

Software

Open Access

MPrime: efficient large scale multiple primer and oligonucleotide design for customized gene microarrays

Eric C Rouchka*^{1,3}, Abdelnaby Khalyfa^{2,3} and Nigel GF Cooper^{2,3}

Address: ¹Department of Computer Engineering and Computer Science, Speed School of Engineering, University of Louisville, Louisville, Kentucky, USA, ²Department of Anatomical Sciences and Neurobiology, University of Louisville School of Medicine, Louisville, Kentucky, USA and ³Bioinformatics Research Group, University of Louisville, Louisville, Kentucky, USA

Email: Eric C Rouchka* - eric.rouchka@louisville.edu; Abdelnaby Khalyfa - a.khalyfa@louisville.edu; Nigel GF Cooper - nigelcooper@louisville.edu

* Corresponding author

Published: 13 July 2005

Received: 02 February 2005

BMC Bioinformatics 2005, **6**:175 doi:10.1186/1471-2105-6-175

Accepted: 13 July 2005

This article is available from: <http://www.biomedcentral.com/1471-2105/6/175>

© 2005 Rouchka et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Enhancements in sequencing technology have recently yielded assemblies of large genomes including rat, mouse, human, fruit fly, and zebrafish. The availability of large-scale genomic and genic sequence data coupled with advances in microarray technology have made it possible to study the expression of large numbers of sequence products under several different conditions in days where traditional molecular biology techniques might have taken months, or even years. Therefore, to efficiently study a number of gene products associated with a disease, pathway, or other biological process, it is necessary to be able to design primer pairs or oligonucleotides en masse rather than using a time consuming and laborious gene-by-gene method.

Results: We have developed an integrated system, MPrime, in order to efficiently calculate primer pairs or specific oligonucleotides for multiple genic regions based on a keyword, gene name, accession number, or sequence fasta format within the rat, mouse, human, fruit fly, and zebrafish genomes. A set of products created for mouse housekeeping genes from MPrime-designed primer pairs has been validated using both PCR-amplification and DNA sequencing.

Conclusion: These results indicate MPrime accurately incorporates standard PCR primer design characteristics to produce high scoring primer pairs for genes of interest. In addition, sequence similarity for a set of oligonucleotides constructed for the same set of genes indicates high specificity in oligo design.

Background

Recent advances in DNA sequencing technologies have resulted in the availability of the assembly of a number of genomes at various stages. As of December 24, 2004, the Genomes OnLine Database [1] lists 1,245 on-going or completed genome projects. Included are the genomes of human [2,3], mouse [4], Norwegian brown rat [5], fruit fly [6], fugu [7], chimpanzee, and zebrafish [8]. In addition,

various projects have been concerned with the curation of the genic regions of these genomes, most notably RefSeq [9].

Advances in microarray technology [10] have made it possible to simultaneously study the expression levels of thousands of genes under a given condition. Since a molecular biologist may be interested in studying a

Table 1: Comparison of properties of MPrime with various primer and oligonucleotide design programs.

Program	Reference	Primers	Oligos	Large Scale Design	Sequence Similarity Avoidance
Primer Master	Proutski and Holmes (1996) [19]	X			
PRIMO	Li, et al. (1997) [17]	X			
GenomePRIDE	Haas, et al. (1998) [14]	X	X	X	X
Primer3	Rozen and Skaletsky (2000) [21]	X		X	
DOPRIMER	Kampke et al., (2001) [15]	X		X	
PrimeArray	Raddatz et al. (2001) [20]	X		X	
GST-PRIME	Varotto et al. (2001) [23]	X		X	
ProbeSelect	Li and Stormo (2001) [25]		X	X	X
PROBE	Pozhitkov and Tautz (2002) [27]		X	X	X
PRIMEGENS	Xu et al. (2002) [24]	X		X	X
OligoWiz	Nielsen et al. (2003) [26]		X	X	X
OligoArray 2.0	Rouillard et al. (2003) [29]		X	X	X
OligoPicker	Wang and Seed (2003) [30]		X	X	X
ROSO	Reymond, et al. (2004) [28]		X	X	X
MPrime	2005	X	X	X	X

particular subset of genes in a given genome using microarray experimentation, it becomes necessary for the scientist to design complementary sequences to uniquely identify the genes of interest for either a cDNA product [10] or synthetic oligonucleotide [11-13] approach.

Custom microarray experimentation involves several steps, each of which must be performed in a timely manner in order to avoid becoming a bottleneck in the system. The steps involved are: determination of target genes; designing primers or oligonucleotides to create products uniquely identifying the target genes; spotting of the products on the microarray slide; detection of the presence of the gene under a given condition; and data analysis. MPrime focuses on the primer and oligonucleotide design stage in the microarray experimentation process.

Comparison to existing primer and oligo design software

Within the past few years, several algorithms for primer design have become available [14-21] including some specifically targeting whole genome analysis for microarrays [22-24]. In addition, more recent publications have become available discussing methods for oligo design [25-30], due to the emergence of custom oligonucleotide arrays. However, many of the primer design programs do not consider product similarity for cross hybridization, in part due to the lack of available genomic sequence at the time of their publication (see table 1).

There are a number of limitations to these approaches. Some of the algorithms find a primer or oligo from a single sequence at a time. Such an approach is time consuming and prone to human error, since the results must be transferred one at a time. Other algorithms may consider

multiple sequences, yet fail to consider a sequence similarity score, resulting in products which do not uniquely identify genes. Some approaches pre-compute a set of products for a given genome, which works well if the complete genome is available, but behaves poorly in an adaptive environment. Very few of these approaches allow the user to specify whether they would like to design either primers or oligos, limiting them to one or the other.

MPrime attempts to overcome each of these limitations by allowing for efficient large scale multiple product design. MPrime approaches primer and oligo design by looking at a user-defined subset of genes of interest identified by either a GenBank [31] or RefSeq [9] accession number, by a gene name, or by the raw fasta sequence data. When the user specifies a particular genome of interest, MPrime will search through that genome to ensure that the products chosen will uniquely identify the gene the user is trying to identify.

Due to the methods employed, MPrime can be useful in large-scale primer and oligo design. Two software packages, GenomePRIDE [14] and GST-PRIME [23], have been described as offering large-scale, genome-wide design of oligomers and primer pairs for the construction of custom microarray products. MPrime provides advantages to each of these approaches.

GST-PRIME allows for an automated, genome-wide design of primer pairs in a similar fashion to MPrime. MPrime improves upon the functionality of GST-PRIME by allowing for the construction of gene-specific oligos for use with oligo-based custom microarrays. In addition, while GST-PRIME allows for the detection of primer pairs

for a cDNA based chip on a whole genome, users of the system are limited to only choosing genes associated with a protein GenBank Identifier (GI) number. This inherently imposes limitations for the user. For instance, computationally predicted genes, genes supported by biological evidence (ESTs for example), or pre-published gene sequences cannot be used. MPrime avoids these complications by allowing the user to obtain the gene information directly from the GI number, or optionally from a raw fasta formatted sequence.

