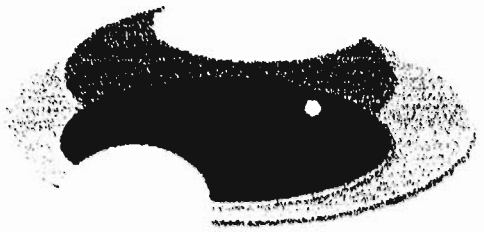
FUNDACION PARA LA INNOVACIÓN AGRARIA

PROGRAMA DE FORMACIÓN PARA LA INNOVACIÓN AGRARIA

APOYO A LA PARTICIPACIÓN EN ACTIVIDADES DE FORMACIÓN

BID-FP-V-2003-1-M-025

MATERIAL ANEXO



INTRODUCCIÓN AL DIAGNÓSTICO MULTIPLE USANDO BIOCHIPS

Pertenece a:_____

Introducción al Diagnóstico múltiple usando Biochips

CONTENIDO

Objetivos

Al final de esta sección Ud. podrá:

1. Explicar cual es el origen de los Biochips
2. Por qué son una nueva y valiosa herramienta para el diagnóstico molecular en el laboratorio clínico.
3. Explicar como opera un Biochips y cómo se relaciona con el diagnóstico mediante PCR múltiple.
4. Explicar el proceso , incluyendo amplificación múltiple y detección mediante biochips..

INTRODUCCION

Los métodos de diagnóstico deben tener cualidades fundamentales como poseer una máxima especificidad, sensibilidad y rapidez. Con el aumento de la especificidad se evitan las reacciones cruzadas y por ende los falsos positivos; y con el aumento de la sensibilidad se intenta detectar la mínima cantidad de patógeno existente con lo que se evitan los falsos negativos. La rapidez con que se entreguen los resultados es fundamental para una oportuna decisión en cuanto a las líneas de acción a seguir.

A principios de los 80 ocurrió una gran revolución tecnológica con el descubrimiento de los anticuerpos monoclonales aplicados para la detección de patógenos. Se pensó que gracias a la alta especificidad y sensibilidad que presentaba esta técnica, se habían solucionado gran parte de los problemas de diagnóstico. Sin embargo, con el correr de los años se comprendió que los anticuerpos monoclonales eran muy útiles para algunos casos, pero no servían para otros, más aún, quedaban muchos casos sin resolver.

A mediados de la década de los 80, apareció otra gran revolución científica la que se denominó reacción en cadena de la polimerasa (PCR). Esta herramienta tecnológica consiste en aumentar artificialmente el contenido del material genético del patógeno mediante la amplificación de una región específica del genoma utilizando la enzima Taq DNA polimerasa. Esta técnica se caracteriza por ser altamente sensible ya que es capaz de detectar cantidades mínimas del patógeno, lo que generalmente corresponde a etapas tempranas o/y crónicas de la enfermedad. Con este descubrimiento también se han llevado a cabo grandes progresos en la especificidad del diagnóstico de patógenos, ya que no presenta reacciones cruzadas, superando con creces a los métodos inmunólogicos. La rapidez con la que se obtienen los resultados es otra característica diferencial de esta técnica.

Cada día van apareciendo más modificaciones de la técnica de PCR, como la automatización o la combinación de ella con otras técnicas, que permiten día a día un diagnóstico más certero.

Recientemente en el campo de las ciencias biomédicas, estamos asistiendo desde hace años a un boom de la biología molecular y más concretamente de la genética y la genómica, gracias a la continua implementación y desarrollo de técnicas experimentales a disposición de los investigadores en los laboratorios. Los biochips representan una de las herramientas recientes con las que cuentan los investigadores para hacer frente a la resolución de los problemas biológicos basados en nuevos enfoques que se orientan a la obtención masiva de información. El desarrollo de estos enfoques integrados para el análisis ha venido de la mano de la capacidad de gestionar y almacenar grandes cantidades de información, por tanto no es de extrañar que la llegada de estos dispositivos haya coincidido con la madurez de la bioinformática en la cual se sustentan la realización de los experimentos en general y el análisis de los datos que de ellos se obtienen en particular. A estos dispositivos también se les conoce con otros nombres como Micromatrices de material biológico, Microarrays, y según el tipo de material inmovilizado como DNA arrays o Chips Genéticos, Protein Chips o Tissue Chips. Estos dispositivos están constituidos formando una matriz con el material biológico que se inmoviliza sobre ellos de forma que se sabe en cada punto de la matriz que es lo que se ha depositado permitiendo el posterior análisis. El número de posiciones en estas matrices puede llegar a alcanzar las decenas de miles. El fundamento de los biochips se encuentra en el desarrollo y miniaturización de las técnicas de afinidad que se conocen y han venido empleando desde hace años como una herramienta común en biología molecular. El desarrollo de los primeros ensayos de afinidad con muestras inmovilizadas sobre soportes sólidos se remonta a los primeros ensayos inmunológicos que se desarrollaron en los años 60´s y en los que se inmovilizaban sobre una superficie de antígenos o anticuerpos para su detección. El siguiente paso en la evolución hacia estos dispositivos se dio en los años 70´s cuando Edwin Southern, comenzó a emplear filtros de nitrocelulosa para que actuasen como soporte sólido para la adhesión de moléculas de DNA. El DNA sí inmovilizado no interacciona con las otras moléculas inmovilizadas pero mantiene su capacidad de

