room temperature and sequenced by Pyrosequencing. A detailed description of the Pyrosequencing reaction conditions has been reported by Gharizadeh et al. (5). Pyrosequencing data were quantified and background corrected using PSQ 96 MA version 2.0.2 software (Pyrosequencing AB).

RESULTS

Algorithm for the Selection of PCR Primers

Designed to accept input via a web-based user interface, the software application SOP³ can be accessed over the Internet at URL <http://biodev.hgen.pitt.edu/sop3/>. The application consists of a database of the human genome sequence downloaded from the UCSC Genome Browser, Build 34, integrated with a human SNP database obtained from dbSNP, Build 118, at NCBI. The software application is written in PHP and enables user-defined queries to be examined against the SNP and sequence databases. As illustrated in the flow-chart (Figure 1), the application accepts input as a gene name, SNP reference sequence number, or chromosomal nucleotide location. Multiple gene names or reference sequences can be queried simultaneously with the upper limit currently set at 1000 SNPs. When a query is presented, the application extracts that information and compares it to the warehoused database. FASTA-formatted sequences of user-selectable length flanking the SNP (maximum of 2000

Figure 1. Representation of the organization of the SOP³ application web site. Data were combined from the University of California Santa Cruz (UCSC) Genome Browser (Build 34) and dbSNP (Build 118) from the National Center for Biotechnology Information (NCBI) into the SOP³ database. The SOP³ application enables the analysis of SNP attributes, including flanking sequence and chromosomal location, during the design of primer sets for PCR and Pyrosequencing. SNP, single nucleotide polymorphism.

Figure 2. Web interface of the output page of the SOP³ primer design application. The primer set was designed for SNP rs3842748 from the human insulin gene. The gene name, SNP reference sequence number, allele description, average heterozygosity value, and location of the SNP are indicated in the upper part of the figure. Primers for Pyrosequencing and PCR, along with the calculated length of the PCR amplification product, are given in the central part of the figure. A nucleotide motif at the 5' end of the forward PCR primer, separated by a space from the rest of the primer sequence, signifies the modification to the primer sequence to avoid self-priming during Pyrosequencing. In the lower area of the figure, the chromosomal DNA sequence surrounding the SNP is illustrated. The polymorphic residues are shaded blue while primers for PCR and Pyrosequencing are shaded yellow and red, respectively. SNP, single nucleotide polymorphism.

Table 1. Inputs, Default Values, and Output for SOP³

Input	Default Value
Search Settings	
Entry box for gene name (locus) or RefSeq	gene name (locus)
Number or chromosomal location	
PCR Primer Design	
T _m ^a	60°C
Flank length (maximum 1000 each side)	1000 nucleotides
Use repeat masker (yes or no)	no
Use extra 5' nucleotide	yes
Maximum PCR size	500 nucleotides
Minimum foldover	5 nucleotides
Unique n-mer length	6 nucleotides
AT test (yes or no)	no
Residue occurrence thresholds (minimum and maximum)	minimum 14, maximum 40
Pyrosequencing Primer Design	
T _m ^a	40°C
Primer minimum distance from SNP	1 nucleotide
Maximum distance from SNP	30 nucleotides
Function Type	
Can limit search by specific SNP attributes such as: UTR, splice site, synonymous or nonsynonymous change, coding region, or intron.	all selected
Order Results and Save	
Results can be saved to file (yes or no)	no
Results can be parsed by 1 or 2 limits: (heterozygosity, gene name, rs#, location, chromosome, or function).	heterozygosity, ascending
Output Display	
Gene name, rs#, and residue change	
Heterozygosity value, primer sequences	
Nucleotide sequence flanking SNP	
Sites of PCR primer annealing	Highlighted in yellow
Sites of Pyrosequencing primer annealing	Highlighted in red
Location of nearby SNPs	Highlighted in blue
Repeat masked regions	Capitalized, highlighted in gray
Predicted program (yes or no)	No
T _m , melting temperature; SNP, single nucleotide polymorphism; UTR, untranslated region; rs#, reference sequence number.	
^a T _m was calculated using the equation $T_m = 16.6 \cdot \log[\text{cation concentration}] + 41 \cdot (\text{fraction of GC}) + 81.5$, where the concentration of cation was estimated to be 0.1 M (14).	

