

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

SNP Genotyping Tech Guide

&A:

Genome Technology
THE GENOME TECHNOLOGY CENTER

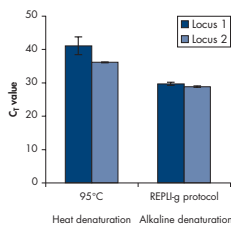


**The GenomeWeb
Intelligence Network**

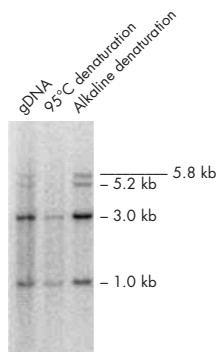
High-quality and unlimited genotyping from precious DNA samples

Genetic analyses often require large amounts of genomic DNA. Since the availability of DNA from individual genomes is often limited, accurate replication of large amounts of high-quality genomic DNA is required. QIAGEN's REPLI-g® Kits and Service use multiple displacement amplification (MDA) to overcome the limitations of small amounts of template DNA.

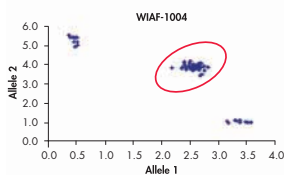
Effect of Heat and Alkaline Denaturation on Loci Representation



High-Molecular-Weight DNA



Reliable SNP genotyping



In contrast to traditional PCR-based whole genome amplification, QIAGEN's REPLI-g Kits and Service provide highly uniform DNA amplification across the entire genome, with minimal sequence bias. Reproducible, high yields of DNA can be obtained from various clinical and non-clinical research samples, including genomic DNA, fresh or dried blood, buffy coat, and tissue culture cells.

A number of kit formats are available to suit all DNA yield and throughput requirements. Typical yields per 50 µl reaction are up to 8 µg (Screening Kit), 10 µg (Mini Kit) and 40 µg (Midi Kit). A reproducible yield of amplified DNA is usually achieved regardless of the quantity of template DNA, enabling immediate downstream genetic analysis, including SNP genotyping, STR analysis, Southern blotting and RFLP analysis, subcloning, and DNA sequencing. The average product length is typically greater than 10 kb, with a range between 2 kb and 100 kb.

DNA denaturation is a critical step in the amplification of whole genomic DNA. Excessive heat can damage the DNA resulting in incomplete and less consistent locus representation. In contrast, the alkaline denaturation (Mini and Midi Kit) or gentle thermal denaturation (Screening Kit) provided by QIAGEN's REPLI-g Kits results in uniform DNA amplification.

The REPLI-g Screening Kit combines a number of features making it ideal for use in manual or automated high-throughput mutation screening, including high volume pipetting steps for increased accuracy, compatibility with 96-well microplates, and a streamlined protocol for fast and easy set up.

The REPLI-g Service provides amplification of unlimited amounts of DNA from limited samples. A stringent quality control assay provides information on the quality of the amplified DNA, enabling you to make reliable predictions for the success of your downstream assay.

Visit www.qiagen.com/goto/REPLI-g to find out more about unlimited genotyping from precious samples!



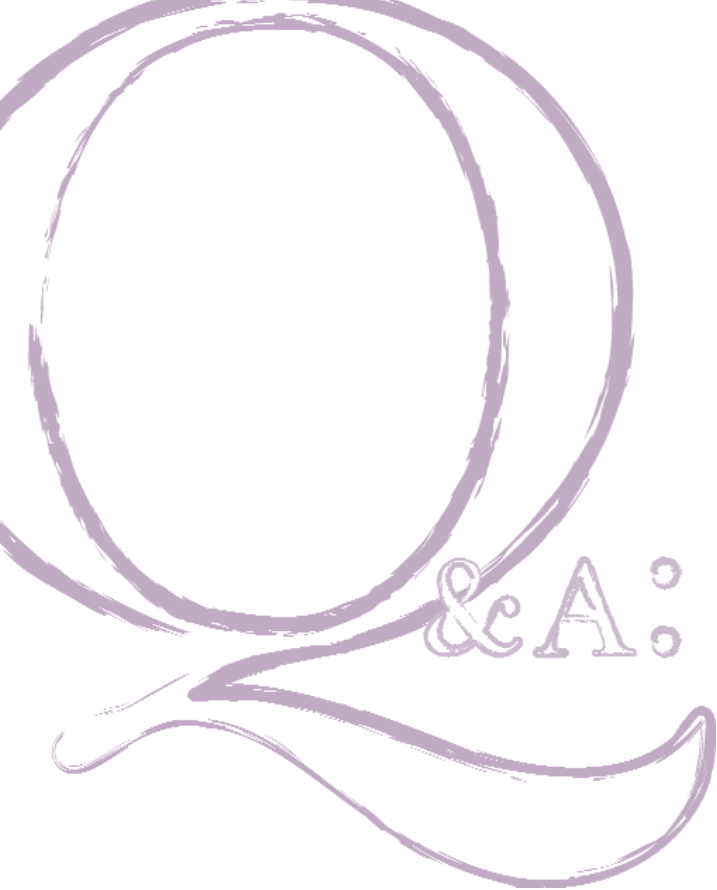


Table of contents

Letter from the Editor	4
Index of Experts	4
Q1: What do you consider when choosing platforms?	6
Q2: What criteria do you use for selecting SNPs or potentially functional sites?	8
Q3: Which factors are most important to achieving high genotyping accuracy?	11
Q4: What quality control measures do you employ?	13
Q5: What methods or tools do you use for the analysis and visualization of data?	16
List of Resources	18

Enabling Better Decisions



Quality & Experience:

- First GLP Compliant Microarray Laboratory
- CLIA-registered for clinical trials
- Ph.D. Staff with over 30 years experience in microarrays

Comprehensive Services:

- RNA Expression: Genome & Exon
- Genotyping: High Density & Focused
- Isolation: Expertise in RNA, DNA and Blood Protocols

Innovation:

- Experienced in FDA submissions
- Leader in standardization efforts
- Novel analysis tools



Licensed Affymetrix Service Provider



Comprehensive Microarray Services

www.expressionanalysis.com

Toll-free 1.866.293.6094

Letter from the editor



We are delighted to present the latest installment in *Genome Technology's* technical reference series. In this issue, experts share their tips, experiences, and insights on SNP genotyping.