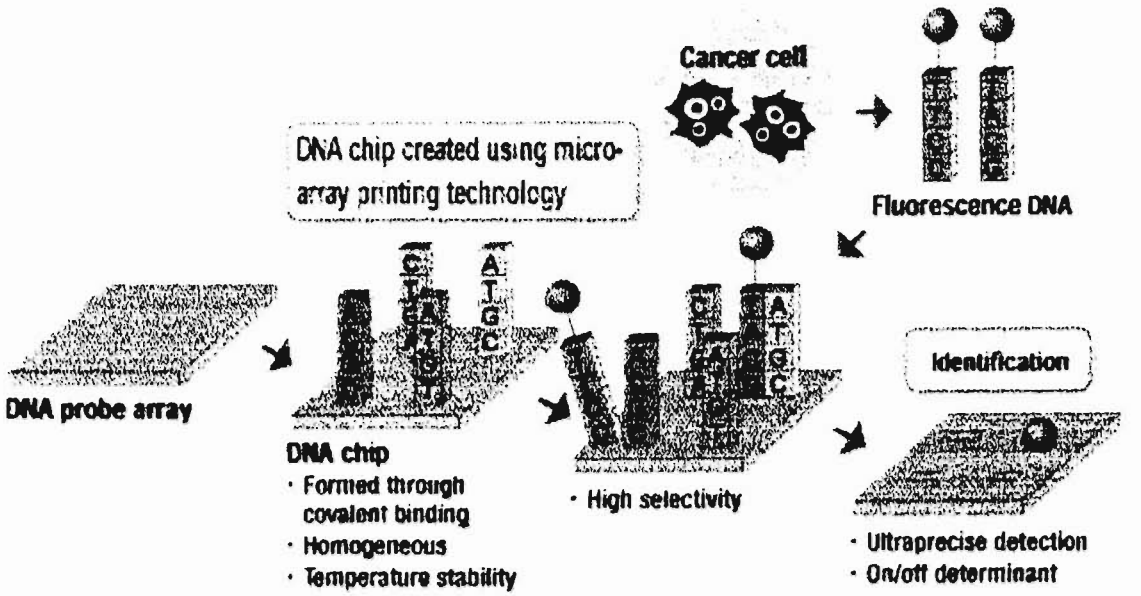
hibridar con moléculas complementarias en disolución. La detección de estas hibridaciones se realizaba mediante la detección de un marcador radiactivo en un revelado por autorradiografía. A este tipo de técnica se la bautizó con el nombre de Southern blot, que después se extendió al campo de la inmovilización de proteínas y RNA. Con la puesta a punto de la técnica de Southern, el siguiente paso en el camino hacia la aparición de los biochips consistió en la construcción de matrices de material biológico inmovilizado por este mecanismo empleando para ello superficies porosas como son las membranas de nitrocelulosa o nylon. Posteriormente se comenzó a trabajar con el empleo de superficies con unos tamaños de poro más reducidos y con soportes sólidos como pueden ser el vidrio y el silicio. Paralelamente con la llegada y desarrollo de las técnicas de miniaturización también se comenzó a disminuir el tamaño de los puntos de material depositado sobre la superficie, consiguiendo de esta manera una mayor densidad de integración en las matrices. La aplicación de las técnicas de miniaturización condujo hasta el desarrollo de las Micromatrices.Uno de los acontecimientos más importantes se produjo a finales de la década de los 80´s cuando en un laboratorio de la compañía entonces llamada Affymax, un grupo de cuatro científicos, Stephen Fodor, Michael Pirrung, Leighton Read y Lubert Stryer, que trabajaba en la síntesis sobre superficies sólidas de péptidos, terminó desembocando en la plataforma GeneChip, que ha sido desarrollada por Affymetrix, una compañía escindida de Affymax en 1993. La importancia de este paso radica en la gran capacidad de miniaturización alcanzada por este sistema. Posteriormente al nacimiento de la tecnología desarrollada por Affymetrix se han ido sucediendo la aparición de nuevas compañías y nuevos desarrollos que han permitido alcanzar el alto grado de diversidad tecnológica existente en la actualidad. Esta técnica comenzó a ser operativa entre el año 1993 y 1995, desde entonces a la fecha ha superado problemas relacionados con sensibilidad y reproducibilidad lo que ha permitido a partir del año 2000 su consolidación como una técnica habitual en investigación, más aún se prevé el inicio de su aplicación clínica para el año 2003. Cabe destacar que más de 200 centros los utilizan habitualmente y colaboran en el desarrollo de la tecnología. Las múltiples aplicaciones de esta tecnología son:

- Expresión diferencial de tumores (Cáncer-Tumores)
- Detección de mutaciones o polimorfismos (Genotipificación)
- Detección de microorganismos (Diagnóstico)
- Screening de fármacos (Farmacogenómica)

La aplicación de esta técnica al diagnóstico de microorganismos es muy interesante ya que abre grandes posiibilidades a realizar diagnóstico múltiple, es decir, detectar varios patógenos a partir de una misma muestra en una misma reacción, lo que significa una característica adicional de gran relevancia para el diagnóstico en sectores productivos donde deben ser chequeadas altas cantidades de muestras y patógenos en periódos de tiempo y a un costo limitado, Esto ha incentivado la creación de empresas cuyo objetivo es lograr el desarrollo de biochips y lectores accequibles para uso masivo, tanto en términos de costos como técnicamente.

¿Cómo trabaja un BIOCHIP?

