

A troubleshooting guide:  
Sequencing experts share  
their tips on handling key  
experimental challenges

# DNA Sequencing Tech Guide

&A:

Genome Technology  
THE DNA INTEGRATED BUSINESS



The GenomeWeb  
Intelligence Network

Finally, your lab can run  
high-performance BLAST,  
without the furnace.



The Biocomputing Workstation for Biologists

TimeLogic's CodeQuest™ is an accelerated biocomputing workstation that puts you in control of your genome exploration, 454 data set analysis and oligo mapping—without having to build and administer your own cluster. Avoid ongoing power & cooling expenses and server maintenance headaches.

Simply plug CodeQuest into a standard outlet and network jack. Your entire lab can immediately run the most advanced analyses at blistering speeds. Don't wait for your IT group—answer your toughest bioinformatics questions now!

#### CodeQuest includes:

- A powerful HP workstation with 2-CPU's, 1 terabyte of data storage, a 19" flat-panel monitor, 1 or 2 DeCypher Engine™ Accelerator cards for ultra-fast processing and PipeWorks™, a drag-and-drop interface for building analysis pipelines
- Tera-BLAST™, HMM and Smith-Waterman applications to enable comprehensive comparisons with GenBank, UniProt and Pfam
- Tera-Probe™ for mapping oligos to the genome with high specificity to select optimal microarray probes and locate SNPs
- GeneDetective™ to generate intron/exon models and putative RNA transcripts for alternative splicing studies



CodeQuest enables you to run large analyses faster. In 1 hour, you can map 26,000 30-mers to the rat genome, compare 2,000 protein sequences to Pfam for family assignment, and run 30,000 sequences past GenBank NR. This performance rivals that of hundreds of CPU's, without the hassles and expenses of running a server farm.

In addition to freeing you from high power, cooling, space, administration and maintenance costs, CodeQuest is priced less than a 10-CPU cluster. Contact us today to bring CodeQuest's performance and usability to your lab.

#### North America

Active Motif, Inc.  
Toll Free: 877-222-9543 x4  
timelogic@activemotif.com

#### Europe

Active Motif SA  
Direct: +32 (0)2 653 0001  
eurotech@activemotif.com

#### Asia

Active Motif Japan  
Direct: +81 (0)3 5225 3638  
japantech@activemotif.com

Additional international distributors: [www.timelogic.com/distributors](http://www.timelogic.com/distributors)

TimeLogic®  
biocomputing solutions

A brand of Active Motif® Inc.



# Table of contents

Letter from the Editor ..... 5

Index of Experts ..... 5

**Q1:** What methods do you find most effective for increasing purity and quality during DNA template preparation? ..... 6

**Q2:** How do you minimize contamination and maximize quality during library construction? 8

**Q3:** What do you do to optimize primer specificity and sensitivity? ..... 9

**Q4:** How do you determine efficient reagent dilution for your sequencing reactions? .... 10

**Q5:** What steps do you take to successfully sequence across homopolymeric regions? .. 11

**Q6:** How do you detect and resolve artifacts in your sequence data? ..... 13

List of Resources ..... 14

**Y-BIO** Bioinformatics Package

Y-Bio's single, tightly integrated interface enables genomics researchers to benefit from unified data management and analysis.

Support for 32- and 64-bit x86 and Power Linux. Also available from Penguin Computing as the *ClusterWare HPC - Life Sciences Suite*.

- For x86, Power Linux.
- Secure Login, Data Access.
- Integrated System Management.
- Workstation, Cluster Support.
- Elegant Web Interface

Proteomics  
T-Coffee  
mpBLAST  
M-Bases  
Mastodon  
NCBI BLAST  
EMBOSS  
Glimmer  
Cluster W  
HAMMER  
Wise

[www.terrasoftsolutions.com](http://www.terrasoftsolutions.com)

IBM Business Partner

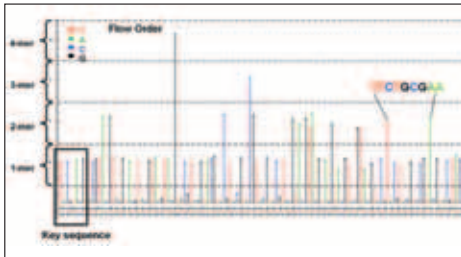
Apple Value Added Reseller



[www.roche-applied-science.com](http://www.roche-applied-science.com)

## Genome Sequencer 20 System

# First to the Finish



Flowgram of a GS 20 read

### Analyze transcriptomes like never before

- Identify miRNA expression
- Perform expression profiling
- Uncover unknown or altered transcripts
- Discover transcription factor binding sites

### with unprecedented speed and sensitivity

Harness the horsepower of the newest revolution in sequencing today — visit [www.genome-sequencing.com](http://www.genome-sequencing.com) call 800 262 4911 (U.S.), or contact your local sales representative.