Once upon a time, the only means to identify particular base changes in the human genome was through the use of good old PCR. But, as ambitious genotypers found out, PCR can only be multiplexed so much before dreaded primer-dimer artifacts threaten to crowd out intended amplicons.

Luckily, times have changed and there is now a relative panoply of genotyping technologies to choose from. Commercially available systems span the gamut

of those that are microarray-based to those that rely on allele-specific hybridization of oligo probes. PCR still has a role, especially restriction-enzyme based and long-range approaches.

With this diversity of technologies, however, comes a host of questions about optimal use. Although new tools have revolutionized the way in which studies are realized — witness the HapMap project or any number of whole-genome association studies — best practices in the genotyping world are still evolving.

We've thus asked the experts to share their advice concerning platform selection, quality control standards, and more. With special thanks to the contributors below, we hope you enjoy the following pages.

— *Jennifer Crebs*

Index of experts



Hsueh-Wei Chang

Faculty of Biomedical Science and
Environmental Biology
Kaohsiung Medical University, Taiwan



David Duggan

Genotyping Technology Center
Translational Genomics
Research Institute



Darryl Irwin

Genotyping Business Unit
Australian Genome Research Facility



Stuart Macdonald

Department of Ecology and
Evolutionary Biology
University of California, Irvine



Heather McKhann

Étude du polymorphisme
des Génomes Végétaux
INRA, France



Louise Nordfors

Department of Molecular
Medicine and Surgery
Karolinska Institute



Jose Luis Royo

Department of Structural Genomics
neoCodex SL



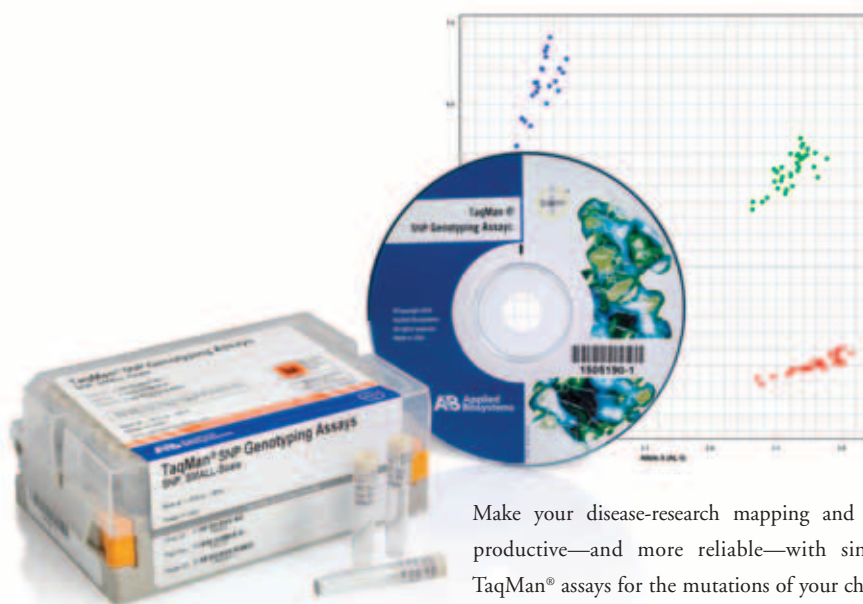
Huanming Yang

Beijing Genomics Institute
Chinese Academy of Sciences

Announcing TaqMan® Drug Metabolism Genotyping Assays for basic and clinical research. For more information, please visit dme.appliedbiosystems.com

Choose your genome. Choose your SNP.

Ready-to-use TaqMan® SNP Genotyping Assays.




Make your disease-research mapping and screening efforts more productive—and more reliable—with single-tube, ready-to-use TaqMan® assays for the mutations of your choice. For unprecedented genomic coverage, select from our library of human mapping and coding SNP assays or have custom assays built for any possible SNP in any genome. A simple, scalable workflow runs on Applied Biosystems real-time PCR systems and gives you the accurate and reproducible results you're looking for. Your SNP assays are ready when you are at: <http://www.allsnps.com>

AB Applied Biosystems



For Research Use Only. Not for use in diagnostic procedures. The PCR process and the 5' nuclease process are covered by patents owned by Roche Molecular Systems, Inc., and F. Hoffman-La Roche Ltd. Applied Biosystems is a registered trademark and AB (Design) is a trademark of Applied Biosystems Corporation or its subsidiaries in the US and/or certain other countries. TaqMan is a registered trademark of Roche Molecular Systems, Inc. © 2006 Applied Biosystems. All rights reserved.



What do you consider when choosing platforms?

In general, we perform SNP genotyping by restriction fragment length polymorphism (RFLP), sequencing, PCR-CTPP. In our published freeware, SNP-RFLPing, the RFLP available restriction enzymes for input of SNP ID (rs#, ss#), gene name, gene ID, and any formats of sequences are provided. Users can get the RFLP information for SNP genotyping online.