A continuación se presenta un diagrama que explica cómo opera un biochip. Este consiste en una superficie no mayor al tamaño de un portaobjeto en la cual se encuentran unidos cientos de oligonucleotidos ubicados matricialmente. Luego se incorporan a la matriz productos de PCR y/u oligonucleotidos marcados -.En el caso de que alguno de los productos incorporados sea complementario con los oligos ligados a la matriz, permanecerán hibridados y mediante algún sistema de detección es posible identificar cual punto fue complementario con el producto incorporado, así se puede determinar una serie de variables simultáneamente en algún ensayo de interés. En el caso de su uso para el diagnóstico, se incorpora producto de PCR de los patógenos de interés y en el caso de que exista producto(s) amplicado(s) deberá(n) hibridar con el ligando correspondiente al patógeno, así se puede utlizar esta matriz con fines diagnósticos múltiples.



Proceso de PCR múltiple.

El diagnóstico múltiple tiene por objeto detectar a partir de una misma muestra en un mismo ensayo, más de un patógeno a la vez. Para ello, se diseñan partidores específicos para los patógenos de interés y además estos deben cumplir con requerimientos determinados para que puedan actuar como partidores de la mejor manera posible, sin interferir con el resto de las reacciones de los otros partidores que ocurren en el mismo tubo. Esto significa que cuando se lleva a cabo una reación de amplificación múltiple, lo que ocurre es que en un mismo tubo pueden estar ocurriendo simultáneamente una serie de reacciones paralelas que van a resultar en productos específicos amplificados . Luego mediante el uso del Biochips ,se pueden detectar simultáneamente una serie de puntos y/o los patógenos de interés a partir de la misma muestra.

El método de PCR múltiple consta de cuatro pasos:

1. Preparación del Reactivo (Pre-PCR)
2. Preparación de la Muestra (Pre-PCR)
3. Amplificación (Post-PCR)
4. Detección (Post-PCR)

Preparación del Reactivo (Pre-PCR).

- **Master Mix (Mezcla Maestra):** es el reactivo primario utilizado en PCR. Este reactivo tiene los componentes necesarios que permiten la amplificación específica de la secuencia objetivo:

1. *Nucleótidos marcados*: Unidades individuales de dATP, dGTP, dUTP y dCTP marcado., .
2. *Cofactor de la enzima*: Mg^{+2} para la polimerasa *Taq* o Mn^{+2} para la Transcriptasa reversa
3. *Polimerasa Taq o Transcriptasa reversa:* Enzima utilizada para la replicación del DNA o RNA.
4. *Buffers*: Se usan para mantener el pH y la concentración de sales adecuados.
5. *Partidores o Primers para patógeno 1, para el patogeno 2 , para el patógeno 3, etc*: Son pequeñas secuencias sintéticas de DNA de una sola cadena que consisten de no más de 20 a 30 bases. Son complementarios a la secuencia de DNA/RNA de cada uno de los patógenos de interés.

Preparación de la Muestra.

Las muestras que se pueden usar son: tejido, sangre, ovas ,fluido ovárico, semen. Las técnicas de preparación de la muestra variarán dependiendo del tipo de ácido nucleico que se desea extraer o aislar.

Fudamentalmente la muestra debe ser lisada para liberar al patógeno, luego se deben separar las proteínas del ácio nucleico y finalmente se debe recuperar el DNA y/o RNA y una vez extraído de la muestra, se añade al tubo de reacción con la Mezcla Maestra.

Amplificación múltiple

Una vez que se ha añadido el DNA/RNA extraído la muestra al tubo de reacción con la Mezcla Maestra, se somete a la reacción de PCR (con Transcripción Reversa si fuese necesario) múltiple dentro del termociclador.

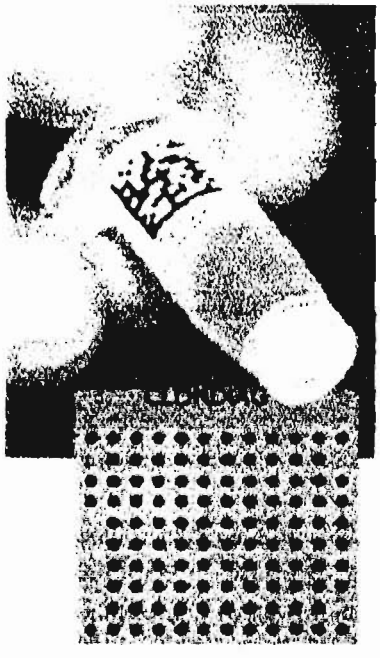
Detección

Una vez que ha finalizado la reacción de PCR múltiple , el(los) producto(s) amplificado(s) se somete(n) a una reacción de hibridación con la matriz diseñada especificamente para la detección del patógeno(s) de interés, esta reacción consiste en una serie de pasos de hibridación y lavado. Finalmente la matriz se somete a la lectura mediante un lector que es capaz de detectar las señales emitidas por el producto marcado en el caso de que haya hibridado con algún ligando presente en la matriz. Esto se traduce en una serie de puntos marcados y al tener conocimiento previo de la matriz diseñada se puede determinar que producto logró hibridar con la matriz.

Así se puede detectar simultáneamente en una pequeña superficie una serie de patógenos de interés

ArrayTube®

Instruction Manual



Release 2002

Table of Contents

General Information

Notes

This booklet describes in detail all of the steps in an ArrayTube® hybridization and detection experiment. Hybridization is performed directly in the ArrayTube (AT) against the (customized) microarray integrated within the microtube. For the detection of the hybridization pattern, the ArrayTube® reader ATR 01 is required. User instructions for the ATR 01 reader are available in the separately provided ***ATR 01 Reader Manual***. Since the detection of the hybridization pattern is done online within the reader (after the enhancement reaction has been initiated), carefully read all of the notes for ATR 01 use before starting your detection experiment. Take all of the notes and recommendations into consideration to ensure reliable and successful experiments. If there are any complications or questions, contact support@clondiag.com.