nucleotides), along with its associated attributes (e.g., SNP location within a genetic structural element and heterozygosity value), are then returned to the application to facilitate primer design for PCR and, if successful, Pyrosequencing. This flanking sequence also indicates the locations and polymorphisms of all additional SNPs (nearby SNPs) contained in the database. The identification of sites for PCR primer annealing are based on providing primers that result in a PCR product of at least 60 bp in length while avoiding repetitive sequence elements and the formation of primer dimer pairs. The search begins for unique “n-mers”

of user-specified length to make up the 3' end of the PCR primers, and the search begins 30 bases on either side of the SNP to leave room for placement of a sequencing primer and includes querying a sequence string incorporating the other possible nucleotides at all other SNP locations. Once a unique region is located, the potential primer location is extended along the reference sequence until the specified melting temperature is achieved (14). The primer is tested to determine if it contains a predefined representation of each nucleotide residue, which is accomplished through the user-defined residue occurrence threshold

setting. As an additional level of specificity, the user can select to incorporate the AT test during primer design. This provides an additional constraint that will reject a potential oligonucleotide if it does not contain at least one A or T residue within the previously defined unique 3'-end region. The primer is then tested for the presence of a palindrome that would result in secondary structure and thus inhibit PCR amplification. Moreover, the forward PCR primer can be designed with additional unique nucleotide sequence at its 5' end to avoid the formation of secondary structure in the biotinylated template strand, which, during Pyrosequencing, can lead to competing sequencing signal (6,12,15). Once a pair of candidate primers is chosen, it can be rejected for the potential formation of primer dimers by searching for reverse complementarity. If this occurs, the program continues to choose a new forward primer, and then tests for dimerization with the reverse primer. If the new forward primer fails the dimerization test, then another reverse primer is chosen. This continues in a stepwise manner, each time moving further away from the SNP location, until a suitable primer pair is found within the specified PCR product length.

Algorithm for the Selection of Pyrosequencing Primers

Pyrosequencing primer design is initiated once a suitable pair of PCR primers has been identified (Figure 1) and occurs under the same basic parameters. Pyrosequencing primers are required to be within a user-selectable number of residues of the site of the SNP (with the maximum being 30 nucleotides). This 30-base maximum helps to streamline high-throughput assays and reduce reagent use. The AT test and the test for residue occurrence thresholds are not performed on the Pyrosequencing primer, but primer dimers and hairpin formation tests are still applied so as to avoid the formation of secondary structure that may interfere with the annealing of Pyrosequencing primer to the biotinylated template strand. Candidate Pyrosequencing primers are chosen on both the 5' and 3' sides of the SNP, and the one closer to the SNP location

is selected. This step is incorporated to reduce the sample processing time and cost of testing. Note that if the Pyrosequencing primer is selected 3' from the SNP, the sequence of the PCR and Pyrosequencing primers printed above the sequence map are given as the reverse complement, as is necessary to correctly biotinylate, PCR amplify, and sequence the appropriate DNA strand during Pyrosequencing. Rules for developing Pyrosequencing primers include the initiation of primer design at a unique user-specified nucleotide sequence motif occurring as close as possible to the site of the SNP and within the PCR-amplified region. As was done for the PCR primers, the unique sequence motif defines the 3' end of the Pyrosequencing primer, which increases primer specificity during the sequencing reaction. The program requires that Pyrosequencing primers have a calculated melting temperature (T_m) of at least 40°C (default value) and lack secondary structure consisting of more than a user-specified number of adjacent nucleotide base pairs (14).