**454** LIFE SCIENCES



Diagnostics

Roche Diagnostics GmbH  
Roche Applied Science  
68298 Mannheim  
Germany

# Letter from the editor



Welcome to the latest installment in *Genome Technology's* technical reference guide series. In this issue, we are delighted to present lab-tested tips from a set of proven authorities in DNA sequencing.

The art of sequencing has gone through many permutations since the days when Sanger, Maxam, and Gilbert essentially established the field. After PCR was developed, things really heated up when Lee Hood thought to attach dyes to sequencing primers, thereby spurring the march toward fully automated sequencing. It was then just a matter of time before Mike Hunkapiller and Hood built the first instrument to read dye-labeled fragments, the ABI 370.

Four-color sequencing really took off with the

advent of the fly and human genome projects. Admittedly exciting to recall, the genesis story of sequencing may appear clean-cut from today's vantage, especially as the present view is cluttered with distinct chemistries and tools jockeying for position in the race to bring out the top next-generation sequencer. But with each new technological advance, experimental tweaks have been necessary to successfully generate reliable reads.

With this in mind, we decided that this guide should address issues fundamental to sequencing. The following pages feature resources and protocols on topics such as maintaining template purity, choosing primers, and determining reagent dilutions. The experts below will also tell you how to spot artifacts, negotiate homopolymeric regions, and more.

— *Jennifer Crebs*

## Index of experts



**Kevin Knudtson**

Director, DNA Facility  
University of Iowa



**Lee Rowen**

Senior Research Scientist  
Institute for Systems Biology



**Margaret Robertson**

Director, Sequencing and Genotyping  
Ernest Gallo Clinic and Research Center  
University of California, San Francisco



**Glenis Wiebe**

Facility Leader, DNA Sequencing  
Max Planck Institute of Molecular Cell  
Biology & Genetics



**Bruce Roe**

George Lynn Cross Research Professor  
of Chemistry and Biochemistry  
University of Oklahoma



**Alice Young**

Deputy Director, Sequencing Group  
NIH Intramural Sequencing Center



# What methods do you find most effective for increasing purity and quality during DNA template preparation?

Successful DNA sequencing is greatly dependent on the purity and integrity of the DNA. Most commercial kits used to prepare plasmids for sequencing are based on the alkaline lysis procedure of Birnboim and Doly (*Nucleic Acids Res*, 7(6):1513-23) and will yield sequencing-quality templates when applied correctly.

There are a number of keys to success when using this protocol to extract plasmid DNA. Do not overgrow the bacterial culture, as the culture should be collected in the log to late-log phase of growth to avoid degradation of the DNA. Completely remove all the growth media following centrifugation. Some bacteria produce nucleases that will degrade the DNA during the preparation process. Thoroughly suspend the bacterial pellet in the lysis solution; vortex if necessary. Mix gently, but thoroughly, when adding the denaturing and neutralizing solutions to avoid fragmentation of the genomic DNA. The supernatant in the resulting mixture should be crystal clear. If the supernatant is turbid, then the bacterial pellet probably was not completely suspended in the lysis solution and the DNA yield will be low. After centrifugation to pellet the cellular debris, avoid collecting any of the debris. Repeat the centrifugation step again if any debris is collected.

The various commercial kits have their own unique methods to concentrate the DNA while reducing or eliminating the salt concentration. We like to perform at least two wash steps to reduce the salt concentration and thoroughly remove proteins and other bound material. Regardless of template type, the DNA should be suspended in a low-salt buffer or water. In addition, the sample should be free of any traces of ethanol, isopropyl alcohol, and phenol as these compounds will inhibit the DNA sequencing chemistry.