Storage and Handling

To secure optimal performance, the ArrayTubes® (AT) are shipped in light-protective foil sealed under inert gas. Within these packages, the AT's can be stored at room temperature for at least 6 months. After opening these packages, we recommend that the AT's be used within a week to prevent a loss in their performance. Keep the tubes inside the foil in which they were shipped in a dry and dark location. Protect the ArrayTubes® from direct sun light exposure. Avoid touching or scratching the microarray on the bottom of the tube. Information on how to store the processed ArrayTube® upon completion of the experiment can be found in the section "Post Assay Procedure " on page 15.

Precautions

- The ArrayTubes® are for research laboratory use only.
- Follow standard lab safety regulations when using the AT's with any hazardous material that may be required in your individual experiments.
- Do not expose the ArrayTube® or the solutions to direct sun light.
- Never scratch or touch the array surface (e.g. during pipetting steps with the pipette tip).
- Do not bring any metallic surface (tweezers, etc.) into contact with the AT and AT solutions.
- Keep your solutions nuclease-free.

Reagents and Materials

a) Supplied Materials

- *AT Hybridization Buffers*

  Ready-to-use solutions.

  We offer three different AT Hybridization Buffers, I, II and III. Depending on your individual assay, buffer I, II or III will be best for your experiment.
- *AT Blocking Mix*

  Ready-to-use powder blend for dissolving in water or a buffer.

b) Additionally required reagents

Solution	Remarks	Recommended Company*
• Streptavidin-gold conjugat (STP.5)	Streptavidin - colloidal Gold labeled (5nm)	British Biocell**
• Silver enhancement kit***		British Biocell**
• SDS – wash solutions	for preparation see below	*Common lab suppliers*
• SSC – wash solutions		

* These suppliers are recommendations only.

** Please ask for your local distributors.

*** An alternative homemade silver solution is available. For more information, please contact us.

c) Preparing AT Assay Solutions

The washing buffer procedure listed below is only a recommendation. Depending on your individual assay, alternative buffer solutions may result in better results.

Name	Procedure
• AT Blocking solution	• Prepare 1-2 ml of 2% solution of AT Blocking Mix in AT Hybridization Buffer. Mix and vortex for dissolving.
• TritonX solution	• Prepare a 0.001% TritonX 100 aqueous solution
• Silver solution (British Biocell)	• Mix 70 µl of enhancer and initiator solution in a ratio of 1:1
• 2xSSC (wash solution)	• Dissolve 17.53 g NaCl and 8.82 g Na-citrate in 100 ml H_2O
• 2xSSC/SDS (wash solution)	• Dissolve 17.53 g NaCl and 8.82 g Na-citrate in 100 ml H_2O, add SDS to a final concentration of 0.2%
• 0.2xSSC (wash solution)	• Dissolve 1.753 g NaCl and 0.882 g Na-citrate in 100 ml H_2O

AT Experiment Procedure:

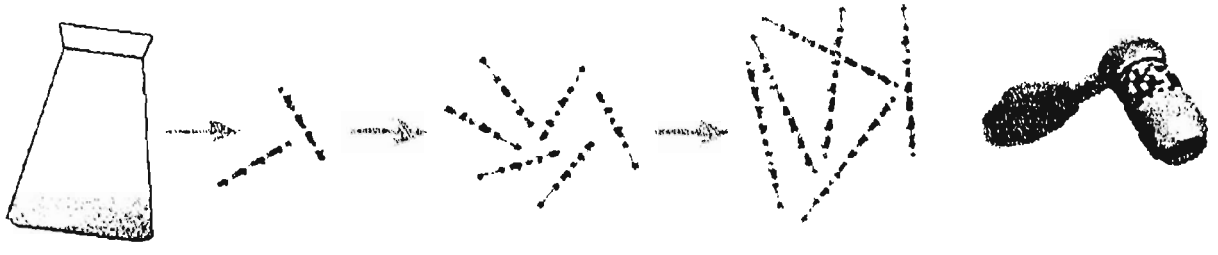
The illustrations on the next two pages give an overview of the methods used in an AT assay experiment. Within the ArrayTube®, containing the individual probe array, the hybridization, washing, conjugation and detection steps are performed. All sample preparation procedures, including labeling, are usually done outside of the ArrayTube®. The main step in the labeling process is adding gold particles to the target molecules. This can be done either by the direct binding of gold to the target or by indirect gold labeling via the biotin-streptavidin interaction, in which gold particles are linked to streptavidin molecules. After hybridization, all target molecules binding to the microarray will carry gold particles. Upon adding a silver enhancement solution, the gold particles will induce the precipitation of elementary silver specifically at those spots where the gold-labeled target molecules are bound. The hybridization pattern can be visualized.

In our instructions, we recommend doing the gold labeling via the indirect method because it is more reliable and less expensive.

AT Experiment Workflow

1. Sample Preparation

2. ArrayTube® Conditioning



Sample | Targets (e.g. RNA) | Marker (e.g. biotinylated primer) | Labeled Targets | ArrayTube® with specific microarray

At first, the sample preparation procedures like extraction, purification, amplification and biotin labeling need to be performed using standard methods that are best suited to your specific target. For biotinylation of the target molecules, use a standard procedure when introducing biotinylated primers or nucleotides during PCR or for direct chemical coupling of biotin to the target molecules.

The second preparation step includes pre-conditioning the AT probe-array.