— *Hsueh-Wei Chang*

Some of the things we consider include throughput and performance as measured by reliability, reproducibility, and accuracy. We also consider cost and things such as ease-of-use.

In addition to that, we may not have the right technology, and so we may ask, “Is there a better technology available elsewhere?” — whether it be from a collaborator or even a fee-for-use service.

[The last] in our consideration of approaches is [asking whether] there are there new, more efficient technologies on the horizon. Efficient is defined by any one or more [aspects such as] throughput, performance, cost, and ease-of-use. So a lot goes into considering or choosing approaches to these types of studies.

In my lab, we actually have a number of SNP genotyping technologies, including technology from Sequenom, Illumina, and Affymetrix. We also have, although we're not using them all that much, Applied Biosystems' TaqMan and SNPlex. At times we still find ourselves using PCR followed by restriction enzyme digestion.

Being a genotyping and genomics research institute, we're closely watching the latest developments in resequencing technology. Resequencing is, after all, the ultimate genotyping

technology. It's just cost prohibitive right now.

— *David Duggan*

The Australian Genome Research Facility is a major national research facility, which has been set up by the federal government to be a provider of genomic services for the Australasian region. This basically means that we are a service provider and as such do not perform our own research studies. Rather, our clients request us to run specific SNPs on their samples and provide them with the data. When choosing platforms for SNP genotyping, we're actually caught in a 'tug-of-war' between two key factors: the quality and reliability of a specific SNP genotyping platform versus the cost. Our clients are largely cost motivated, but at the same time, we need to ensure that the data that's generated is accurate. Our primary motivator is the accuracy and reliability of the platform, as well as its throughput (because we are a high-throughput laboratory), followed very closely by the cost.

— *Darryl Irwin*

SNP genotyping platforms each have associated costs and benefits, and will vary from project to project and from lab to lab.

Arguably, the most important consideration is the scale of the project: How many SNPs are to be assayed, and how many individuals are to be tested? Say you want to test 10 SNPs in 10,000 DNA samples. Highly multiplexed technologies are not required, but one would probably like to avoid running 10,000 lanes on a sequencer. An appropriate alternative is FP-SBE, or fluorescence polarization detection of single-base extension products (Chen et al. 1999). This



method has the advantage that the detection happens right in the microtiter plate (no electrophoresis is required), the assays are fairly robust, and the initial cost of the oligonucleotide probe is low. If you want to genotype 10,000 or more SNPs, then the ultra-high-throughput platforms offered by Illumina (Shen et al. 2005) and Affymetrix (Matsuzaki et al. 2004) are ideal.

The real difficult decisions arise when you want to genotype 100 to 200 SNPs in 1,000 to 2,000 individuals, for example in a high-power candidate gene-based association mapping experiment. There is no obvious commercially available technology that is fully compatible with a 200 SNP, 2,000 sample project: 200 SNPs is not sufficient to make use of the power of the ultra-high-throughput systems, and you need access to a large number of PCR machines to make singleplex assays workable. This is exactly the problem we faced in our lab, and so we developed an open-source genotyping system based on the OLA (Oligonucleotide Ligation Assay, Landegren et al. 1988), using 16-plex genotyping reactions, and array-based genotype detection (Macdonald et al. 2005). This permitted us to collect genotypes on large panels of 2,000 individuals, and to assay 200 SNPs cheaply, efficiently, and with high accuracy.

— *Stuart Macdonald*

At the French National Genotyping Center, a number of platforms are already in place, including mass spectrometry, TaqMan, Amplifluor, SNPlex, and Illumina. The first consideration to find the appropriate platform for a given project is the number of samples and the number of SNPs to be

genotyped. This allows us to choose among the different technologies. Then among the appropriate platforms, the accuracy of the genotyping method is the second consideration.

For a large number of SNPs in a single gene, with a small number of samples, sequencing is very efficient. For a small number of SNPs on a relatively large number of samples, we have found TaqMan to be the most accurate method among the three we have tested (mass spectrometry using the GOOD assay, Amplifluor, and TaqMan; Giancola et al., 2006). For a large number of SNPs and a large number of samples, we have chosen SNPlex and have obtained very good results. Starting with 384 SNPs, Illumina is the system chosen.

— *Heather McKhann*

Accuracy, cost, user-friendliness, and possibility of technical support from the company providing the instruments.