GenomePRIDE is the only known software package that incorporates high throughput primer pair and oligo design. MPrime distinguishes itself from GenomePRIDE by taking into account several factors, including GC content, secondary structure loop formation, and the presence of a GC clamp. While omitting these calculations makes the detection of primers and oligos much more efficient, it can also potentially lead to problems with both PCR product amplification and probe to product binding affinities. In addition, MPrime is freely available via a web interface, whereas GenomePRIDE has an associated cost of 150 euros for academic institutions.

Implementation

Selection of genic regions

The first step in primer design is to decide the regions to study. The interests of our collaborating lab lie in collating sets of mammalian programmed cell death genes related to neurological diseases such as Alzheimer's, Parkinson's, and Huntington's for use in microarray analyses. Therefore, an interface has been incorporated into MPrime in order to allow the user to retrieve genic regions by either a GenBank accession, gene name, or by a keyword. Retrieval by a keyword allows the user to have a general idea of the region of interest without having to spend time searching for an exact gene name or GenBank accession. Potential candidates are presented to the user. MPrime allows the user to refine their results by deciding on which candidate regions will be included before proceeding to primer or oligo design.

Primer design calculations

Although there are potentially a number of applications for the use of MPrime, it was originally designed with the intent for the construction of primers for custom cDNA microarray chips. In order to obtain the primer products with a high throughput, multiplex PCR can be implemented as the underlying biological experimentation. To ensure that the multiplex PCR experiments function properly and the microarray experimentation produces results that can be meaningfully analyzed the primer products should adhere to the properties of uniformity, sensitivity, and specificity.

Product uniformity

Multiplex PCR [32] is a technique allowing for the amplification of two or more targets in an organism of interest by incorporating multiple sets of PCR primers into a single reaction. Since each set of primers is subject to the exact same reaction conditions, they must be defined uniformly in order to react to the thermo cycling in a similar fashion. The properties which need to be relatively uniform include primer melting temperature, primer G+C concentration, and primer length.

MPrime requires the user to input a range of primer melting temperatures along with an optimal melting temperature. Any primers designed are required to fall inside of this range. There are a number of different approaches to calculating the melting temperature (T_m) of an oligonucleotide sequence, including an arbitrary method based on the formula $T_m = 2 * (A + T) + 4 * (G + C)$ [33]; a long probe method based on the length and GC concentration [34]; and a nearest neighbor method based on dinucleotide interactions [35,36]. MPrime incorporates the nearest neighbor approach using the formula of Rychlik, et al. [36] where the T_m is calculated as follows:

$$T_m \equiv \frac{-1000 * \Delta H}{-10.8 - \Delta S + R \ln\left(\frac{c}{4}\right)} - 273.15 - 16.6 * \log_{10} M \quad (1)$$

Where ΔH is the enthalpy for helix formation, ΔS is the entropy for helix formation, c is the molar concentration of the primer (set at 250 pM); M is the molar concentration of Na⁺ (set at 50 mM) and R is the gas constant (1.987 cal/degree * mol).

Calculation of the primer G+C concentration is a straightforward calculation of the percentage of the primer oligo containing G's or C's. MPrime requires the user to input the minimum, maximum, and optimal G+C concentration for the primers. In addition, the user can specify a GC clamp length, or the number of G's or C's required at the 3' end of the primer. The GC clamp helps to promote primer end stability, resulting in more efficient and specific bonding of the primer to the amplicon template.

In addition, MPrime requires the user to enter information concerning the primer length. The minimum, maximum, and optimal primer lengths are required. The primer length uniformity along with the uniform G+C content helps to maintain a uniform primer melting temperature.

Product sensitivity

PCR amplification can potentially fail due to either self or paired annealing. There are three main cases to be considered. In the first case, if either of the individual primers have a large number of consecutive complementary bases

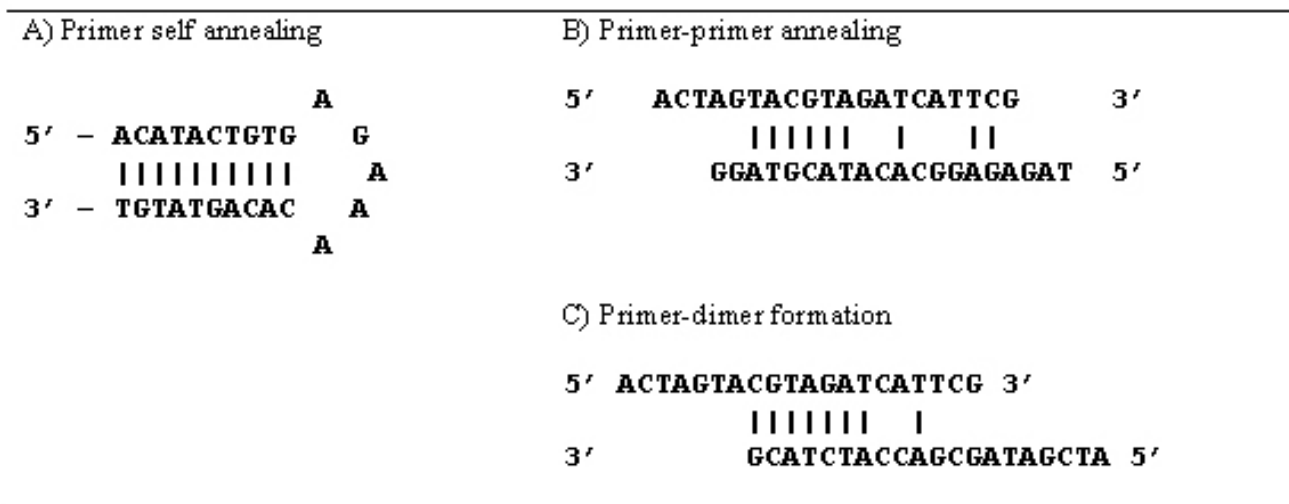


Figure 1
Primer secondary structure formation. Shown in figure 1A) is the secondary structure stem loop formation for the primer ACATACTGTGAGAAACACAGTATGT. Figure 1B) illustrates an example double-stranded structure formation between the primers ACTAGTACGTAGATCATTCG and GGATGCATACACGGAGAGAT. Note the run of six straight matches between the two sequences. Shown in figure 1C) is the primer-dimer formation between the primers ACTAGTACGTAGATCATTCG and GCATCTACCAGCGATAGCTA. Note the 3' end of the second primer has a run of seven straight matches to the first. The scores for each of these is based on the formula $4 * |G + C| + 2 * |A+T|$. For figure 1A) there are 6 A-T base pairs, and 4 G-C base pairs, yielding a score of 28. For figure 1B), there are 6 A-T base pairs, and 3 G-C base pairs, yielding a score of 24. For figure 1C), there are 4 A-T base pairs and 3 G-C base pairs on the 3' end of the second sequence, yielding a score of 20.

within their sequences they can potentially self-anneal to form a secondary structure such as a stem loop (see figure 1A). In the second case, complementary bases in paired primers can interact in their mid-regions to form partial double stranded structures (see figure 1B). The final case occurs when the primers interact with each other at the 3' end of either primer to produce primer-dimers (see figure 1C).