— *Kevin Knudtson*

We don't prepare plasmid DNA, only PCR products from genomic DNA. When we prepare PCR template for sequencing, we set up the PCR reactions in a separate clean room, using either a Beckman robot with filter tips or a Deerac Equator that is a non-contact pipettor. In our lab we consider it critical to keep pre-PCR and post-PCR manipulations in a different room to avoid contamination.

— *Margaret Robertson*

We use an automated procedure for cell lysis and plasmid extraction (specifically, an Eppendorf/ Brinkman robot that uses filter plates for DNA separation). However, non-automated procedures using standard cell lysis and plasmid extraction protocols also work.

— *Lee Rowen*

Template purity is the most important factor in obtaining good-quality sequence data, and as a core laboratory, this is something that we struggle with on a regular basis. We work with many types of templates from a variety of sources, often prepared using different techniques. Salts, organics, RNA, proteins, polysaccharides or chromosomal DNA from plasmid preparation, or primers and unincorporated nucleotides from PCR reactions all interfere with the sequencing reaction. We encourage customers to keep this in mind before submitting samples to us.

For plasmid preparations using commercial miniprep kits, we suggest many of the standard precautions:

1. Avoid overloading the columns, which may lead to decreased yield or plasmid quality. For low-copy plasmids we recommend adding no more than 6 ml of overnight

culture despite some manufacturers' claims that up to 10 ml is acceptable.

2. Prior to eluting the DNA from the column, increase the spin time or speed to ensure complete ethanol removal.
3. Elute in water (preferred) or the provided low ionic strength Tris buffer. Buffers containing EDTA, such as TE, should not be used as this may interfere with the sequencing reactions.
4. Performing two smaller elution steps with pre-warmed elution buffer (70°C) may improve yield, particularly with low-copy or larger plasmids.

For our larger-scale projects, we have had success working with the TempliPhi kit from GE Healthcare, which is reliable and not very labor-intensive. We typically do half reactions (5ul denature buffer and 5ul premix) and skip the final enzyme denaturation step.

For PCR products, we recommend assessing the quality of the DNA template by standard agarose gel electrophoresis. If the PCR product is sufficiently pure (i.e. a single product of the expected size), it is enough to purify the product directly, using either a spin-column or exonucleaseI/shrimp alkaline phosphatase (ExoI/SAP) cleanup method. If multiple PCR products are generated, it is necessary to excise the band of interest from the gel and use an agarose gel purification method.

— *Glenis Wiebe*

Being a high-throughput sequencing lab, our emphasis is on throughput over quality (as long as the sequence quality is high). For instance, the 384-well Agencourt Sprint prep method that we use for

plasmids gives us small quantities of DNA that aren't stable for more than a couple of months, but the templates perform beautifully in BDT sequencing in ABI 3730 DNA sequencers. We are careful to grow cultures to a consistent OD by using precise incubation times and shaking speeds. We examine template aliquots from preps by agarose gel analysis as a post-processing QC to ensure DNA integrity and yield.

In my experience preparing plasmids for small-scale sequencing, growth conditions are critical elements for success. Consistent culture conditions are extremely valuable in obtaining consistent yields and performance. RNA can be hard to see on a gel, so UV quantitation can be misleading (I have seen concentrations exaggerated by as much as 10-fold by UV measurements). We recommend using a fluorometric method employing Hoechst dye 33258. This dye is specific for double-stranded DNA, so RNA will not be detected. For transfection work, purification of plasmids using ion-exchange columns can be critical, but these expensive preps are not required for sequencing. Silica-based methods or old-fashioned manual alkaline-lysis preps are completely adequate for producing top-notch sequencing results.

One key to getting higher yields and lower protein contamination is proper mixing during the denaturation and neutralization steps of alkaline lysis. While protocols emphasize gentle inversion for mixing to minimize gDNA contamination, sharp snapping of the wrist during the mixing steps will produce higher-quality DNA. Contaminating *E. coli* DNA is rarely detectable by gel analysis and never reaches levels that affect DNA sequencing using plasmid-specific primers.