Web Interface

Default settings (summarized in Table 1) are presented for PCR and Pyrosequencing primer melting temperatures, minimum allowed secondary structure, and required length of the unique sequence motif at the 3' end of the Pyrosequencing primer. The application accepts input for the length of additional sequence harvested from the warehoused genomic databases in order to obtain extended sequences flanking the SNP for use during the design of PCR primers. However, the current version of SOP³, version 1, does not accept user input of individual sequences because the scope of the application is to aid in the development of primers for projects in which large batches of primer sets must be developed. The output can be parsed by gene name, SNP reference number, heterozygosity value, chromosome, or chromosomal location or function. The application allows one to choose the SNPs for which to develop PCR and Pyrosequencing primers based on the SNP location within each genetic structural element. For example, queries can be defined to

Figure 3. Calculated Pyrosequencing results for human insulin-associated SNP rs3842748. Oligonucleotide primers for PCR and Pyrosequencing are indicated in Figure 2. The SOP³ application provided a recommended order of nucleotide dispensation events and pyrogram for each (A and B) homozygous and (C) heterozygous genotype. The nucleotide sequence examined is indicated in each panel. SNP, single nucleotide polymorphism.

Table 2. Selected SNPs Developed for Pyrosequence-Based Typing Assays Using the SOP³ Primer Design Application

RefSeq	Gene Name	Description	Chromosome	Gene Element ^a	Alleles	Heterozygosity
rs913746	COL4A2	Collagen, type IV, α 2	13	Intron	AC	0.463
rs2272946	COL4A5	Collagen, type IV, α 5	X	Exon (nonsyn)	GT	Not reported
rs3747408	COL4A5	Collagen, type IV, α 5	X	Exon (syn)	AG	Not reported
rs2056402	ITGA2	Integrin, α 2	5	Intron	AT	0.198
rs3212504	ITGA2	Integrin, α 2	5	Intron	CG	0.326
rs27890	ITGA2	Integrin, α 2	5	Intron	CT	0.203
rs26679	ITGA2	Integrin, α 2	5	Intron	CG	0.5
rs1421941	ITGA2	Integrin, α 2	5	Intron	CT	0.395
rs3212594	ITGA2	Integrin, α 2	5	Exon (syn)	AG	0.192
rs17468	ITGB1	Integrin, β 1	10	UTR	CT	0.413
rs1316757	ITGB1	Integrin, β 1	10	Intron	AC	0.418
rs2230396	ITGB1	Integrin, β 1	10	Exon (syn)	AC	0.187

SNP, single nucleotide polymorphism.
^aGene element symbols are synonymous (syn), nonsynonymous (nonsyn), and untranslated region (UTR).

return PCR and Pyrosequencing primers for SNPs associated with introns, exons, as well as 5' and 3' untranslated regions, among others. SNPs associated with amino acid coding changes can also be specified, based on whether the change is synonymous or nonsynonymous. The entry of queries from which primers are to be designed is accomplished by typing gene names, reference sequence numbers, or chromosomal regions into the text box.

Figure 2 illustrates an example of the output page of the SOP³ primer design application. PCR and Pyrosequencing primers were developed for SNP rs3842748, which occurs within an intronic region of insulin. As indicated in the output (Figure 2), the gene name, SNP identifying reference sequence number, and nucleotide polymorphism are provided. The application also returns the dbSNP attributes for average heterozygosity and chromosomal location. Sequences of primers for Pyrosequencing and PCR are indicated along with the expected nucleotide length of the PCR amplification product. In the lower part of Figure 2 is the application's output of the FASTA-formatted sequence flanking the SNP. The predicted pyrograms are also provided with the application output and are shown for SNP rs3842748 (Figure 3) for homozygous (Figure 3,

A and B) and heterozygous (Figure 3C) individuals. Pyrograms plot nucleotide dispensation event versus luminescence from the Pyrosequencing reaction on the x- and y-axes, respectively. The predicted pyrograms, as illustrated in the figure, indicate the level of resolution provided by Pyrosequencing for distinguishing various alleles as well as suggesting a potential order of nucleotide dispensation events for each PSBT assay. The SOP³-generated nucleotide dispensation order is determined by combining the sequence of fixed and polymorphic residues, as given in the dbSNP database.