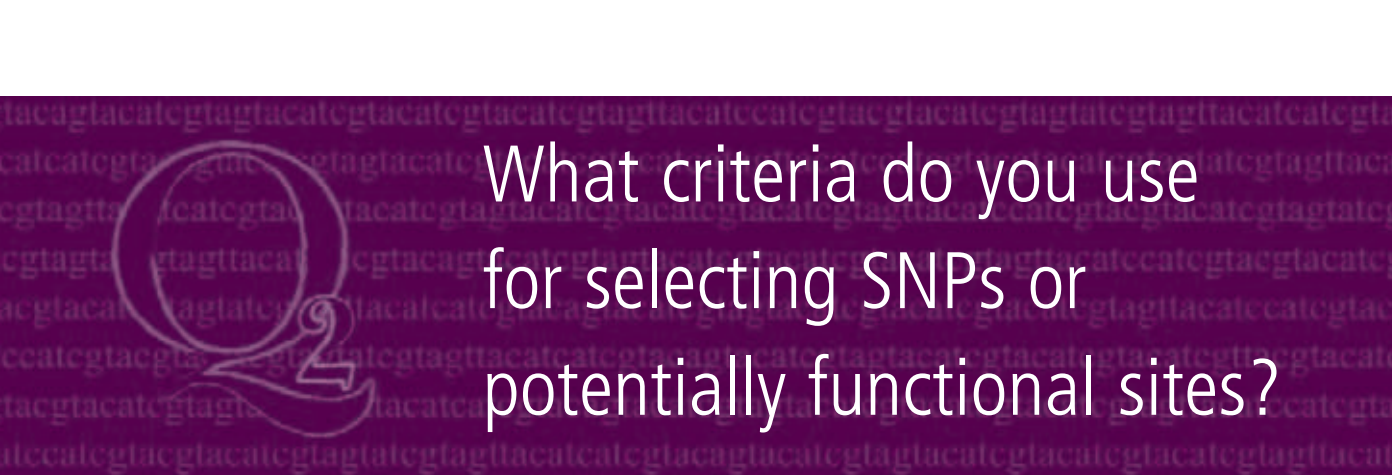
— *Louise Nordfors*

I choose the platform according to the requirements: (i) many SNPs in a small subset of patients or (ii) few SNPs in thousands of patients. [It also depends on] the budget and the price-per-genotype calculated by the technical director.

— *Jose Luis Royo*

Accuracy always first, then cost and flexibility, and others. Although all are important, the accuracy is the most important for general considerations; cost and flexibility are important for large-scale genotyping projects.

— *Huanming Yang*



What criteria do you use for selecting SNPs or potentially functional sites?

Coding nonsynonymous, 5' locus (promoter), splicing site with suitable heterozygosity frequency.

Of course, the SNP frequency difference for different ethnicities is also considered. In NCBI dbSNP, they provide detailed information for almost every SNP now. Users can select the limitation in dbSNP to find your SNPs [of interest]. Usually, the inputted gene name should be commonly used in HUGO and is researchable in Entrez Gene, [in which] chromosome location is suggested to remove noise. Some related SNPs which do not belong to the input target gene will [be] provided if the limitation is not selected.

— *Hsueh-Wei Chang*

To begin with, the criteria depends on the study's question or hypothesis. For example, if we're working on a linkage analysis project, we may be interested in fine mapping a positional candidate gene region, and our criteria would depend on the size of that region. If we had 10 megabases of DNA to fine map, we may not necessarily jump to SNP genotyping. We may put in some microsatellite markers into that region prior to doing any SNP genotyping. But if the region is sufficiently small and/or absent of any microsatellites, we would most definitely use SNP genotyping. In that case, and even in the larger case, we would specify some criteria depending on the size, like maybe one SNP every 100 kb. If it's a big region or we're doing even finer mapping, one SNP every 10 kb. We may even put some minor allele frequency parameters on that selection as well.

We typically turn to three areas when selecting functional SNPs. The first would be the literature. If

there's a region of the genome we're interested in, or a gene we're interested in, [we ask whether there] is anything in the literature — PubMed or otherwise — that would highlight some functional SNPs we should look at.

We would also consider expert curated SNPs. Maybe there are SNPs — some people call them your favorite SNPs — that don't yet reside in a database for one reason or another. They could also be included.

Finally, when we're actually selecting our functional SNPs or putative functional SNPs, we turn to many of the public databases like NCBI or UCSC's genome browser. They've become rather rich sources of genetic variation, especially SNP information, including functional SNPs or putative functional SNPs such as non-synonymous SNPs. We'll look to those two sources, as well as others, for non-synonymous SNPs.

— *David Duggan*

Our clients provide the identity of the SNPs to be assayed to us; we don't actually determine which SNPs we'll run because we are a contract service provider. We thoroughly check the supplied SNP sequences both before and after designing the assays to ensure the reference sequence is correct and the designed assays have a high probability of success. To do so, we largely use Sequenom's RealSNP database and software, which actually compares the supplied sequence back to the publicly available human genome sequence and checks the designed primers to ensure these do not fall within a known repeat region or a copy number variant. We also check the SNPs against publicly available

databases to review potential alleles and minor allele frequency.

— *Darryl Irwin*

All else being equal, one would prefer to design assays for those SNPs most likely to convert to working assays. For each query SNP, the sequence around it will be available from at least two (and hopefully more) alleles. There are several properties of the flanking sequence that will influence whether the query SNP yields accurate genotypes: GC content, sequence repetitiveness, and sequence polymorphism are three of the important ones. SNPs in very AT-rich sequence and those within highly repetitive regions are less likely to result in converting assays, due to instability of hybridization between genotyping probes and their targets. Also, if the sequence is very polymorphic around the query SNP, it may not be possible to design allele- or SNP-specific genotyping probes. Segregating polymorphisms can and do interfere with the hybridization of genotyping probes to their targets, and can adversely impact the accuracy of the genotype data.

— *Stuart Macdonald*

For mapping using SNPs, it is important to have SNPs that are evenly distributed throughout the genome. We have made consensus maps for *Arabidopsis* recombinant inbred lines with a recurrent parent. In this case, it was important to choose SNPs that distinguish the recurrent parent from all the other accessions. For functional studies, SNPs that cause non-conservative amino acid

changes are prioritized. Another possibility when functional information is lacking is to choose a subset of SNPs to genotype that represent the different haplotypes. Once SNPs have been chosen, there may be technical constraints for their genotyping; additional adjacent SNPs are usually not allowed and for multiplex methods, there must be compatibility between the SNPs.