— *Alice Young*



# How do you minimize contamination and maximize quality during library construction?

There are a number of critical steps that can affect the ability to construct a quality library. Construction of a high-quality cDNA library starts with obtaining good-quality RNA. The use of degraded or partially degraded RNA will give a library that does not fully represent the genome. Following cDNA synthesis, careful fractioning of the resulting cDNA should be performed to eliminate very small products. Failure to do so will lead to a library that only contains small inserts that will necessitate the collection of considerably more independent recombinants to have complete coverage of the genome.

Also, the inclusion of a library-specific sequence tag in the poly(T) reverse primer during cDNA synthesis will help detect cross-library contamination if more than one library is to be constructed. Vector DNA must be completely digested and phosphatased because uncut vector will transform the bacterial host very efficiently, resulting in a library that contains few or no inserts.

Library quality can be assessed by DNA sequencing to look for the presence of an insert, the poly(A) tail, a hexamer known to be a signal for cleavage and polyadenylation, and the unique library-specific sequence tag. The absence of inserts suggests the vector was not cut completely or there was a poor insert-to-vector ratio for the ligation step. The absence of poly(A) tails and polyadenylation signals suggests RNA degradation. The presence of the wrong sequence tags indicates contamination from a previous library preparation.

— Kevin Knudtson

“Presence of the wrong sequence tags indicates contamination from a previous library preparation.”

— Kevin Knudtson

Take extreme care to not shear any genomic DNA during BAC isolation by minimizing agitation at the various mixing steps (rolling the centrifuge bottles rather than shaking).

Employ a second acetate precipitation step during the BAC isolation — see our website at [http://www.genome.ou.edu/BAC\\_isoln\\_200ml\\_culture.html](http://www.genome.ou.edu/BAC_isoln_200ml_culture.html).

— Bruce Roe

a) Pay attention to bacterial growth conditions, optimizing media and duration.

b) We use a Kurabo AutoGen 740 for BAC or fosmid DNA isolation, again optimizing conditions (e.g. lysis, centrifugation) to reduce *E. coli* contamination.

— Lee Rowen

Library construction is not something that we do in house, but rather outsource to commercial vendors. I can, however, stress the importance of carrying out the recommended quality checks throughout the entire procedure to ensure that the final product is of the highest quality. For cDNA libraries, we verify the quality of the starting RNA and determine the final cloning efficiency and average insert size (typically greater than 95 percent and 1.5 kb, respectively). Within the scope of our EST projects, we also assess library quality and diversity based on the number of redundant clones in the library, by performing BlastN searches against all sequenced ESTs.

— Glenis Wiebe

(continued on p.14)





# What do you do to optimize primer specificity and sensitivity?

The users of our core have had good success using Primer3 and the commercially available primer design programs to design their primers. We encourage our users to design their primers to have a  $T_m$  of at least 45°C, but primers with a  $T_m$  of 55°C to 60°C tend to give successful sequencing results on a more consistent basis.

Primers should not contain stretches of any one particular base. Repetitive G's or C's should especially be avoided. Primers should be "stickier" on their 5' ends than on their 3' ends. A "sticky" 3' end as indicated by a high GC content could potentially anneal at multiple sites on the template DNA. However, we encourage the use of a G/C clamp, especially when sequencing through AT-rich regions.

Primers should not possess complementary sequences (palindromes), as this will result in a primer that will fold back on itself, leading to an unproductive priming event that decreases the overall signal. Primers should not contain sequences of nucleotides that would allow one primer to anneal to another (primer dimer formation). If possible, run a computer search against the vector and insert DNA sequences to verify that the primer and especially the 8-10 bases of its 3' end are unique.

— *Kevin Knudtson*

We use Primer3 to select PCR primers. Primer concentration is kept low at 2uM, as are dNTPs and Taq. We try to keep Taq, primers, dNTPs, and DNA at

low concentrations without compromising efficiency. We have a routine optimization procedure where we perform touchdown PCR (the first 10 cycles vary by one degree of the annealing temperature from 60-50°), then 25 cycles at 50°. Any PCR dropouts go on to a "Betaine" optimization that usually recovers about 60 percent to 70 percent of the dropouts. After that we have to work very hard to define conditions for the remaining assays. This can take a lot of hands-on time. It is probably more efficient to redesign primers, which we usually do if an assay continues to fail.

— *Margaret Robertson*

"Primer design programs are a must to ensure that the primer will anneal at a unique site in your clone."