Multiple SOP³-Generated Primer Sets

The application was used successfully during primer design when tested against a variety of SNPs within loci correlated with risk toward developing diabetes or diabetes-associated complications (Table 2). Reference SNP identifiers, listed in Table 2, were used as input to query the SOP³ application. For each SNP that was examined, the application generated a list of candidate primers from which a trio of optimized primers was chosen (primer sequences are given in Supplementary Table S1). Using the 12 SNPs listed in Table 2 as an example, the web-based application

completed its analysis within 10 s. PCR amplification from genomic DNA was performed using primers developed for 48 SNPs and resulted in 38 (approximately 80%) validated Pyrosequencing assays. Selected examples are illustrated in Table 2. Figure 4 presents the results for SNP rs2056402 located within the gene encoding the α 2 subunit of integrin. Pyrosequencing traces from human genomic DNA samples are illustrated for homozygous (Figure 4, A and B) and heterozygous (Figure 4C) individuals selected from our pool of healthy volunteer donors. Of the

assays that failed, roughly half were due to poor amplification during PCR, while the rest yielded Pyrosequencing data in which an additional unrelated sequence was present, perhaps due to co-amplification of a pseudogene (data not shown). Table 2 summarizes information from the SOP³ application's output regarding gene name, reference sequence, genetic element location, allele, and average heterozygosity for a subset of the primers examined. The efficiency of DNA amplification using PCR primers designed by the SOP³ application is comparable with those reported for web-based software applications developed for designing PCR primers. These applications have reported that 90% or more of the PCR primers tested succeeded in generating PCR amplification products from genomic DNA (16,17). However, SOP³ is the first software application that can automatically generate PCR and Pyrosequencing primers surrounding one or multiple SNPs in a given genomic region.

DISCUSSION

PSBT of alleles consisting of SNPs as well as regions of polymorphic

DNA has been performed during gene mapping studies to associate disorders in human populations with specific genetic markers (4,6,7,18–21). The methodology is able to score hundreds of alleles daily, which makes it suitable for SNP screening studies involving large numbers of samples. Pyrosequencing provides distinct advantages for genetic typing in that samples can be assayed in 96-well trays, making it compatible with laboratory automation, thus increasing the rate of sample analysis, and the sensitivity of the technology requires little PCR-amplified material for each sequencing reaction (1). PSBT strategies can be designed using a minimal number of nucleotide dispensation events so that an entire 96-well tray can be assayed at a rate of roughly 1 min per base and approximately 10 min for a 10 nucle-

otide sequence, which is sufficient for the analysis of most SNP and deletion/insertion-containing alleles.

The web-based software application SOP³ was developed to enhance the use of PSBT for SNP scanning projects in which many SNP assays need to be developed. To that aim, the restrictions applied to PCR and Pyrosequencing primer design have been taken into consideration. For example, PSBT during SNP scanning projects requires PCR primers designed such that the 3' end of the amplified template does not fold to form a hairpin capable of initiating DNA extension during Pyrosequencing (7,12). Increased Pyrosequencing signal is proportional to the presence of increased concentration of DNA sequencing template in the reaction, a requirement achieved by the design of robust PCR

primers. Likewise, Pyrosequence signal is increased by the placement of the Pyrosequencing primer as close to the site of the SNP as possible while obeying the constraints associated with stringent annealing of primer at the Pyrosequencing reaction temperature (1,5). Additional stringency has been achieved by the addition of single-stranded binding protein (SSBP) to the Pyrosequencing reaction as reported previously (7,22).