3. Target Hybridization

4. Washing + 5. Blocking

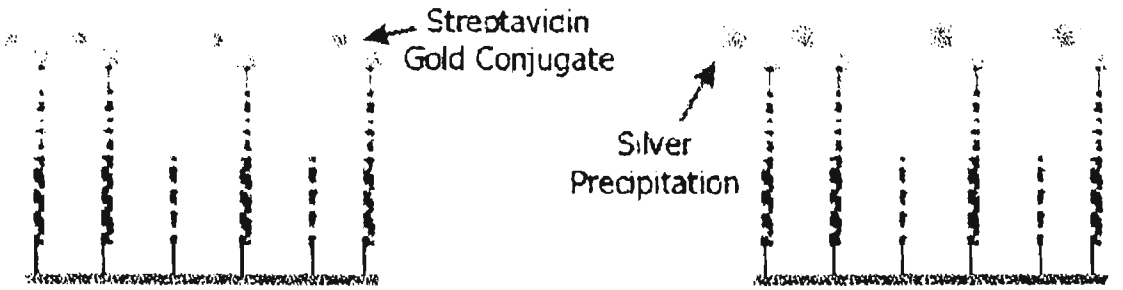
Probes on ArrayTube® Chip

Hybridizated Target

Target hybridization and the washing and blocking steps are performed within the ArrayTube® reaction vial.

6. Conjugation

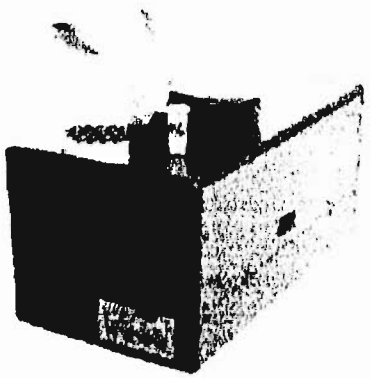
7. Silver Enhancement



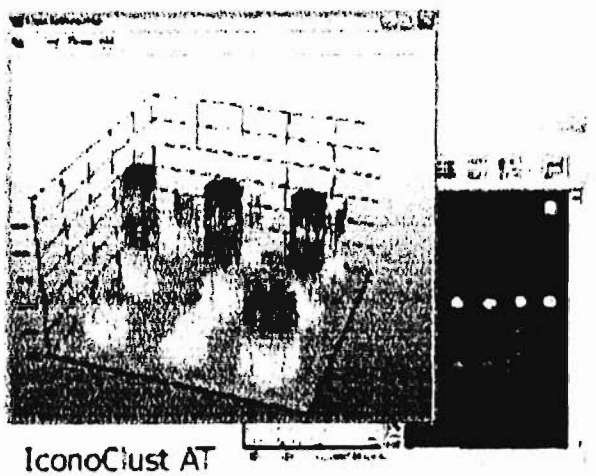
After hybridization, a streptavidin gold conjugate is added which binds to the biotinylated target molecules. The gold particles induce a silver precipitation reaction that results in a detectable hybridization pattern.

8. Image Detection

9. Image Analysis



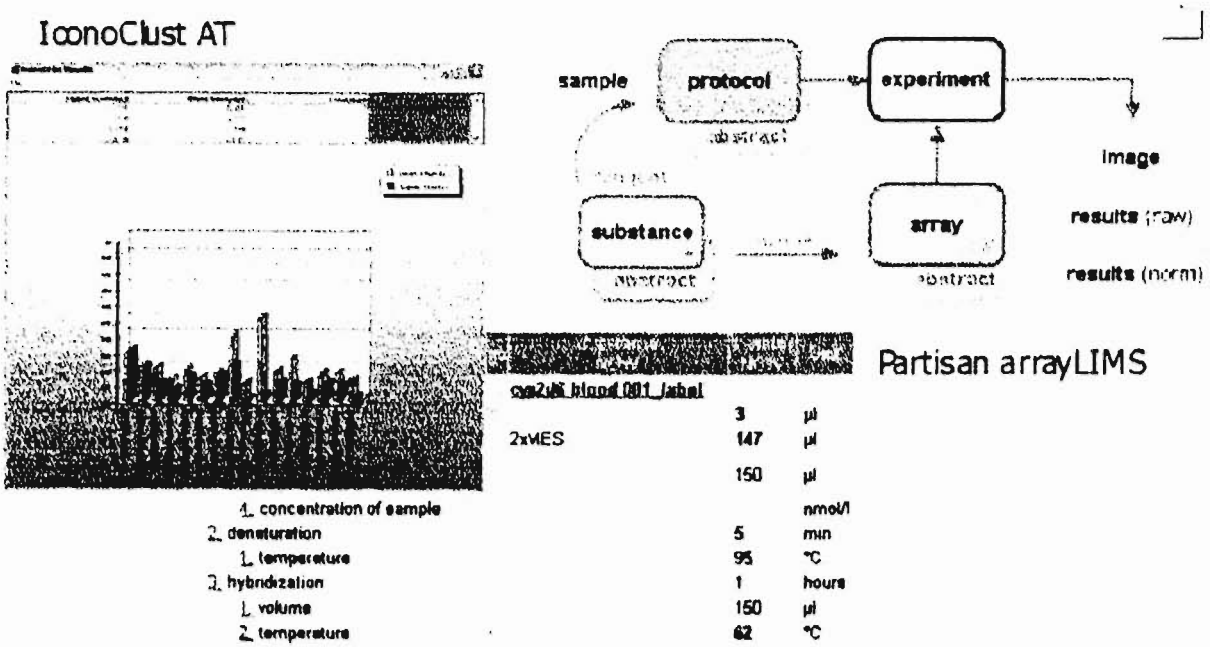
ArrayTube® Reader



IconoClust AT

The ArrayTube® Reader detects as a time series the AT hybridization pattern, which has been recorded and analyzed by IconoClust-AT software.