— *Alice Young*

Use PrimOU, a primer picking program based on MIT's Primer and UT Southwestern's Primo. Freely available from our website at <http://www.genome.ou.edu/informatics/primou.html>.

— *Bruce Roe*

For sequencing PCR products, we use the same primer for both PCR and sequencing after purifying the PCR products. For plasmids, we use either universal or custom primers, following ABI protocols for concentration. Generally, our primers are between 18 and 20 bases. It's been our experience that choosing a primer without going through the rigmarole of an oligo design program usually works just fine. We choose a primer site about 100 bases upstream of the region we want sequence for, so as to avoid errors that occur at the beginning of a sequence read.

— *Lee Rowen* (continued on p.14)



# How do you determine efficient reagent dilution for your sequencing reactions?

We took an empirical approach to determine the best reagent dilution to use in our sequencing reactions for our core users. Essentially we processed two to three users' samples using different dilutions of BigDye and we are currently using the highest dilution that continued to give successful sequencing results.

— *Kevin Knudtson*

Since we are re-sequencing for rare mutations it is imperative that we don't dilute the BigDye v3.1 reagent too much. Phred scores have to be higher than 30 to pass our quality metrics in our sequencing pipeline, so good quality data is a must.

I test out dilutions of reagents using several exons of different lengths (400 to 600 bases) that contain known polymorphisms of different sequence combinations (e.g. C/T, or G/C) and empirically determine the performance of the various dilutions by inspecting the traces for quality, signal strength, and accuracy.

— *Margaret Robertson*

Since we're the folks that originally published reagent dilution on our website, we do the dilutions and then run reactions with both pUC and a set of pUC-based shotgun clones at each BigDye or ET mix dilution, with sequencing on either the ABI 3700 or 3730. See our website at <http://www.genome.ou.edu/PredispensedDilutedSeqMixBD3orET.html>.

— *Bruce Roe*

We run a 96 plate with, say, four batches of 24 templates of the sort we typically sequence, varying

the reagent dilution. After the sequencing run, we analyze the data for read length (alignment to a consensus sequence) and Phred quality scores.

— *Lee Rowen*

In a core facility, the templates we receive are extremely varied in type, quality, and concentration. We continuously monitor failure/repeat reaction rates to help determine what types of problems are occurring, and how they are best dealt with. In our experience, using 1/8th dilution (1  $\mu$ l of BigDye Terminator v3.1 in 10  $\mu$ l final volume) in the reaction provides a good balance between minimizing sample failure and keeping costs down. With larger-scale projects where we have control over template preparation, we have been able to use as little as 1/80th (0.1  $\mu$ l) of the sequencing reaction kit.

— *Glenis Wiebe*

For ABI 3730 DNA sequencers this was accomplished by setting up reactions at various dilutions and volumes. A robust reaction can be achieved using a 6  $\mu$ l reaction containing 1/3  $\mu$ l of the BDT version 3.1 sequencing reaction mix and supplemental 5x buffer. We use very small amounts of DNA (20 ng to 50 ng for plasmids 5 kb to 7 kb). Lower dilutions can be used, but are dependent on the quality and concentration of your DNA. For ABI 3130 DNA sequencers we use 10  $\mu$ l reactions containing 4  $\mu$ l BDT version 3.1. These conditions are probably overkill, but we are generally sequencing very difficult templates on these instruments.

— *Alice Young*



# What steps do you take to successfully sequence across homopolymeric regions?

Homopolymeric regions, Alu repeats, direct and inverted repeats, di- and trinucleotide repeats, and strong hairpins are among the template features that can be difficult to successfully sequence through. Sometimes difficult templates need to be addressed on a case-by-case basis depending on the base content of the region.

One of the first steps we take is to design a new sequencing primer that binds approximately 100 bases upstream of the difficult region. Also, we will include a heat denaturation step of 98°C for five minutes prior to the cycle sequencing reaction. In many cases, the heat denaturation step and moving the primer closer will permit the DNA polymerase to extend through the difficult region without changing the cycle sequencing chemistry. We have also had limited success in sequencing through GC-rich difficult regions by using a cocktail of three parts BigDye terminator v3.1 to one part dGTP BigDye terminator v3.0. A more comprehensive discussion on sequencing through difficult regions can be found in chapter three of *DNA Sequencing: Optimizing the Process and Analysis*, edited by Jan Kieleczawa.