— *Heather McKhann*

Literature searches in related research areas give input on interesting candidate genes and SNPs. Databases such as NCBI and HapMap provide tools for selecting frequent SNPs and showing haplotype blocks and tag SNPs.

— *Louise Nordfors*

Searching for candidate SNPs on the basis of their putative functionality is a lottery. I try to capture the haplotype diversity using SNPs according to (i) Celera, (ii) Perlegen, and (iii) HapMap, usually in non-conserved, intronic regions and taking into account the informativity (frequency of 30 percent to 50 percent) and covering the different haplotype blocks.

— *Jose Luis Royo*

SNP scoring, such as the criteria given by Illumina SNP scoring algorithm, and htSNPs, generated by the HapMap Project. Generally speaking, we will consider genotyping SNPs at a 10 kb to 50 kb density for general studies on human diseases.

— *Huanming Yang*

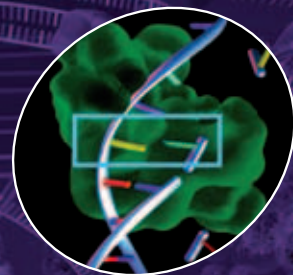
SNP Genotyping

as easy as

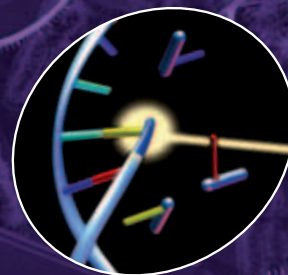
1, 2, 3



1 - Hybridize



2 - Extend



3 - Detect

12-Plex

48-Plex

GenomeLab™ SNPstream®

SNP Genotyping Systems

In an automated, scalable package, the GenomeLab SNPstream Genotyping System combines multiplexing abilities with single-base primer extension technology to provide accurate, cost-effective, medium to high throughput analysis. This system provides a scalable solution for processing a range of 4,600 to over 3 million genotypes per day. Just the kind of power you need to accelerate genetic discoveries, validation and testing of signature panels.

- Two levels of multiplexing reduces cost per genotype and per sample at any throughput
- As little as 0.04 ng DNA per genotype required
- Fast and easy optimized primer design with Autoprimer.com
- Automated image processing into genotype calls

www.beckmancoulter.com/123



Which factors are most important to achieving high genotyping accuracy?

DNA quality. Purity is the number one factor. Molecular weight is also important. However, while it may be important, it has become less of an issue today, especially since we're moving from RFLPs to SSTRs and now to SNPs. The region of the genome that we're interrogating is actually getting smaller and smaller, thus there's really not the requirement to have extremely high molecular weight DNA samples like there were when we were doing Southern blots.

— *David Duggan*

The number one factor is DNA quality. If you start with poor quality DNA, you're going to get a poor-quality result. As these reactions are becoming more highly multiplexed, the quality of the starting material is becoming more and more important. We have quality

control steps in our process where we thoroughly check the DNA before we put it in the reactions to make sure that it is high quality.

We do make some DNA quality recommendations to our clients, though we don't recommend a specific technology for DNA extraction. We recommend the A260/280 ratio being 1.8-2.2 and the A260/230 ratio being 1.6-2.4. We also recommend that the DNA is high molecular weight and available for amplification. This is a key point, the DNA may be present and quantifiable by OD but it may not amplify. To check this we recommend using a real-time PCR system such as SYBR green to check DNA activity and normalize the concentration.

We also recommend that there is minimal EDTA in the elution buffers, because the EDTA can interfere, particularly when it's at high concentration. If [EDTA] is up around 0.5 mM, then it can interfere with the activity of some of the enzymes that we use, so we recommend that they should elute in either water, 10mM TRIS or reduced TE.

— *Darryl Irwin*

The whole process of genotyping (from organism to genotype calling) should be made as routine and streamlined as possible. Some form of liquid-handling robot is essential for anything other than modest amounts of genotyping, and will markedly reduce errors introduced by manual pipetting. Reducing the number of pipette steps and

sample movement among plates is also important to minimize the possibility of contamination across wells. To minimize handling of our valuable DNA samples, using our liquid-handling robot we create many replicate sets of aliquoted DNA plates. Each set of plates holds the entire panel of DNA samples, and each well represents one sample, and contains sufficient DNA for a single PCR reaction. The DNA aliquot plates are then dried and stored at -80°C until PCR. In this way we eliminate the need to repeatedly freeze/thaw DNA, and radically streamline the process of setting up PCR reactions.

— *Stuart Macdonald*
(continued on p.18)

“We believe that the more factors taken into consideration for improving accuracy, the higher accuracy will be achieved.”

— *Huanming Yang*



Unwind.

ROSETTA
SYLLEGO™

Genetic Data Management and Analysis System


Introducing the Rosetta Syllego™ system, a genetic data management and analysis solution. The path to breakthrough science just became more straightforward. With the Rosetta Syllego system, you can better plan and execute genetic studies and ensure high-quality data to prevent loss of time in downstream analyses. It also helps you connect relevant genotypic and phenotypic study data and catalog these data and analysis results for drawing solid conclusions. All so you can focus on what really matters—advancing your research and drug development programs. Go to www.rosettabio.com/syllego for more information.

1.800.737.6583 (U.S.)
+1.206.926.1220 (Outside the U.S.)
sales@rosettabio.com



ROSETTA
BIOSOFTWARE

Copyright © 2006 Rosetta Inpharmatics LLC. All rights reserved. Syllego is a trademark of Rosetta Inpharmatics LLC. The Rosetta Biosoftware convergence icon is a registered trademark of Rosetta Inpharmatics LLC.