10. Result Evaluation and Data Management



The array image results can be evaluated and prepared for further processing with IconoClust-AT. As an option for your experiment, our software system Partisan arrayLIMS, which manages the whole experiment data workflow, is available.

AT Experiment Steps

All of the AT experiment steps are summarized in Table 1. The required procedures are described on the following pages.

Table 1

Step	Experiment Step*	AT Component*	Notes
1.	Sample Preparation and Biotinylation		Performed by using any standard procedure that is best suited to the user's specific target.
2.	AT Conditioning	AT	
3.	Target Hybridization	AT	
4.	Washing	AT Hybridization Buffer	Supported by AT assaying protocols.
5.	Blocking	AT Blocking Mix	
6.	Gold labeling (conjugation reaction)		
7.	Silver Enhancement	AT	Supported by AT assaying protocols. Silver Enhancement and Detection are parallel processes.
8.	Detection	AT ATR 01 IconoClust-AT	
9.	Image Analysis	IconoClust-AT	
10.	Post Assay Procedure	AT	

*All of the AT experiment steps can be supported by PARTISN arrayLIMS, CLONDIAG®'s powerful array laboratory information management system. PARTISAN arrayLIMS organizes all data around microarray experiments and can easily be used with the AT-System.

Before starting

Prepare all solutions according to the instructions on p. 5. Prepare a 60°C water bath or a heating block for use in step 4. Prepare a heating block at 95°C for denaturation of the hybridization sample. Turn on the AT Reader and adjust its working temperature to 25°C. Before performing step 8, start the appropriate IconoClust-AT functions required for AT online detection (see also ***ATR 01 Manual***).

1. Sample Preparation

Perform sample preparation, purification and amplification by following the procedures that you would normally use and that are most appropriate for your specific target. For target labeling, we recommend using indirect labeling via the biotin-streptavidin-gold complex. Follow standard biotinylation procedures like introducing biotinylated primers or nucleotides during PCR or having direct chemical coupling of biotin molecules to targets.

2. AT Conditioning

Before performing hybridization, the AT has to be pre-conditioned. Wash the AT 2-4 times every step with 100 µl of the hybridization buffer for 5 min at 500rpm* each time.

*Listed rpm values are recommendations only.

3. Sample Hybridization

Required materials:	**Volume**
Sample in hybridization buffer	100 µl
2 x SSC/SDS	500 µl
2 x SSC	500 µl
0.2 x SSC	500 µl

- **Preparing the AT hybridization sample**

3.1 Prepare an aliquot of your labeled sample in 100 µl of hybridization buffer to a final concentration of ≥ 100 pM. We recommend that you optimize the concentration of the sample by testing different aliquots.

- **Hybridization (Outside of the AT)**

3.2 Heat the hybridization sample at 95° C for 2 min.

3.3 Load the sample into the AT and incubate it at 60°C for 1h at 500 rpm. You might need to optimize the incubation time and temperature depending on the sample used.

Caution: Never rinse the ArrayTube with distilled water after hybridization because it will result in denaturation.

4. Washing after hybridization

4.1 Remove the sample solution.

4.2 Add 500 µl of 2 x SSC/SDS wash solution. Wash the tube for 10 min at 30°C and at 500 rpm.

4.3 Discard the solution. Add 500 µl of 2 x SSC wash solution. Wash the tube for 10 min at 20°C at 750 rpm.

4.4 Remove the solution. Add 500 µl of 0.2 x SSC wash solution. Wash the tube for 10 min at 20°C at 750 rpm.

4.5 Dry the tube in a Speed Vac.

5. Blocking

Required materials	**Volume**
AT Hybridization Buffer	100 µl
AT Blocking Mix	2 % (v/wv) in AT Hybridization Buffer

5.1 Prepare a fresh 2% (v/wv) solution of AT Blocking Mix in the hybridization buffer and shake it at 500 rpm.

5.2 Add 70 µl of the blocking buffer solution to the tube. Incubate it for 15 min at 30°C.

6. Gold Labeling (Conjugation Reaction)

If using the recommended indirect target labeling, the biotinylated target molecules will have already hybridized to the probe arrays. For labeling, the streptavidin-gold conjugate now has to be added following the procedure below:

Required materials:	**Volume**
Streptavidin.5 gold conjugate (conjugation solution)	250 pg/µl
2 x SSC/SDS	500 µl
2 x SSC	500 µl
0.2 x SSC	500 µl

- **Conjugation**

6.1 Add the conjugation solution into the AT, which contains the blocking solution of step 5. Incubate it for 15 min at 30°C.

6.2 Carefully remove the complete solution with a pipette.

- **Washing**:

6.3 Wash with 500 µl 2 x SSC/SDS for 10 min at 30°C

6.4 Wash with 500 µl 2 x SSC for 10 min at 20°C

6.5 Wash with 500 µl 0.2 x SSC for 10 min at 20°C

Note: After the AT is preconditioned and the sample is prepared, we recommend processing it within a day. During this time, keep the AT and the sample on ice until further processing.

7. Silver Enhancement

When adding the silver enhancement solution, the gold particles attached to the hybridized target molecules will immediately induce the precipitation of elementary silver corresponding to the hybridization pattern.