— *Kevin Knudtson*

Since we are sequencing PCR products, it is often not possible to sequence through a long homopolymer because of enzyme slippage. In the primer design stage we try to avoid homopolymers greater than eight. In regular core sequencing of plasmids or BACs, we can use a polyT primer with an anchored base or a mixed 3' base to try to get through the region.

— *Margaret Robertson*

Follow the protocols that we have on our website:

- [http://www.genome.ou.edu/protocol\\_enhancing\\_PCR.html](http://www.genome.ou.edu/protocol_enhancing_PCR.html)
- [http://www.genome.ou.edu/seq\\_very\\_difficult\\_regions.html](http://www.genome.ou.edu/seq_very_difficult_regions.html)
- [http://www.genome.ou.edu/phi29\\_protocol.html](http://www.genome.ou.edu/phi29_protocol.html)

These are especially useful for GC-rich regions. If they are AT-rich regions, we replace the BigDyes with the d-Rhodamine mixes. Also we PCR the homopolymer stretch to determine the exact number of bases present in the repeats in conjunction with the above methods.

— *Bruce Roe*

We try a different chemistry, such as BigDye Primers (now hard to get) or dGTP reagent kits.

— *Lee Rowen*

We use a 3:1 mixture of BigDye v3.1 and dGTP BigDye v3.0 in most reactions. For a problematic long homopolymeric region, the first step is to sequence the opposite strand. If this doesn't provide the necessary information, we try a combination of approaches such as lowering the extension temperature from 60°C to 55°C, using an oligo dT primer with degenerate bases at the 3' end, or designing a new primer closer to the homopolymeric region.

— *Glenis Wiebe*

We avoid these like the plague. Our medical resequencing amplimers are carefully designed to avoid these regions in both the primers and the internal sequence of the PCR products. Good luck!

— *Alice Young*

# Next generation sequencing services

At the 454 Sequencing™ Center, we can sequence *anything*—and we can do it *now*—in record time, at a cost never before possible.

Our next generation, ultra-high throughput DNA sequencing platform is simple, incredibly fast, and a tenth of the cost of traditional methods.

Learn what's truly possible. Access 454 Sequencing technology and our expertise through the professional services of the 454 Sequencing Center. Call us today (877-890-GNOM) or contact us online at [www.454.com/services](http://www.454.com/services).

## 454 Sequencing Center

Access the power of revolutionary 454 Sequencing technology today



Forty times faster. **40x**

One tenth the cost. **1/10**

99% accurate\*. **99%**

BACTERIA • VIRUSES • FUNGI • LARGE EUKARYOTIC GENOMES •  
PCR AMPLICONS • AMPLIFIED TAG SEQUENCING •  
ENVIRONMENTAL SEQUENCING • POPULATION BIOLOGY •  
cDNA • BAC CLONES OR POOLS

\* Single-read accuracy. Consensus accuracy in excess of 99.99% for non-repeat regions.

454 Sequencing Center  
16 Commercial Street, Branford, CT 06405  
Tel: (877) 890-GNOM or (203) 871-2300  
Fax: (203) 315-2670  
E-mail: [info@454.com](mailto:info@454.com)

Visit [www.454.com](http://www.454.com)

**454** LIFE SCIENCES



# How do you detect and resolve artifacts in your sequence data?

Initially we examine the plate report generated by our LIMS software, Geospiza Finch Suite, for potential issues in the sequencing results. We also import the plate record into Applied Biosystems' Sequence Scanner v1.0 program to perform individual assessments and examination of the sequencing quality of each sample. Once an artifact has been identified, the daunting task can be trying to figure out its source. Is the artifact due to the instrument, chemistry, polymer, capillary, primer, template, or something else? The DNA Sequencing Research Group of the Association of Biomolecular Resource Facilities has developed a tool [known as the] DSRG Sequencing Troubleshooting Web Resource that investigators can use to help troubleshoot their sequencing problems.

— *Kevin Knudtson*

We rely a lot on automated data analysis, but have LIMS tools that will flag data that is poor quality or that needs to be inspected by a human.