One problem that can occur in product design is that either the products or the corresponding probe sequences could form secondary or tertiary structures, thus causing them to fail to interact, resulting in an apparent low signal. In order to overcome these problems of unwanted annealing between primers and products, MPrime incorporates a scoring scheme for paired-end and self-end annealing to reduce primer dimers; paired annealing to reduce partial double stranded structures; and self annealing to reduce secondary structure formation of single stranded sequences within both products and individual primers. These scoring schemes are adapted from Kampke, et al. [15] where each G-C base pair is assigned a score of 4 and each A-T base pairing is assigned a score of 2.

Product specificity

It is important that each primer set results in only the product of interest, and that each product uniquely identifies the gene being studied. In order to build in these constraints, MPrime searches the primers against a RefSeq [9] database for the organism being studied using wublastn [37]. The highest scoring sequences should be the genic region of interest; all other similar, suboptimal alignments will result from spurious hits. In order to assure that the product designed is specific to the gene of interest, the blast scores for all database matches are considered, and a specificity score is computed as the percent identity of the match multiplied by the percent of the product covered by the match. The blast results are then resorted based on this specificity score. The native locus should have a score of 1 (100% identity * 100% coverage). The next highest scoring segment is then taken to give an idea of the specificity of the designed product. For instance, if the product is completely unique, the second highest score will be 0. If the next highest scoring segment has an 80% identity of 60% of the sequence, the score would be 0.48. The product specificity score is weighted and calculated into the overall primer score. Note that in some instances, a gene will have duplicate entries in Ref-

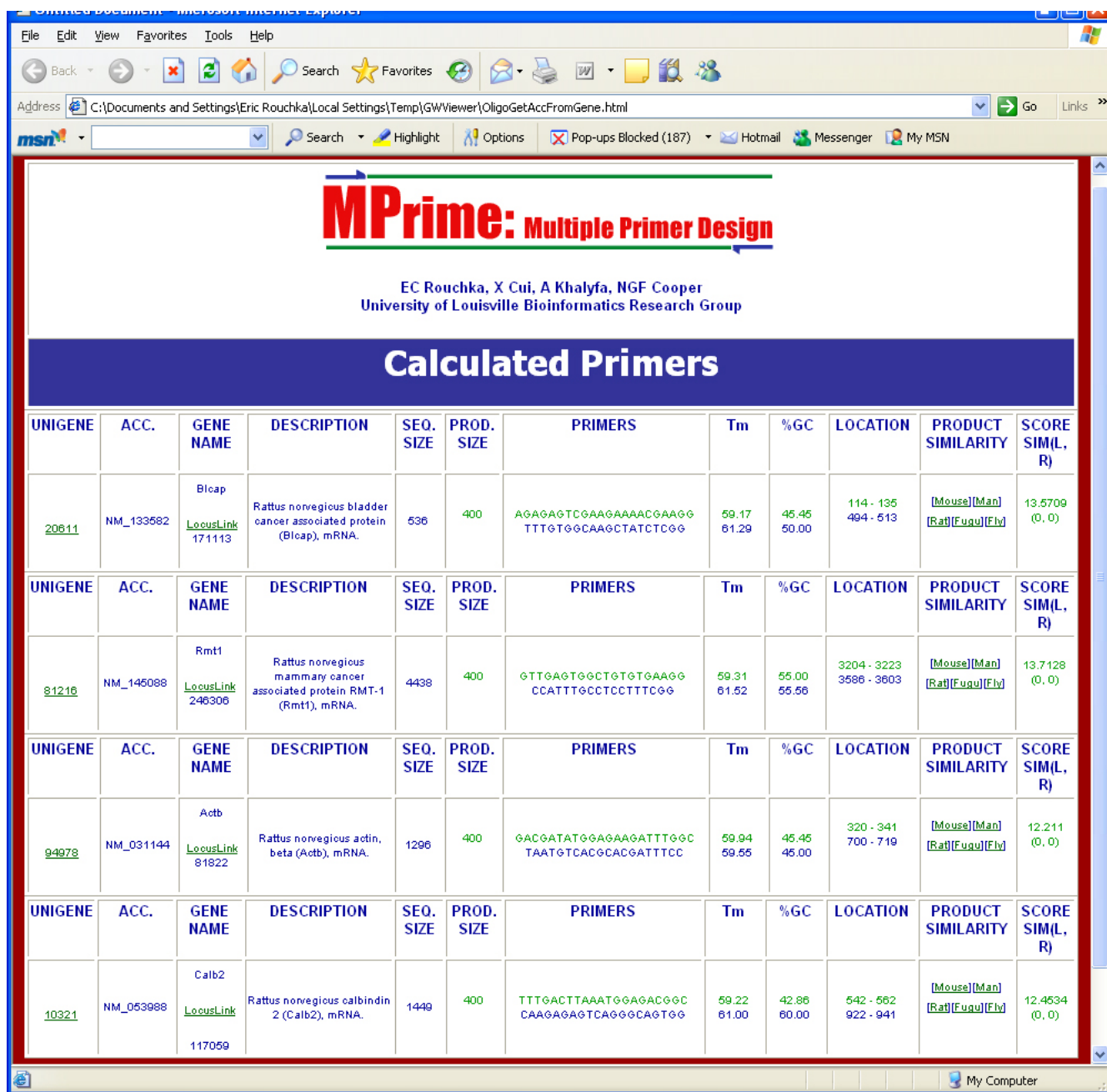


Figure 2
Web interface and example of primer pair output. The initial inputs were for the rat genome with GenBank accession NM_053986, gene ACTB, and keyword "cancer". The default values for the primer size, GC%, melting temperature, product size, and GC clamp length were used. The user was able to refine these choices to include the four genic regions indicated in the results. A similar screen is produced when oligonucleotides are chosen as well (not shown).

Seq. This may result in the product specificity score of 1 for all possible products. In such a case, the best scoring primer set is reported along with the product specificity score. In these cases, it is up to the user to determine whether or not such a primer set should be used.

MPrime incorporates an overall scoring scheme for each pair of primers. The basis behind this scheme is to find the best scoring primer pair with the smallest deviation from the overall optimal values. The smaller the score is, the more likely it is that the primer pair will function as

Table 2: Sample set of genes and resulting primer pairs using MPrime 1.0. The top sequence in each primer pair is the forward primer, the second sequence is the reverse primer.