Note: The AT reader and the appropriate IconoClust-AT functions for online AT image detection need to be started at this point. See also the **ATR 01 Manual.**

Before Recording:

- Start the IconoClust-AT program. Set the correct directory for saving the images, the collecting interval time and the number of readings (see p. 12 in the ATR 01 Manual) . For your first experiments, we recommend recording 120 images with a time interval in between of 20 sec. The total recording time would then be 40 min, which usually results in a complete precipitation reaction.

Note: Depending on your assay, shorter time intervals with less readings might be useful.

- Place the ArrayTube® in the reader slot. If required, adjust the tube's position with the live preview mode of IconoScan.

Required materials:	**Volume**
Enhancer Solution	35 µl
Initiator Solution	35 µl
0.2 x SSC	500 µl

- **Preparation of the Silver Enhancement Solution**

7.1 Pipette 35 µl of the enhancer and initiator solution into a separate microreaction tube.

7.2 Vortex the reagents, add 500 µl of 0.2 SSC and incubate the solution at 25°C for 20-40 min with 550 rpm.

Note:

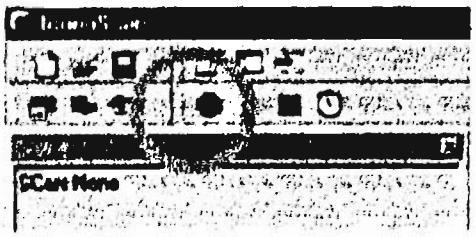
- Always prepare a fresh silver enhancer solution on ice.
- For temporary storing, keep the enhancer and initiator solution in a dark location at 2-8°C prior use.

- **Silver Enhancement**

Carefully add 70 µl of the freshly prepared silver enhancement solution into the AT. (Try to prevent air bubbles from forming within the solution. If bubbles do form, carefully try to remove them by repeated up- and down pipetting of the AT solution).

Next, begin the recording of the precipitation reaction (see also ATR 01 Manual, p. 13).

Click the record button on the upper menu bar of the image acquisition module IconoScan.



Depending on the assay conditions, a gray precipitation pattern will become visible after 5-10 min. The full development of the precipitation pattern usually needs 15 - 20 min, and ends with the saturation of the reactions after 30 - 40 min.

Note: Do not move or vibrate of the AT-Reader during recording!
Do not change the focus position anymore after starting the record process!

After finishing the image recording, pull the AT out of the reader. If you would like to store the ArrayTube®, follow with the post assay procedures described below.

8. Post Assay Procedure

The precipitation pattern on the AT array is quite stable after the post assay procedure have been performed and can still be imaged after a couple of months. The post assay procedure should be done shortly after recording.

Required materials	**Volume**
H_2O_{bidest} or 0.2 SSC	500 µl

8.1 Suction off the silver solution.
8.2 Wash the ArrayTube with 250 µl H_2O_{bidest} or SSC for 1 min two times.
8.3 Dry the tube in a Speed Vac for 15 min without vacuum.

Mutual Confidential Disclosure Agreement

DATED August 22, 2003

Between

DIAGNOTEC S.A.

and

CLONDIAG chip technologies GmbH

Lobstedter Strasse 103-105

07749 Jena/Germany

Whereas in furtherance of their business relationship each of the parties proposes to disclose to the other certain confidential and proprietary information in connection with mutually beneficial business opportunities; and

WHEREAS for their mutual protection **DIAGNOTEC** and **CLONDIAG** wish to set out the terms and conditions for the use and maintenance of the confidential information of the other party:

Now therefore, in consideration of the mutual covenants and agreements contained herein and intending to be legally bound, the parties agree as follows:

1. As used in this agreement, the term "Confidential Information" means any information concerning the current products, future products, business plans, marketing plans or research and development of either party, or any third party proprietary information given to either party, whether disclosed in written, oral or other media form, to the other party or its employees, but does not include information which:

    (i) is known to the receiving party before receipt thereof from the other party, as evidenced by the receiving party's records, and was not acquired, directly or indirectly, from the disclosing party; or

    (ii) is disclosed to the receiving party in good faith by a third party who had a right to make such disclosure; or

    (iii) is made public by the originating party, or is established to be a part of the public domain otherwise than as a consequence of a breach by the receiving party of its obligations hereunder.

2. Each party agrees that its obligations hereunder shall continue for so long as the parties maintain their business relationship, and for a period of five years thereafter; provided that this time limitation on the obligations of the parties shall not be deemed to reduce the term of copyright or other statutory protections.

3. Neither party will use or derive any benefit from or disclose to any person, firm or corporation, any part of the Confidential Information of the other, except with the prior written consent of the other.

4. Each party shall disclose and grant access to the Confidential Information of the other party only those of its bona fide, full-time employees who shall have a legitimate need to know the Confidential Information for the purposes specified in this Agreement. Each party shall employ the same safeguards to keep the Confidential Information of the other party confidential as it employs to safeguard its own trade secrets, including, without limitation, causing each of its employees with a need to know to enter into a general confidentiality agreement for the protection of the other party's Confidential Information herewith.

5. Each party shall use the Confidential Information of the other party strictly for the purposes specified in this agreement. Neither party shall use or allow the use of the Confidential Information of the other party for any purposes without the prior written consent of the other party.

6. Confidential Information shall not be copied by either party without the express written consent of the other, except for such copies as each party may reasonably require for its use pursuant to this Agreement.

7. Upon written request of the other party, each party shall return promptly to the requesting party all originals and copies of any and all Confidential Information which they have received.

8. The parties hereby acknowledge and expressly agree that the disclosure of Confidential Information without the express written consent of the disclosing party will cause irreparable harm to the disclosing party, and that any breach or threatened breach of this Agreement by the receiving party will entitle the disclosing party to injunctive relief, in addition to any other legal remedies available to it, in any court of competent jurisdiction.