— *Margaret Robertson*

It's best not to generate any artifacts by following established protocols. For example, the ABI BigDye Version 3.1 includes some additives that virtually eliminate many of these artifacts. We also sequence each base at least three times following Sanger's "Rule of Three" with at least one of the times being from the opposite strand.

— *Bruce Roe*

Systematic errors are best detected by comparing the sequences of thousands of templates using different platforms or chemistries (e.g. ABI vs. MegaBace vs.

454; BigDye terminator vs. BigDye primer chemistry). If only one sequencing platform/chemistry is used, then systematic error may not be detected. Random errors can be detected using redundancy (e.g. shotgun sequencing) or by comparison to a consensus sequence. In many cases (e.g. compressions) sequences off of the forward and reverse strand may differ, and a visual examination of the data in a trace editor can help an experienced person make the correct call.

— *Lee Rowen*

Because we are a lower-throughput operation, we are able to visually inspect every chromatogram. This allows us to quickly note any sequencing artifacts that arise, and convey them to the customer. In some cases we are able to edit the base-calling, and no further intervention is required. Otherwise, we may re-run the sample and repeat the sequencing reaction as necessary.

— *Glenis Wiebe*

Statistics are generated for each sequencing run. Each day a summary report is e-mailed to key personnel. Array views are routinely examined on the sequencers, especially when there appears to be anything abnormal showing up in our statistics. We scrutinize these displays for signal intensity, resolution of peaks, and background. Our finishing group is responsible for bringing the assembled sequence quality up to standards. In doing so, they examine many traces. They once detected a cross-contamination we were getting from a robot that was not properly washing tips between samples.

— *Alice Young*

## Q2: How do you minimize contamination and maximize quality during library construction?

(continued from p.6)

We make two types of libraries - small-insert (3-5 kb) shotgun libraries from BACs or fosmid libraries from BACs. In both cases our BACs come to use as glycerol stocks. These stocks were fingerprinted in our mapping lab where the fingerprint is one of several parameters used in selecting which BACs to sequence. When we reprep the BACs for library construction we re-fingerprint the DNA and check for a match. During library construction, BACs are handled in groups of 16 so swaps are possible. When the sequence data comes through, the virtual fingerprint is compared to the real fingerprint. To minimize *E. coli* contamination we use a BAC prep method that came out of Wash U. During the denaturation and neutralization steps of alkaline lysis there is no mixing that would cause shearing of the genomic *E. coli* DNA.

— *Alice Young*

## List of resources

Our panel of experts referred to a number of Web resources, which can be found below.

## Websites

### Primer3:

[http://frodo.wi.mit.edu/cgi-bin/primer3/primer3\\_www.cgi](http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi)

### DSRG Sequencing Troubleshooting Web Resource:

<http://www.abrf.org/index.cfm/stwr.home>

### Mutation Surveyor from Softgenetics:

<http://www.softgenetics.com/mutationSurveyor.htm>

### Vector NTI Suite (Invitrogen):

<http://www.invitrogen.com/content.cfm?pageid=10130>

### Sequencher (Gene Codes Corp.):

<http://www.genecodes.com/sequencher>

## Q3: What do you do to optimize primer specificity and sensitivity?

(continued from p.9)

We tend to follow the standard rules: the length should be 18 to 24 nucleotides, with a G/C content of approximately 40 percent to 60 percent, and a melting temperature between 48° and 72°C. Primers with long runs of a single base should be avoided, especially runs of three or more G's or C's. Primers should not form secondary structures or primer dimers; the presence of 3' hairpins or 3' complementarity is particularly detrimental. It has been suggested that the 3' end of a primer should have a G or C as final base to act as a "clamp" when the primer anneals to the template DNA; however, we have not found this to be very important. In cases where the template is known, the primer should be checked for second-site hybridization. Finally, degenerate primers generally do not work well for sequencing.

Regarding primer purity, as long as the coupling efficiency of the synthesis was high, post-synthesis desalting is usually sufficient. However, since primers of poor quality do not work well for sequencing, we recommend that our customers order HPLC or PAGE purified primers.

— *Glenis Wiebe*

Primer design programs are a must to ensure that the primer will anneal at a unique site in your clone. When primers cannot be designed using the standard parameters, first try lengthening the primer by a base or two, then allow slightly long homopolymer runs. I really like to have a G or C at the 3' or proximal 3' end of the primer. More than two Gs or Cs at the 3' end can lead to false priming.