The current version of the SOP³ application, version 1, was developed specifically for use in SNP screening projects involving human genomic DNA samples for the association of genetic markers with disease loci. Searching by gene name alone may omit some SNPs with sequence information

Figure 4. Observed Pyrosequencing results for SNP rs2056402. Primers for PCR and Pyrosequencing were chosen using the SOP³ application. Genomic DNA was obtained from healthy donor volunteers. Samples genotyped were (A) homozygous AA, (B) homozygous TT, and (C) heterozygous. The nucleotide dispensation order is indicated along the x-axis while the Pyrosequence is illustrated for each sample. Nucleotide disposition A6, following the polymorphic residues, was intended to serve as a negative control for background signal. SNP, single nucleotide polymorphism.

but without an associated locus name. At this point, only di-allelic SNPs are considered. Tri-allelic, or complicated insertion-deletion, are not accurately represented by the current version of the computer program. Primers sets will be designed by the software, however, the third (or fourth) nucleotide possibility of a complicated polymorphism will not be taken into account when searching for unique sequence regions at which to initiate primer selection. The creation of a version of the software for use with the mouse and other model organism genomes is underway and will be added during future updates of the web site. Moreover, the database of SNPs used in version 1 of the application has been downloaded from dbSNP at the NCBI. This database contains values for the average heterozygosity of each SNP as provided in dbSNP. Other databases, such as that at the Hapmap project, also contain lists of SNPs that can be warehoused and developed for use with the SOP³ application. These can be added in future versions and will allow primer design to focus on validated SNPs from these publicly available sources. The principal advantage of the current version of the SOP³ application is that it provides a single site for developing primers for PCR and Pyrosequencing, of which previous outputs have been tested and resulted in successful genotyping of human genomic DNA. Additional changes incorporated in future versions of the software application that will improve PSBT of alleles during genome scanning projects may include the development of databases for tracking primer design, which would allow users to report whether primers suggested by the SOP³ application resulted in useful Pyrosequencing data. Data warehousing of genomic sequences and polymorphisms has increased the efficiency of the design of useful primer sets for PCR and Pyrosequencing by providing a single application for generating these oligonucleotides for laboratory testing.

ACKNOWLEDGMENTS

We thank Patrick Hnidka and Kelly Downing for administrative assistance. This work was supported by funds from

Children's Hospital of Pittsburgh (S.R.) by grant no. MCB0316255 (P.V.B.) from the National Science Foundation (NSF), 1R01LM007994-01 (P.V.B.) from the National Institutes of Health-National Library of Medicine (NIH-NLM), U19-AI056374-01 from the Autoimmunity Centers of Excellence (S.R. and M.T.), R01DK24021 (M.T.) from the National Institutes of Health, and ERHS #00021010 (M.T.) from the Department of Defense.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing interests, commercial or otherwise, regarding the data presented in the manuscript.