Gene Name	Product Size	Primer Pairs
GAPDH	400	AATGTGTCGGTCGTGGATCT GGGTCTGGGATGGAAATTGT
Actb	400	GGCGCTTTTGACTCAGGATT AGTTGGGGGACAAAAAAA
Sdf4	400	AGACCTGCCAACCACTCATC TGCTTGCCAAAAAATCTCACT
Ubb	400	TTCTGTGAGGGTGTGAGGGT TTTATCCTGGATCTTGGCCT
Ywhaz	400	ACAATGTTCTTGCCCATGT AGGAAGAGGAGGAGGAAGGA
Hprt	400	TTATGCCGAGGATTTGGAAA AACCTTAACCTTTTGGGGC
Tubd1	400	TAAGATGCTGGGTGTCGGTA TAAGAGCTGGCTGTTGCTGA
Tuba1	400	GACCCTCGCCATGGTAAATA AATCCACACCAACCTCCTCA
Ppicap	400	ACTCCCTCCCTCTTTCCTG CAGCAGAGAAAAGCTCCACC
Hspca	400	AGGAAACCCAGACCCAAGAC ACACCAAACCTGGCCAATCAT
Htf9c	400	TGTGAATTCCTGGTCGGAGT TCTTCTCTGCCCTCCTCC

desired. Each of the parameters entered is weighted evenly, with the exception that a larger weight is associated to product specificity. Similar scoring schemes have been proposed in Kampke, et al. [15] and are used in Primer3 [21].

Oligonucleotide calculations

MPrime also has the capability of calculating oligonucleotides for sequences of interest. The scoring scheme for determining the optimal oligonucleotides is very similar with the exception that there are no primer-primer interactions that can occur. With oligonucleotides, the similarity scores are weighted slightly higher, and are based on the actual length of the oligos desired. Typically, the 3' end of genes tends to be more variable and therefore more unique in their sequence. In addition, MPrime has the added feature of allowing the user to specify a region in which the oligo must be located.

Sequence input into MPrime

Sequences can be input into MPrime in one of four ways: by gene name, by GenBank accession number, by keyword, or by the raw fasta sequence data. The user can specify whether they are interested in sequences from the human, rat, mouse, fruit fly, or zebrafish genome. MPrime will then search for the sequence of interest, and return sequence records from RefSeq and the appropriate GenBank databases matching the search string. If the GenBank

accession is given, MPrime first checks to see if there is a gene matching exactly.

Once the regions of interest are decided upon, the sequences are retrieved from local copies of either RefSeq or GenBank. For cDNAs, these sequences are then searched in order to retrieve primer pairs adhering to a set of guidelines. The primer length, G+C content, melting temperature, self-hybridization characteristics, and primer-primer hybridization characteristics must be taken into account in order to allow the primers to effectively be used in PCR experiments [15]. In addition, information on the sequence product is taken in as well, including the product length, G+C content, and melting temperature. The MPrime software was developed using both Perl and C++.

Since the end goal is to create products for incorporation into gene microarray chips, it is extremely important to provide relatively uniform product lengths in order to provide equal concentrations of sequence products and to be able to distinguish between the expression levels of different genes within a single experiment. The MPrime interface allows the user to enter in a range of values along with an optimal product length.

After potential primer pairs are created, they are searched against the appropriate database of genomic sequence in order to ensure the primer pairs and subsequent products are unique. If the products on the microarray chip are to uniquely identify genes, then it is necessary to check that the resulting products do not represent domains repeated in several genes.

Validation of MPrime

Sequence products produced by primer pairs selected from MPrime were validated using two independent techniques: (i) polymerase chain reaction and (ii) sequencing using the Beckman Coulter CEQ8000 Genetic Analysis System. Oligonucleotides constructed from MPrime were checked for their specificity by a nucleotide sequence search against *Mus musculus* sequences.

Polymerase chain reaction (PCR)

After primer pairs are determined, the sequence products are amplified using a 96 well plate DNA Engine Tetrad Thermal Cycler (MJ Research, Inc., Boston, MA). A Multi-PROBE II Liquid Handler (Packard Instruments, Boston, MA) is used to help automate the process of creating the 96 well trays used in PCR amplification. The resulting products are then spotted on customized microarrays using a BioChip Arrayer (Packard Instruments).

PCR was performed using mouse brain total RNA (Clontech, Palo Alto, CA) and reverse transcribed into cDNA.

RT-PCR was performed for a set of housekeeping genes and the primers used for RT-PCR were the same as those designed by MPrime program. Briefly, 5 µl of total RNA was used reverse transcribed using SuperScript II (Invitrogen, Carlsbad, CA) as described by the manufacture protocols. PCR reaction including 10 µl of cDNA template in a 100 µl reaction volume was amplified using DNA Engine Tetrad (MJ Research, Inc., Boston, MA). The thermal cycling profile consisted of 95°C for 3 min, 94°C for 30 s, 60°C for 30 s, 72°C for 30 s and 72°C for 7 min. Cycling kinetics were performed using 30 cycles. Amplified PCR reactions were separated on 2% agarose gel in the presence of ethidium bromide for visualization.

Sequencing

Purified mouse housekeeping genes products were sequenced using the CEQ 8000 Genetic Analysis System (Beckman Coulter Inc., Fullerton, CA). PCR was carried out using left primer and target cDNA from mouse brain as described above for 25 cycles. Sequences were then compared using a blast search against the corresponding GenBank and/or RefSeq entry for verification.

Oligonucleotide specificity

MPrime-computed oligonucleotides were compared against sequences from the organism *Mus musculus* using NCBI Blast [38]. The nucleotide-nucleotide Blast program blastn was used with default parameters. The advanced search option limiting to the *Mus musculus* dataset was used. Similar sequences reporting a bit score of 99.7 or higher (five mismatches in 70 bases) were reported as significant alignments for the purpose of oligonucleotide specificity.

Results

MPrime has allowed us to create multiple primers for related genic regions in a short span of time which are reported back to the user in the form of a web interface that can be stored as a tab delimited file that can be read into an Excel spreadsheet (figure 2). In our hands, detection of primers on a set of 138 rat actin, myosin and various muscle related genes with an average size of 2.7 KB finished in approximately 80 seconds on a dual processor AMD Athlon system with 1 GB of memory (results not shown). Analysis of the primers designed using the MPrime strategy compare well to the widely used Primer3 [21] software (results not shown).

A set of eleven randomly chosen mouse housekeeping genes (GAPDH, Actb, Sdf4, Ubb, Ywhaz, Hprt, Tubd1, Tuba1, Ppicap, Hspca, and Htf9c) was chosen for study. The resulting MPrime primer pairs detected for each of these genes is located in Table 2. The results of the PCR amplification showing products on the order of 400 bases for each of the primer pairs for these genes are shown (Fig-

ure 3). Sequencing of these products and subsequent similarity searches led to the identification of uniquely identifiable regions for each of these eleven product pairs. Sample results for the stromal cell derived factor 4 (Sdf4) are shown in terms of both the sequence (Figure 4) and the resulting sequence similarity results (Figure 5). The remaining ten sequences show similar results in terms of both their sequence composition and sequence similarity, validating that the sample set of primers produces the desired products. Thus, MPrime provides an effective alternative to finding primers one gene product at a time.