What quality control measures do you employ?

We always perform sequencing after RFLP assay for some samples to ensure high reproducibility. Alternatively, another restriction enzyme is selected to test the RFLP result again. Real-time PCR can provide high-throughput genotyping, but it is expensive in our experiments.

— *Hsueh-Wei Chang*

The first quality control measure would be the DNA purity. DNA purity is measured, for example, by A260/280 standard UV spectra or, more recently, fluorescence (e.g., picogreen). Also, DNA quality — although that's becoming, depending on the question, less of an issue — is measured by agarose gel electrophoresis. Lately, we've also found ourselves switching to the Agilent Bioanalyzer, which consumes considerably less DNA than an agarose gel does.

Next are quality control and quality assurance measures that are provided to us by the technology providers. For example, in the Affymetrix 500K protocol, there are several QC and QA steps along the procedure that would tip us off to something being wrong with the sample or the batch of samples. That includes, in the case of the Affymetrix 500K, an agarose gel and a quantification of PCR products following the PCR amplification step. There's also another agarose gel following the fragmentation reaction. Then, finally, in regards to the technology provider QCs and QAs, there are post-hybridization statistics that we look at, including the MDRs, MCRs, and SNP call rates.

The third of four measures that we have in place includes the more traditional replicate and duplicate controls on each plate that we process. On every

single 96-well microtiter plate, we include positive and, if the technology allows, negative controls in the form of water. In every 96-well microtiter plate, we include replicates and duplicates. By replicate, I mean the same DNA is found on each 96-well microtiter plate. So, if we have a collection of a thousand samples spread out over 10 96-well microtiter plates, there's at least one DNA in common on all 10 of those plates. That is used to look at inter-plate reproducibility and reliability. Then, on each one of those 96-well plates, we also have duplicate controls to account for reliability and reproducibility within a plate. In the case of duplicate controls, we have no less than two randomly selected DNAs per 96-well plate.

When it's available, we can look at the concordance of our SNP genotype with that of public data, such as the International HapMap project.

Those are the four measures we have in place, and all are used at some point in [the] process to assess the quality and control of the actual sample, the experiments, and the data at the end.

— *David Duggan*

We have a variety of quality control steps. We're NATA accredited to international standard ISO 17025, and as such quality is a priority for us.

The first QC step when we get the DNA is to test that the DNA is at the right OD ratios that the clients have been advised to provide. We also do real-time PCR to quantify the DNA and ensure that even if DNA is quantified by OD it is also amplifiable. We use SYBR Green on the real-time PCR platform for this.

Once our DNA has passed, we design and order our SNP assay as previously described. When those



assays arrive, we firstly check the primer sequence and quality by mass spectrometry. We then run the assays on a small number of client samples, as well as our in-house single DNA controls and our in-house pooled DNA controls. Quite a number of SNPs in the public databases are non-polymorphic, [so] by running these assays on pooled DNA we can advise our clients of expected minor allele frequencies prior to incurring genotyping charges from their full project. We also include negative controls to monitor for potential contamination.

On larger projects, we request that clients leave specific plate wells empty. When they arrive, we place a combination of oligonucleotides for which we know the masses. When the data is being analyzed we check the contents of these wells [by mass spec] to ensure there hasn't been a lab mix-up.

— *Darryl Irwin*

On the microtiter plates, among the DNA samples of unknown genotype, we place some blanks. These are valuable to indicate any assay-specific oddities during PCR amplification or genotyping. We also use various DNA controls of known genotype on the plates. If the assay discriminates poorly among individuals with different known genotypes, this indicates that the assay is problematic. In the past we have also conducted resequencing of subsets of individuals and SNPs to ensure that the error rate of our entire SNP genotyping pipeline is low.

— *Stuart Macdonald*

Initially, we tested every method (GOOD assay,

Amplifluor, TaqMan, sequencing SNPlex) using duplicate plates. With the exception of the GOOD assay, we found very high repeatability. Subsequently, each plate is genotyped one time except in cases where there are obvious problems. Prior to genotyping, the quality and quantity of the DNA is assessed. We have found that very low concentrations of DNA may affect genotyping results and further, it is important that the samples are homogeneous in concentration. Then, in each plate, a number of controls are essential, including negative controls (water) and positive controls (previously sequenced individuals with all possible genotypes).

— *Heather McKhann*

Results are checked and confirmed regularly using different methods: SNPlex, pyrosequencing, TaqMan, and sequencing.

— *Louise Nordfors*

First, checking the Hardy-Weinberg equilibrium law. Then we perform 5 percent to 10 percent re-extraction from blood and re-typing with double-blind check. For association studies, 2 percent mistyping is assumable. If more, you must discard the marker.

— *Jose Luis Royo*

Five to 10 percent duplicates of the total samples used for QA. If analyzing pedigree samples, we can also estimate the genotyping quality from the pedigree information. And 1 percent of the total samples are always [included] as negative controls.

— *Huanming Yang*

Relax. We've already built a world-class microarray facility.

So you don't have to.

Looking to expand your capacity or for a backup to your internal capabilities?

Rely on Gene Logic for superior services and results. Our services include:

- SNP genotyping from DNA, blood, or tissue
- Gene expression profiling from tissue, cells, whole blood, or RNA
- Optimized gene expression profiling of whole blood
- qRT-PCR
- RNA isolation from low-yield tissues
- Bioinformatic analyses

Why build it when you can outsource it?