REFERENCES

- Ronaghi, M., S. Karamohamed, B. Pettersson, M. Uhlen, and P. Nyren. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* 242:84-89.
- Ronaghi, M., M. Uhlen, and P. Nyren. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281:363-365.
- Fakhrai-Rad, H., N. Pourmand, and M. Ronaghi. 2002. Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. *Hum. Mutat.* 19:479-485.
- Alexander, A., L. Nichol, S. Ringquist, A. Styche, W. Rudert, and M. Trucco. 2002. Pyrosequencing sheds light on HLA genotyping. *Hum. Immunol.* 63(10 Suppl):S95.
- Gharizadeh, B., T. Nordstrom, A. Ahmadian, M. Ronaghi, and P. Nyren. 2002. Long-read pyrosequencing using pure 2'-deoxyadenosine-5'-O'-(1-thiotriphosphate) Sp-isomer. *Anal. Biochem.* 301:82-90.
- Ringquist, S., A.M. Alexander, W.A. Rudert, A. Styche, and M. Trucco. 2002. Pyrosequence based typing of alleles of the HLA-DQB1 gene. *BioTechniques* 33:166-175.
- Ringquist, S., A.M. Alexander, A. Styche, C. Pecoraro, W.A. Rudert, and M. Trucco. 2004. HLA class II DRB high resolution genotyping by pyrosequencing: comparison of group specific PCR and pyrosequencing primers. *Hum. Immunol.* 65:163-174.
- Garcia, C.A., A. Ahmadian, B. Gharizadeh, J. Lundberg, M. Ronaghi, and P. Nyren. 2000. Mutation detection by pyrosequencing: sequencing of exons 5-8 of the p53 tumor suppressor gene. *Gene* 253:249-257.
- Chen, S.H., C.Y. Lin, C.S. Cho, C.Z. Lo, and C.A. Hsiung. 2003. Primer design assistant (PDA): a web-based primer design tool. *Nucleic Acids Res.* 31:3751-3754.
- Emrich, S.J., M. Lowe, and A.L. Delcher. 2003. PROBEmer: a web-based software tool for selection optimal DNA oligos. *Nucleic Acids Res.* 31:3746-3750.
- Stein, L.D. 2003. Integrating biological databases. *Nat. Rev. Genet.* 4:337-345.
- Ronaghi, M., B. Pettersson, M. Uhlen, and P. Nyren. 1998. PCR-introduced loop structure as primer in DNA sequencing. *BioTechniques* 25:876-878.
- Mullis, K.B. and F.A. Faloona. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* 155:321-210.
- Schildkraut, C. and S. Lifson. 1965. Dependence of the melting temperature of DNA on salt concentration. *Biopolymers* 3:195-208.
- Utting, M., J. Hampe, M. Platzer, and K. Huse. 2004. Locking of 3' ends of single-stranded DNA templates for improved Pyrosequencing performance. *BioTechniques* 37:66-73.
- Varotto, C., E. Richly, F. Salamini, and D. Leister. 2001. GST-PRIME: a genome-wide primer design software for the generation of gene sequence tags. *Nucleic Acids Res.* 29:4373-4377.
- Haas, S.A., M. Hild, A.P.H. Wright, T. Hain, D. Talibi, and M. Vingron. 2003. Genome-scale design of PCR primers and long oligomers for DNA microarrays. *Nucleic Acids Res.* 31:5576-5581.
- Ferraris, A., E. Rappaport, R. Santacroce, E. Pollak, I. Krantz, S. Toth, F. Lysholm, M. Margaglione, et al. 2002. Pyrosequencing for detection of mutations in the connexin 26 (GJB2) and mitochondrial 12S RNA (MTRNR1) genes associated with hereditary hearing loss. *Hum. Mutat.* 20:312-320.
- Ahluwalia, R., R. Freimuth, H.L. McLeod, and S. Marsh. 2003. Use of pyrosequencing to detect clinically relevant polymorphisms in dihydropyrimidine dehydrogenase. *Clin. Chem.* 49:1661-1664.
- Palmieri, O., S. Toth, A. Ferraris, A. Andriulli, A. Latiano, V. Annese, B. Dallapiccola, M. Vecchi, et al. 2003. CARD15 genotyping in inflammatory bowel disease patients by multiplex pyrosequencing. *Clin. Chem.* 49:1675-1679.
- Haglund, S., M. Lindqvist, S. Almer, C. Peterson, and J. Taipalensuu. 2004. Pyrosequencing of TPMT alleles in a general Swedish population and in patients with inflammatory bowel disease. *Clin. Chem.* 50:288-295.
- Ronaghi, M. 2000. Improved performance of pyrosequencing using single-stranded DNA-binding protein. *Anal. Biochem.* 286:282-288.

Received 19 July 2004; accepted 5 August 2004.

Address correspondence to:

Massimo Trucco
Division of Immunogenetics
Rangos Research Center, Children's Hospital of Pittsburgh
3460 Fifth Avenue
Pittsburgh, PA 15213, USA
e-mail: mnt@pitt.edu