MPrime-calculated oligonucleotides of length 70 for the set of eleven housekeeping genes are shown (Table 3). Each oligo was designed to occur in the last 500 bases on the 3' end of the sequences. The oligos were searched against the *Mus musculus* sequences in NCBI's database using NCBI Blast [38] with the default parameters. Alignments resulting in five or fewer matches over the entire length of the oligo were reported and summarized (Table 4). The resulting sequence similarities and bit scores are given. A perfect bit score for an oligo of length 70 using the default blast parameters is 139. The results show that ten of the eleven housekeeping genes map uniquely to a single RefSeq gene sequence, with the exception being GAPDH. Nine of these ten also map to a single genomic locus, with the exception of Ubb, which maps to three separate loci (including the native locus). Ubb, Hprt, Tubd1, Tuba1, Hspca, and Htf9c also map to additional sequences not shown in the table. In each case, these additional sequences represented either redundant mRNA or cDNA sequences for the corresponding gene, or their corresponding chromosomal location. For these ten sequences, the results suggest MPrime produces highly specific oligonucleotide sequences.

For GAPDH, there were not any oligonucleotides found in the 500 bases on the 3' end of the RefSeq sequence of length 70 that were unique to GAPDH. The highest scoring oligo resulted in 37 matches with a perfect bit score of 139. This is reflected in the high sequence similarity score reported by MPrime. For these 37 matches, fourteen were to predicted similar mRNAs; five were similar to GAPDH; and four were GAPDH mRNA. The remaining fourteen sequences were genomic matches to chromosome 2 (2 sequences) chromosome 3 (3), chromosome 4 (2), chromosome 7 (2), chromosome 8 (3), chromosome 10 (1), and chromosome 15 (1).

MPrime uses RefSeq as the default database for searching for gene sequences. In order to understand more fully what happens with GAPDH, RefSeq was examined more closely. A search of RefSeq with the constraint "*Mus musculus* similar to glyceraldehyde-3" "*Mus musculus*"

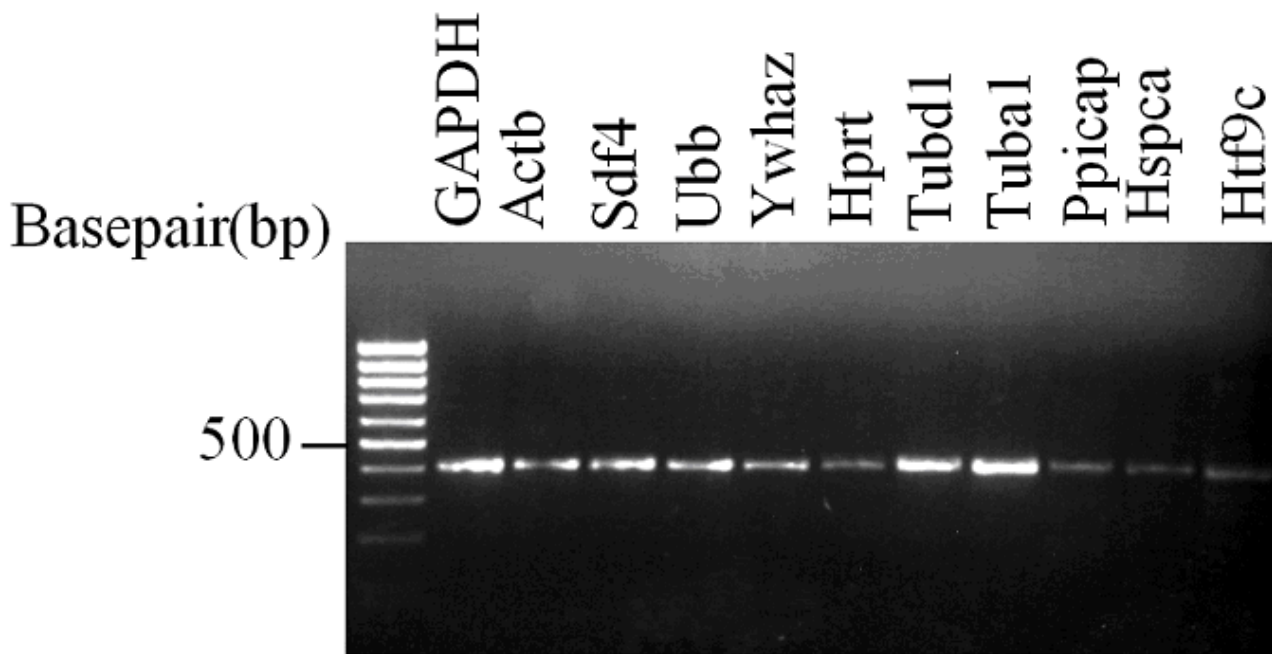


Figure 3
PCR amplification results for eleven mouse housekeeping genes. The resulting bands indicate single products with product sizes near 400 bases in length.

```
TANCATGGGAGGAAAATGAGATCNTGCCCCAGACCACCTGAATGGGGTGAAGNCTGGAGATGGATTGGAC
ACCTCAACAAGGACTTCCATCAGGAGGTTTTCTGGGAAAGGACATGGATTGGGGTTGANTTNAAGACTCA
CAGACCCACGGANAAACCCGGAGGAAGCTGATGGTCATCTTTTCCAAGGTAGACGTGAACACTGACCGGAG
GATCAGCGCTAAGGAGATGCAGCACTGGATTATGGAGAAAACAGCAGAGCACTTCCAGGAGGCCGTCAAGG
AAAACAAACTGCCTTCAGGGCTGTGGACCCTGACGGTGACGGCCATGTTTCTGGGATGAATATAAAGTG
AAGTTTTTGGCAAGCCNNAAAA
```

Figure 4
 Sequencing results for the primer product for mouse stromal cell derived factor 4 (Sdf4) sequence.

[ORGN] results in 118 different sequences, many of which completely cover the GAPDH sequence used by MPrime. Since the GAPDH sequence is not unique within the set of RefSeq sequences, a unique 70-mer could not be detected.

Discussion

In addition to the detection of primer products and single, long oligomers (on the order of 40–70 bases) that uniquely identify a gene of interest, individual researchers may want to design multiple probes for a single gene. Such an approach can help to limit the effects of cross-hybridization, the presence of multiple gene isoforms,

and single nucleotide polymorphisms. An approach to design multiple short oligonucleotide probes (on the order of 25 bases long) has previously been presented [39]. MPrime presents the user the option of reporting back more than one resulting primer product or oligonucleotide for each gene. These results are reported in order based on their optimal score, with a score of 0 having the least deviation from the optimal user-selected conditions.