GENE  LOGIC

610 Professional Drive
Gaithersburg, Maryland 20879
Phone: 1.800.GENELOGIC or
+1.301.987.1709
Fax: 301.987.1701
Email: info@genealogic.com
Web: www.genealogic.com



What methods or tools do you use for the analysis and visualization of data?

Excel and SPSS analysis.

—*Hsueh-Wei Chang*

We are using home-grown visualization software to, for example, display coverage of a gene of interest in a candidate gene study. In some cases, we plug it into graphical user interfaces like that offered by the HapMap.

Secondly, we are using commercial software for linkage analysis. We do a lot of linkage analysis with Affymetrix's 10K. We are using commercial software as well as relying on our statistical collaborators to analyze the linkage data.

We're relying on many more statistical collaborators and evolving software for whole-genome scans. While we have ideas about how to analyze that data today, this is really an evolving area of research, and I rely on my statistical collaborators for direction in analysis.

Some of us researchers play a key role in providing statisticians with direction, and some of us are taking a stab at the analysis itself, but that's a little dangerous. We're not statisticians and we could be creating more havoc than help. Just as we saw with the microarray world 10 years ago, we're seeing software companies bring, in some cases, user-friendly software to market that a lot of statisticians are having a hard time swallowing. Because ease-of-use doesn't mean statistical robustness. Just because you can select an algorithm or a test, it doesn't mean

it's the right test. It doesn't substitute for having a statistician on your team.

—*David Duggan*

For analyzing the genotyping data, we actually use the Sequenom Typer software for the Sequenom platform, and we use the Applied Biosystems GeneMapper software for the Applied Biosystems platform. We supply back to our clients the genotypes only. Clients may request the mass spectra or electropherograms; however, these are not

routinely supplied as the majority of our clients do not have the software to analyze these.

We also analyze a scatter plot of the peak areas of the two alleles to ensure that our genotypes are clustered very clearly. They fall into three separate groups: homozygous wild type, heterozygous, and homozygous mutant. This

plot gives us a lot of information about the quality of the DNA that's going through and the accuracy of the genotypes. We always expect to see three separate groups, but on occasion, we have observed a split in the heterozygous groups, which indicates that there is a copy number variant and this is also advised back to our clients.

We find the Sequenom platform is a very useful tool for looking at copy number variants because it can perform quantitative genotyping to a high level of accuracy. Sequenom's platform can perform

"Some of us researchers play a key role in providing statisticians with direction, and some of us are taking a stab at the analysis itself, but that's a little dangerous."

—*David Duggan*

quantitative genotyping to look at allelic frequencies in pooled DNA samples. We can use the same methodology to analyze individual DNA samples for copy number variants with little additional cost.

— *Darryl Irwin*

The genotypes we collect are all derived from intensity measurements on spotted arrays. We use the ArrayVision software (GE Healthcare) to semi-automatically extract the intensity data, and with this data in hand use custom scripts written in the freely available statistical programming language R to cluster the individuals and call genotypes. These scripts are described in Macdonald et al. (2005). The advantage of R over licensed commercial software is that scripts can be easily customized. Since there is a wide community of academics developing R scripts and depositing them in accessible databases, often there are useful scripts and functions readily available for downstream analysis of the genotype data.

— *Stuart Macdonald*

In general we use the methods provided with a given

platform. For Amplifluor and TaqMan, we have used Applied Biosystem's SDS2.0 software, and for SNPlex, the Genemapper software. For the GOOD assay, software called SNPmaster has been developed for internal use at the French National Genotyping Center. Following analysis, genotyping data is submitted to a database located at the INRA's Unité de Recherche Génomique-Info.

— *Heather McKhann*

We have created our own database for storing the data. Statistical analyses are mainly performed in JMP IN, which also provides graphs, etc.

— *Louise Nordfors*

The traditional statistical ones: Excel sometimes, and mainly SPSS, STATA, Episheet, and EPI-Info.

— *Jose Luis Royo*

Illumina Gencall, Sequenom MassArray typer, and PerkinElmer SNPscorer. All work reasonably well.

— *Huanming Yang*

It's Friday morning.

You've got 1000 samples genotyped and you need to run analysis and get your results by Monday morning.

Possible?

If you have Progeny LAB it is...you'll even make your Friday dinner appointment.

Don't believe us?

Try it yourself for free at www.progenygenetics.com/lab.html. Even better, ask one of our customers.

PROGENY LAB

Unlimited Samples. Unlimited Genotypes. Unlimited Power.



Progeny LAB provides multi-user integration of phenotypic and genotypic data in a completely secure environment.

- Perform whole-genome association studies
- Manage marker sets for both SNPs and STRs
- Complete sample and plate management
- Import large arrays such as 500K
- Import native genotype files from varied platforms
- Export to analysis for association or linkage

Progeny Software, LLC | 574.968.0822 Phone | info@progenygenetics.com | www.progenygenetics.com

Q3: Which factors are most important to achieving high genotyping accuracy?

(continued from p.11)

This depends very much on the method being used. We found that the GOOD assay, which separates alleles based on their molecular weight, is highly accurate. However, as it comprises many steps, there is a loss in repeatability. In fluorescent methods, the key factor is the design of the primers/probes. Depending on the method, this may be done by the manufacturer, in which case there is little control over design; in certain cases, the researcher may design their own primers/probes, for example for TaqMan, using PrimerExpress.

— *Heather McKhann*

A highly accurate genotyping method together with careful research about the gene and SNPs to be investigated to avoid, for example, multiple sequence variation sites and genotyping of pseudogenes. Of course, highly skilled and careful technicians are crucial.