— *Alice Young*

## Acknowledgments

Many thanks to Scott Bloom of the Institute for Systems Biology for advising on the answers submitted by Lee Rowen.



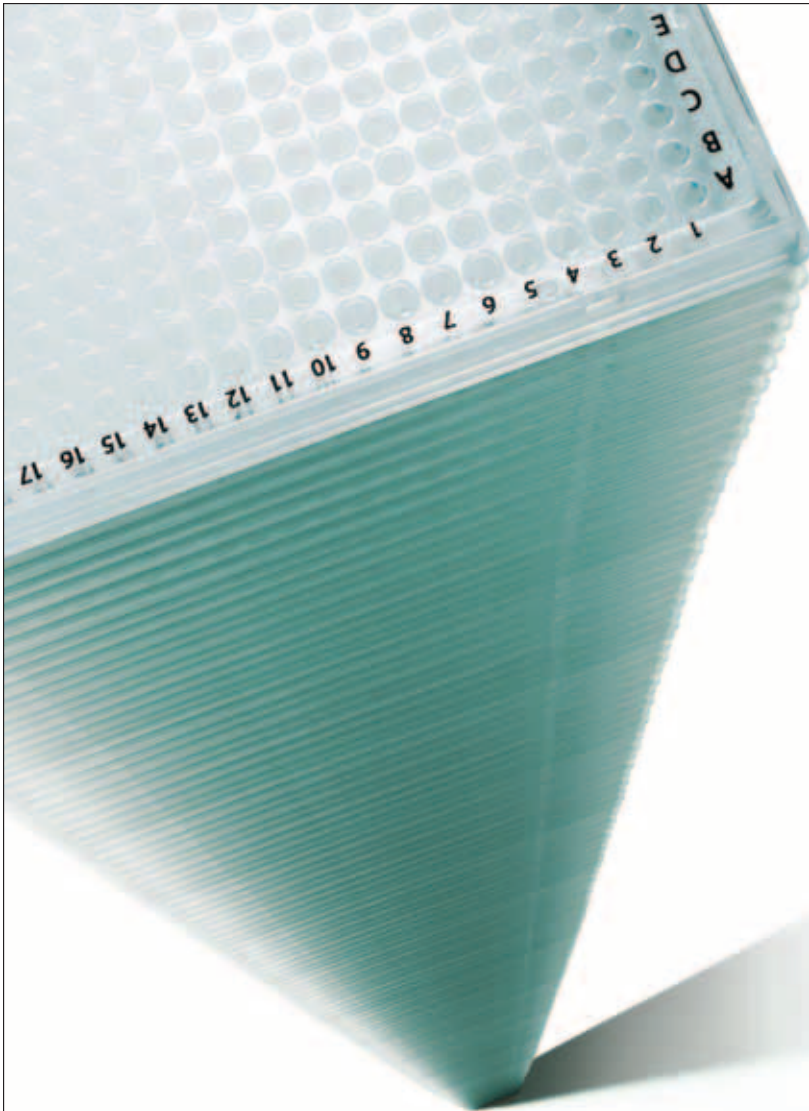
Polymorphism Resequencing • Whole Genome Methylation Analysis •  
DNA Methylation Sequencing • Microarray SNP Genotyping •  
Expression Analysis • Custom Gene Silencing Constructs



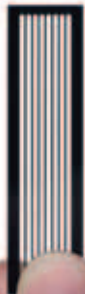
Phone: 800-720-4363 • Email: sales@seqwright.com  
2575 West Bellfort, Houston, TX 77054-5025

[www.seqwright.com](http://www.seqwright.com)





To generate a billion bases, you'd go through thousands of plates.



Fortunately, you don't have to any more! Based on our novel Clonal Single Molecule Array™ technology and proprietary reversible terminator chemistry, the Solexa Genome Analysis System is designed to generate more than one billion bases of data in a single run—at less than 1% the cost of capillary methods. Say goodbye to the piles of plates! To learn more about the Solexa Genome Analysis System, visit [www.solexa.com](http://www.solexa.com).

© 2006 Solexa, Inc. All rights reserved.