Researchers may additionally be interested in looking at multiple overlapping probes for each gene to produce tiled arrays that can help to detect the presence of different


```

 >gi|46195814|ref|NM_011341.3| GU Mus musculus stromal cell derived
factor 4 (Sdf4), mRNA
      Length = 5193

Score = 573 bits (289), Expect = e-161
Identities = 350/364 (96%), Gaps = 8/364 (2%)
Strand = Plus / Plus

Query: 7   gggaggaaaatgagatcntgccccagaccacctgaatggggtgaagmctggagatggat 66
          |||
Sbjct: 577 gggaggaaaatgagatcatgccccagaccacctgaatggggtgaag-ctggagatggat 635

Query: 67   tggacacctcaacaaggacttccatcaggaggtttctctgggaaaggacatggattgggg 126
          |||
Sbjct: 636 -ggacacctcaacaaggacttccatcaggaggtttctctgggaaaggacatggat-gggt 693

Query: 127  ttganttnaagactcacagaccacgganaaacccggaggaagctgatggtcatcttttc 186
          |||
Sbjct: 694 ttga--tgaagactcagag--ccacggagaagcc-ggaggaagctgatggtcatcttttc 748

Query: 187  caaggtagacgtgaacactgaccggaggatcagcgctaaggagatgcagcactggattat 246
          |||
Sbjct: 749 caaggtagacgtgaacactgaccggaggatcagcgctaaggagatgcagcactggattat 808

Query: 247  ggagaaaacagcagagcacttccaggaggc cgtcaaggaaaacaaactgcacttcagggc 306
          |||
Sbjct: 809 ggagaaaacagcagagcacttccaggaggc cgtcaaggaaaacaaactgcacttcagggc 868

Query: 307  tgtggaccctgacggtgacggccatgttctctgggatgaatataaagtgaagttttggc 366
          |||
Sbjct: 869 tgtggaccctgacggtgacggccatgttctctgggatgaatataaagtgaagttttggc 928

Query: 367  aagc 370
          |||
Sbjct: 929 aagc 932
    
```

Figure 5
 Comparison of the PCR product for the mouse stromal cell derived factor 4 (Sdf4) obtained from the MPrime PCR primers (Figure 4) to the RefSeq sequence NM_011341.3 [GenBank:46195814].

isoforms or single nucleotide polymorphisms present in an organism [40]. This option is not currently available in MPrime, but is a planned addition in a future release.

Conclusion

MPrime has proven to be an efficient and effective tool for both primer and oligonucleotide design. The methodol-

ogy behind MPrime allows molecular biologists to construct a large number of primers and/or oligonucleotides for genes that need to have consistent properties. In addition, multiple sub-optimal results can be reported and tested. Preliminary tests on a set of housekeeping genes indicate that primer products obtained by MPrime-designed primer pairs produce uniquely identifying

Table 3: Computed 70-mer oligos for eleven mouse housekeeping genes using MPrime 1.3 default parameters.

Gene name	70-mer oligonucleotide sequence
GAPDH	ACTTTGTCAAGCTCATTTCCTGGTATGACAATGAATACGGCTACAGCAACAGGGTGGTGGACCTCATGGC
Actb	AATAGTCATTCCAAGTATCCATGAAATAAGTGGTTACAGGAAGTCCCTCACCTCCCAAAAAGCCACCC
Sdf4	TGTTAAAAGAAAACATGAAGAGAGCTGTGGCTCTAGCTCAGTGGTCGAACGCTGCCAGCAAGTAAAACG
Ubb	CCTCCGTCTGAGGGGTGGCTATTAATTATTCGGTCTGCATTCCAGTGGGCAGTGATGGCATTACTCTGC
Ywhaz	CTGTCACCGTCTCCCTTTAAAATCCTTCCTCCTCCTTCCTCCTCCTCCTCCTCCTCACATAATGATGG
Hprt	TTTTAGAAATGTCAGTTGCTGCGTCCCCAGACTTTTGATTTGCACTATGAGCCTATAGGCCAGCCTACCC
Tubd1	TTTGGACAGCTTTGCATTGTTGGAGCAAGTTGTTGCCAGTTATGGTAGTCTTGGACCCTAAGCCAAGAGG
Tuba1	GCTCCACAGGGATGTTTATTGTGTTCCAACACAGAAAGTTGTGGTCTGATCAGTTAATTTCTATGTGGC
Ppicap	ATTGTATTCAAATGAAAATTTACTAGAAGTTTCAGCCAGCACTCACTCCAGGACTGAGAGTCCCAGGGC
Hspca	TTAAAACAACCTGACAGGAATCCCCAAGTGGCTTGTTCCTCAAAGTCCCGAGAACAACCCCTAAGTTTCC
Htf9c	CAGACTCCGCACTGTGAGATGCTTATCCTGTTTGTGAGAGGATGCAACAACCCCCAATGGCATAGAAGCCC

Table 4: Blastn similarities to MPrime computed oligos. mRNA sequence from which the RefSeq sequence was derived.

Gene name	GenBank identifier	Score (bits)	Notes
GAPDH	47607489	139	RefSeq entry for GAPDH
Actb	6671508	139	RefSeq entry for Actb
	49865 ¹	139	mRNA for beta actin
	26104752	139	beta actin cDNA
	191660	139	Mouse A-X beta actin mRNA
	22316184	99.6	Chromosome X genomic sequence
Sdf4	46195814	139	RefSeq entry for Sdf4
	47568754 ¹	139	mRNA for Sdf4
	21953005	139	Chromosome 4 genomic sequence
	26097254	131	Unknown EST
Ubb	6755918	139	RefSeq entry for Ubb
	55177 ¹	139	mRNA for Ubb
Ywhaz	31981422	139	RefSeq entry for Ywhaz
	29748001 ¹	139	mRNA for Ywhaz
	1304165	139	mRNA for phosholipase A2
Hprt	7305154	139	RefSeq entry for Hprt
	193984 ¹	139	mRNA for Hprt
	38348687	139	Chromosome X genomic sequence
Tubd1	9790052	139	RefSeq entry for Tubd1
	5813775 ¹	139	mRNA for Tubd1
	17221230	139	Chromosome 11 genomic sequence
Tuba1	6755900	139	RefSeq entry for Tuba1
	202207 ¹	139	mRNA for Tuba1
	202222	139	Tuba1 gene, 3' end
	6755143	139	RefSeq entry for Lgals3bp (Ppicap pseudonym)
Ppicap	297032 ¹	139	mRNA for mama
	17017768	139	Chromosome 11 genomic sequence
	26348284	139	Ppicap cDNA
	397799	139	CyCAP mRNA
	42476088	139	RefSeq entry for Hspca
Hspca	12835986 ¹	139	Hspca mRNA
	6680314	139	RefSeq entry for Htf9c
Htf9c	318976 ¹	139	Htf9c mRNA
	3818382	139	Chromosome 16 genomic sequence

regions. Oligonucleotides designed by MPrime also appear to produce highly specific segments as indicated by similarity searches to a number of databases. As more organisms become sequenced and customized gene microarrays become cheaper, the availability of interactive tools such as MPrime to design sequences for custom use will become even more important.

Availability and requirements

Project name: MPrime: multiple primer design

Project home page: <http://kbrin.a-bldg.louisville.edu/Tools/MPrime/>

Operating system: Platform independent (developed using Perl and C++)

Other requirements: MPrime is a freely available to both academic and commercial users as a web-based form.