— *Louise Nordfors*

Automation in all processes. In DNA extraction, working in 96-well or 384-well formats and with robots preparing the PCRs and the following post-process.

— *Jose Luis Royo*

Basically, we believe that the more factors taken into consideration for improving accuracy, the higher accuracy will be achieved. However, too many factors may lead to difficulties in sample handling and automation. Right now, we prefer ASO- (allele-specific oligonucleotides) and LSO- (locus-specific oligonucleotides) based hybridization, primer extension and ligation.

What's more, primer design algorithms, oligo synthesis protocols, redundancy of reactions, and quality of DNA samples are also very important in ensuring maximum accuracy in SNP genotyping methods.

— *Huanming Yang*

List of resources

Our panel of experts referred to a number publications, which we've compiled in the list below.

Publications

Chen X, Levine L, Kwok PY (1999) **Fluorescence polarization in homogeneous nucleic acid analysis.** *Genome Res.* 9:492-498.

Giancola S, McKhann HI, Bérard A, et al. (2006) **High throughput genotyping in plants: comparison of three current technologies.** *Theor Appl Genet.* 112(6):1115-1124

Landegren U, Kaiser R, Sanders J, Hood L (1988) **A ligase-mediated gene detection technique.** *Science.* 241:1077-1080.

Macdonald SJ, Pastinen T, Genissel A, et al. (2005) **A low-cost open-source SNP genotyping platform for association mapping applications.** *Genome Biol.* 6:R105.

Matsuzaki H, Dong S, Loi H, et al. (2004) **Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays.** *Nat Methods.* 1:109-11.

Sauer S, Lechner D, Berlin K, et al. (2000) **A novel procedure for efficient genotyping of single nucleotide polymorphisms.** *Nucleic Acids Res.* 28:E13

Shen R, Fan JB, Campbell D, et al. (2005) **High-throughput SNP genotyping on universal bead arrays.** *Mutat Res.* 573:70-82.

Syvänen AC (2001) **Accessing genetic variation: genotyping single nucleotide polymorphisms.** *Nat Rev Genet* 2:930-942.

Acknowledgments

Many thanks to Aurélie Bérard of INRA for advising on the answers submitted by Heather McKhann.

Finest in Fine Mapping

The New iPLEX™ Assay for SNP Genotyping to follow up Genome Wide Scans

- ✓ Scalable Throughput
- ✓ Cost Effective
- ✓ Accurate
- ✓ Simple Workflow



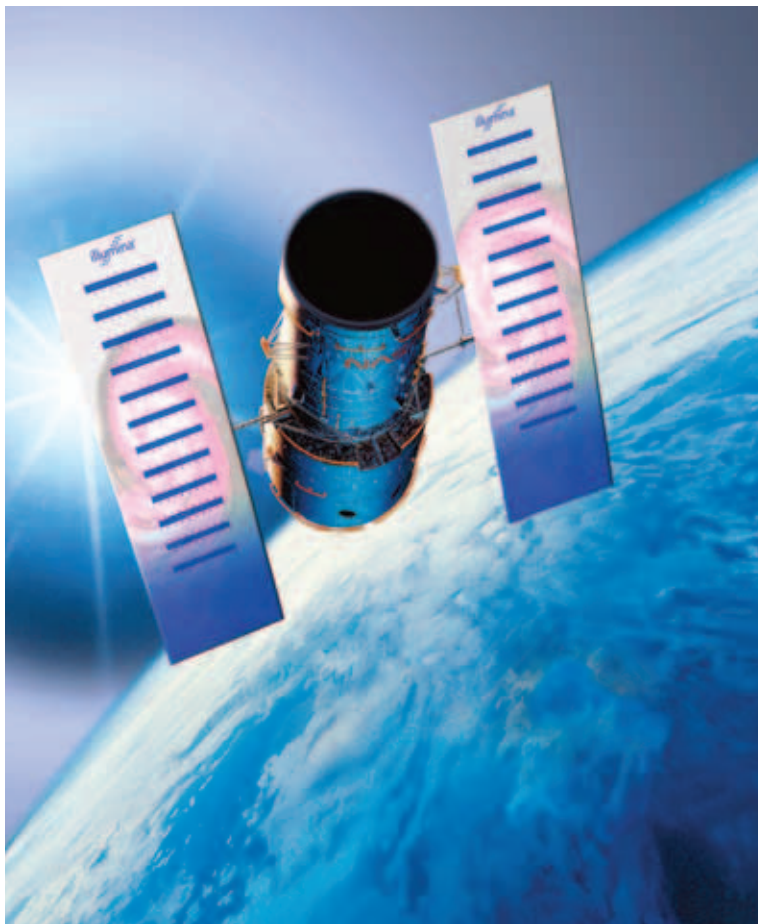
iPLEX is the Leading Solution for Fine-Mapping:

- FAST and ACCURATE RESULTS for < 5 cents per genotype
- Ideally suited for studies using 10s to 100s of SNPs per sample
- Highest available DATA QUALITY
- FLEXIBLE assay design and rapid turn around time
- VERSATILE system that supports DNA Methylation and Quantitative Gene Expression Analysis
- SNP Genotyping Services available

www.sequenom.com

SEQUENOM®

THE SEARCH FOR INTELLIGENT SNP SELECTION IS OVER



INTRODUCING INFINIUM™ WHOLE-GENOME GENOTYPING

- Unrestricted genome-wide access
- Industry-leading call rates, accuracy and reproducibility
 - Fully integrated 100K, 250K and 500K solutions—
with automation and LIMS options

Look into Infinium.

Visit www.illumina.com/infinium