List of abbreviations used

Actb: Beta actin

BLAST: Basic Local Alignment Search Tool

bp: Base pair

BRG: University of Louisville Bioinformatics Research Group

cDNA: Complementary DNA

DNA: deoxyribonucleic acid

GAPDH: Glyceraldehyde-3-phosphate dehydrogenase

GB: One billion bytes

Hprt: Hypoxanthine phosphoribosyltransferase

Hspca: Heat shock protein 1, alpha

Htf9c: HpaII tiny fragments locus 9C

KB: One thousand nucleotide bases

NCBI: National Center for Biotechnology Information

oligo: oligonucleotide

PCR: Polymerase chain reaction

Ppicap: Peptidylprolyl isomerase C-associated protein

RT-PCR: Reverse transcription polymerase chain reaction

Sdf4: Stromal cell derived factor 4

Tuba1: Tubulin, alpha 1

Tubd1: Tubulin, delta 1

Ubb: Ubiquitin B

Ywhaz: Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide

Authors' contributions

ER coded the solution, and was involved in determining how to design primer pairs and oligonucleotides. In addition, ER was responsible for drafting the manuscript. AK participated in the design of the primer pair and oligonucleotide construction. In addition, he carried out the PCR amplifications and the DNA sequencing. AK was involved in testing and validation of MPrime. NC participated in the design of the primer pair in oligonucleotide construction as well as testing and validation of MPrime.

Acknowledgements

The authors wish to thank members of the University of Louisville Bioinformatics Research Group (BRG) for all of their support and important feedback. Support from NIH:NCRR grant P20 RR16481 (NC) is gratefully acknowledged.

References

- Bernal A, Ear U, Kyrpides N: **Genomes OnLine Database (GOLD): a monitor of genome projects world-wide.** *Nucleic Acids Res* 2001, **29**:126-127.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, Levine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RV, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la BM, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglu S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Har-

- mon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de JP, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
3. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, bu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di FV, Dunn P, Eilbeck K, Evangelista C, Gabrieli AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kechariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooshep S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan A, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M: **The sequence of the human genome.** *Science* 2001, **291**:1304-1351.
 4. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara R, Eyraes E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D, Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Graffham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, Karolchik D, Kasprzyk A, Kawai J, Keibler E, Kells C, Kent WJ, Kirby A, Kolbe DL, Korf I, Kucherlapati RS, Kulbokas EJ, Kulp D, Landers T, Leger JP, Leonard S, Letunic I, Levine R, Li J, Li M, Lloyd C, Lucas S, Ma B, Maglott DR, Mardis ER, Matthews L, Mauceli E, Mayer JH, McCarthy M, McCombie WR, McLaren S, McLay K, McPherson JD, Meldrum J, Meredith B, Mesirov JP, Miller W, Miner TL, Mongin E, Montgomery KT, Morgan M, Mott R, Mullikin JC, Muzny DM, Nash WE, Nelson JO, Nhan MN, Nicol R, Ning Z, Nusbaum C, O'Connor MJ, Okazaki Y, Oliver K, Overton-Larty E, Pachter L, Parra G, Pepin KH, Peterson J, Pevzner P, Plumb R, Pohl CS, Poliakov A, Ponce TC, Ponting CP, Potter S, Quail M, Reymond A, Roe BA, Roskin KM, Rubin EM, Rust AG, Santos R, Sapojnikov V, Schultz B, Schultz J, Schwartz MS, Schwartz S, Scott C, Seaman S, Searle S, Sharpe T, Sheridan A, Shownkeen R, Sims S, Singer JB, Slater G, Smit A, Smith DR, Spencer B, Stabenau A, Stange-Thomann N, Sunget C, Suyama M, Tesler G, Thompson J, Torrents D, Trevaskis E, Tromp J, Ucla C, Ureta-Vidal A, Vinson JP, Von Niederhausern AC, Wade CM, Wall M, Weber RJ, Weiss RB, Wendl MC, West AP, Wetterstrand K, Wheeler R, Whelan S, Wierzbowski J, Willey D, Williams S, Wilson RK, Winter E, Worley KC, Wyman D, Yang S, Yang SP, Zdobnov EM, Zody MC, Lander ES: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**:520-562.
 5. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, Okwuonu G, Hines S, Lewis L, DeRamo C, Delgado O, Dugan-Rocha S, Miner G, Morgan M, Hawes A, Gill R, Celer, Holt RA, Adams MD, Amanatides PG, Baden-Tillson H, Barnstead M, Chin S, Evans CA, Ferriera S, Fosler C, Glodek A, Gu Z, Jennings D, Kraft CL, Nguyen T, Pfannkoch CM, Sitter C, Sutton GG, Venter JC, Woodage T, Smith D, Lee HM, Gustafson E, Cahill P, Kana A, Doucette-Stamm L, Weinstock K, Fectel K, Weiss RB, Dunn DM, Green ED, Blakesley RW, Bouffard GG, de Jong PJ, Osoegawa K, Zhu B, Marra M, Schein J, Bosdet I, Fjell C, Jones S, Krzywinski M, Mathewson C, Siddiqui A, Wye N, McPherson J, Zhao S, Fraser CM, Shetty J, Shatsman S, Geer K, Chen Y, Abramzon S, Nierman WC, Hlavik PH, Chen R, Durbin KJ, Egan A, Ren Y, Song XZ, Li B, Liu Y, Qin X, Cawley S, Worley KC, Cooney AJ, D'Souza LM, Martin K, Wu JQ, Gonzalez-Garay ML, Jackson AR, Kalafus KJ, McLeod MP, Milosavljevic A, Virk D, Volkov A, Wheeler DA, Zhang Z, Bailey JA, Eichler EE, Tuzun E, Birney E, Mongin E, Ureta-Vidal A, Woodward C, Zdobnov E, Bork P, Suyama M, Torrents D, Alexandersson M, Trask BJ, Young JM, Huang H, Wang H, Xing H, Daniels S, Gietzen D, Schmidt J, Stevens K, Vitt U, Wingrove J, Camara F, Mar AM, Abril JF, Guigo R, Smit A, Dubchak I, Rubin EM, Couronne O, Poliakov A, Hubner N, Ganten D, Goesele C, Hummel O, Kreitler T, Lee YA, Monti J, Schulz H, Zimdahl H, Himmelbauer H, Lehrach H, Jacob HJ, Bromberg S, Gullings-Handley J, Jensen-Seaman M, Kwitek AE, Lazar J, Pasko D, Tonellato PJ, Twigger S, Ponting CP, Duarte JM, Rice S, Goodstadt L, Beatson SA, Emes RD, Winter EE, Webber C, Brandt P, Nyakatura G, Adetobi M, Chiaromonte F, Elnitski L, Eswara P, Hardison RC, Hou M, Kolbe D, Makova K, Miller W, Nekrutenko A, Riemer C, Schwartz S, Taylor J, Yang S, Zhang Y, Lindpaintner K, Andrews TD, Caccamo M, Clamp M, Clarke L, Curwen V, Durbin R, Eyraes E, Searle SM, Cooper GM, Batzoglu S, Brudno M, Sidow A, Stone EA, Venter JC, Payseur BA, Bourque G, Lopez-Otin C, Puente XS, Chakrabarti K, Chatterji S, Dewey C, Pachter L, Bray N, Yap VB, Caspi A, Tesler G, Pevzner PA, Haussler D, Roskin KM, Baertsch R, Clawson H, Furey TS, Hinrichs AS, Karolchik D, Kent WJ, Rosenbloom KR, Trumbower H, Weirauch M, Cooper DN, Stenson PD, Ma B, Brent M, Arumugam M, Shteynberg D, Copley RR, Taylor MS, Riethman H, Mudunuri U, Peterson J, Guyer M, Felsenfeld A, Old S, Mockrin S, Collins F: **Genome sequence of the Brown Norway rat yields insights into mammalian evolution.** *Nature* 2004, **428**:493-521.
 6. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Bertram BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de PB, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrieli AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nel-

- son DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC: **The genome sequence of *Drosophila melanogaster***. *Science* 2000, **287**:2185-2195.
7. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S: **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes***. *Science* 2002, **297**:1301-1310.
 8. **UCSC Genome Browser** 2005 [<http://genome.ucsc.edu/>].
 9. Pruitt KD, Tatusova T, Maglott DR: **NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins**. *Nucleic Acids Res* 2005, **33**:D501-D504.
 10. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray**. *Science* 1995, **270**:467-470.
 11. Gerhold D, Rushmore T, Caskey CT: **DNA chips: promising toys have become powerful tools**. *Trends Biochem Sci* 1999, **24**:168-173.
 12. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: **Expression monitoring by hybridization to high-density oligonucleotide arrays**. *Nat Biotechnol* 1996, **14**:1675-1680.
 13. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP: **Light-generated oligonucleotide arrays for rapid DNA sequence analysis**. *Proc Natl Acad Sci U S A* 1994, **91**:5022-5026.
 14. Haas S, Vingron M, Poustka A, Wiemann S: **Primer design for large scale sequencing**. *Nucleic Acids Res* 1998, **26**:3006-3012.
 15. Kampke T, Kieninger M, Mecklenburg M: **Efficient primer design algorithms**. *Bioinformatics* 2001, **17**:214-225.
 16. Li LC, Dahiya R: **MethPrimer: designing primers for methylation PCRs**. *Bioinformatics* 2002, **18**:1427-1431.
 17. Li P, Kupfer KC, Davies CJ, Burbee D, Evans GA, Garner HR: **PRIMO: A primer design program that applies base quality statistics for automated large-scale DNA sequencing**. *Genomics* 1997, **40**:476-485.
 18. McKay SJ, Jones SJ: **AcePrimer: automation of PCR primer design based on gene structure**. *Bioinformatics* 2002, **18**:1538-1539.
 19. Proutski V, Holmes EC: **Primer Master: a new program for the design and analysis of PCR primers**. *Comput Appl Biosci* 1996, **12**:253-255.
 20. Raddatz G, Dehio M, Meyer TF, Dehio C: **PrimeArray: genome-scale primer design for DNA-microarray construction**. *Bioinformatics* 2001, **17**:98-99.
 21. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers**. *Methods Mol Biol* 2000, **132**:365-386.
 22. Fernandes RJ, Skiena SS: **Microarray synthesis through multiple-use PCR primer design**. *Bioinformatics* 2002, **18 Suppl 1**:S128-S135.
 23. Varotto C, Richly E, Salami F, Leister D: **GST-PRIME: a genome-wide primer design software for the generation of gene sequence tags**. *Nucleic Acids Res* 2001, **29**:4373-4377.
 24. Xu D, Li G, Wu L, Zhou J, Xu Y: **PRIMEGENS: robust and efficient design of gene-specific probes for microarray analysis**. *Bioinformatics* 2002, **18**:1432-1437.
 25. Li F, Stormo GD: **Selection of optimal DNA oligos for gene expression arrays**. *Bioinformatics* 2001, **17**:1067-1076.
 26. Nielsen HB, Wernersson R, Knudsen S: **Design of oligonucleotides for microarrays and perspectives for design of multi-transcriptome arrays**. *Nucleic Acids Res* 2003, **31**:3491-3496.
 27. Pozhitkov AE, Tautz D: **An algorithm and program for finding sequence specific oligonucleotide probes for species identification**. *BMC Bioinformatics* 2002, **3**:9.
 28. Raymond N, Charles H, Duret L, Calevro F, Beslon G, Fayard JM: **ROSO: optimizing oligonucleotide probes for microarrays**. *Bioinformatics* 2004, **20**:271-273.
 29. Rouillard JM, Zuker M, Gulari E: **OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach**. *Nucleic Acids Res* 2003, **31**:3057-3062.
 30. Wang X, Seed B: **Selection of oligonucleotide probes for protein coding sequences**. *Bioinformatics* 2003, **19**:796-802.
 31. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank**. *Nucleic Acids Res* 2005, **33**:D34-D38.
 32. Chamberlain JS, Gibbs RA, Ranier JE, Nguyen PN, Caskey CT: **Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification**. *Nucleic Acids Res* 1988, **16**:11141-11156.
 33. Wallace RB, Shaffer J, Murphy RF, Bonner J, Hirose T, Itakura K: **Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch**. *Nucleic Acids Res* 1979, **6**:3543-3557.
 34. Meinkoth J, Wahl G: **Hybridization of nucleic acids immobilized on solid supports**. *Anal Biochem* 1984, **138**:267-284.
 35. Breslauer KJ, Frank R, Blocker H, Marky LA: **Predicting DNA duplex stability from the base sequence**. *Proc Natl Acad Sci U S A* 1986, **83**:3746-3750.
 36. Rychlik W, Spencer WJ, Rhoads RE: **Optimization of the annealing temperature for DNA amplification in vitro**. *Nucleic Acids Res* 1990, **18**:6409-6412.
 37. **WU BLAST** 2005 [<http://blast.wustl.edu/>].
 38. **NCBI BLAST** 2005 [<http://www.ncbi.nlm.nih.gov/BLAST/>].
 39. Rahmann S: **Fast large scale oligonucleotide selection using the longest common factor approach**. *J Bioinform Comput Biol* 2003, **1**:343-361.
 40. Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engle P, McDonagh PD, Loerch PM, Leonardson A, Lum PY, Cavet G, Wu LF, Altschuler SJ, Edwards S, King J, Tsang JS, Schimmack G, Schelter JM, Koch J, Ziman M, Marton MJ, Li B, Cundiff P, Ward T, Castle J, Krolewski M, Meyer MR, Mao M, Burchard J, Kidd MJ, Dai H, Phillips JW, Linsley PS, Stoughton R, Scherer S, Boguski MS: **Experimental annotation of the human genome using microarray technology**. *Nature* 2001, **409**:922-927.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:

http://www.biomedcentral.com/info/publishing_adv.asp