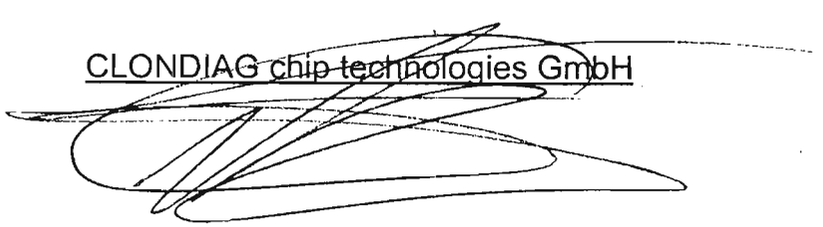
9. Nothing contained in this Agreement shall be construed as granting or conferring any rights by license or otherwise in any Confidential Information, except for the right to use the Confidential Information strictly in accordance with the provisions of this agreement.

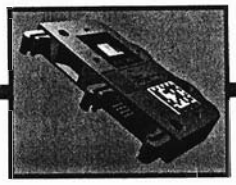
10. This agreement shall apply to each and every country in the world and shall be governed by and construed in accordance with the laws of the Province of Ontario, Canada, and shall benefit and be binding upon the respective successors and assigns of the parties hereto.

IN WITNESS HEREOF, the parties hereto have executed this Agreement as of the date first noted above.

DIAGNOTEC

CLONDIAG chip technologies GmbH





CLONDIAG®
chip technologies

AP System

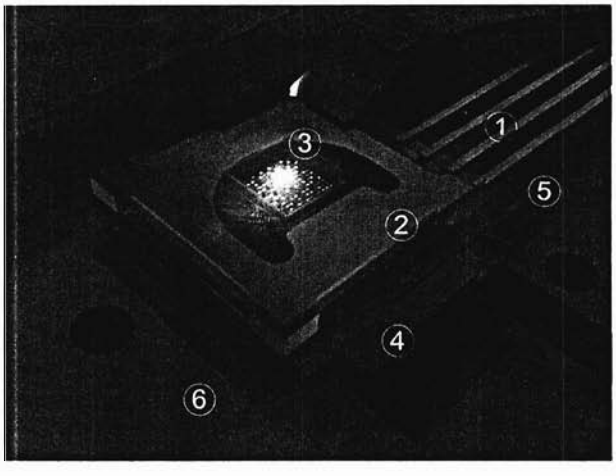
The Assay Processor AP
Integrated Platform for Genomic Diagnostics at the Point-of-Care

CLONDIAG®'s Assay Processor (AP) is a complete platform for array-based genotyping. Due to its robust performance and unprecedented simple structure, the AP has the potential to become the first test stripe format for use in pharma-associated diagnostics at the point-of-care.

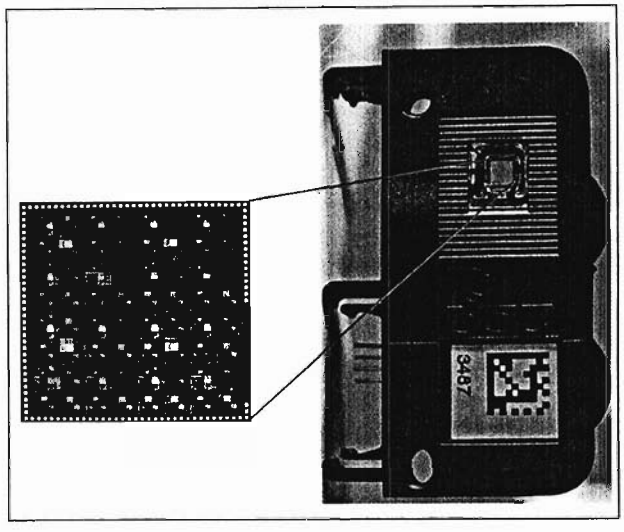
Device for a new type of genotyping assay: fast, robust and inexpensive

The performance of inexpensive, fast and robust genotyping assays is essential for modern diagnostic and therapeutic approaches. Proven the abilty to run genotyping assays at the point-of-care, personalized medicine will become a common standard. Systems for point-of-care genotyping will have to combine all steps necessary to perform a genetic analysis. With the AP-System, CLONDIAG® has developed the new key system for complete array-based genotyping.

Image of the AP principle setup with power lines (1), gasket (2), microarray (3), guidance for filling needle (4), heater & sensor chip (5) and cartridge base (6).



The Assay Processor - microreactor with array, integrated heater and sensor chip.

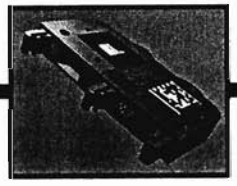


Integrated solution

The AP-System is based on single reactor assays combining DNA array analysis with microfluidics performance. We developed the AP to conduct sample preparation, target amplification and labeling, analysis of the target for specific markers and interpretation of raw data within a single system. Thus, the AP-System is a complete platform for performing complex diagnostic assays in a fast, accurate and inexpensive way.

Multifunctional reaction cartridge

Combining target amplification and array hybridization, which relies on precise temperature control, is feasible. To perform both processes, the AP functions as a miniaturized pressure cooker with a reaction volume of about 10 µl. The reactor consists of a chip with thin film temperature sensors and heaters and the probe array. A gasket serves as a septum for the delivery of the reagents and the sample material. Thus, amplification and detection occur simultaneously in a single reactor.



CLONDIAG® chip technologies

AP System

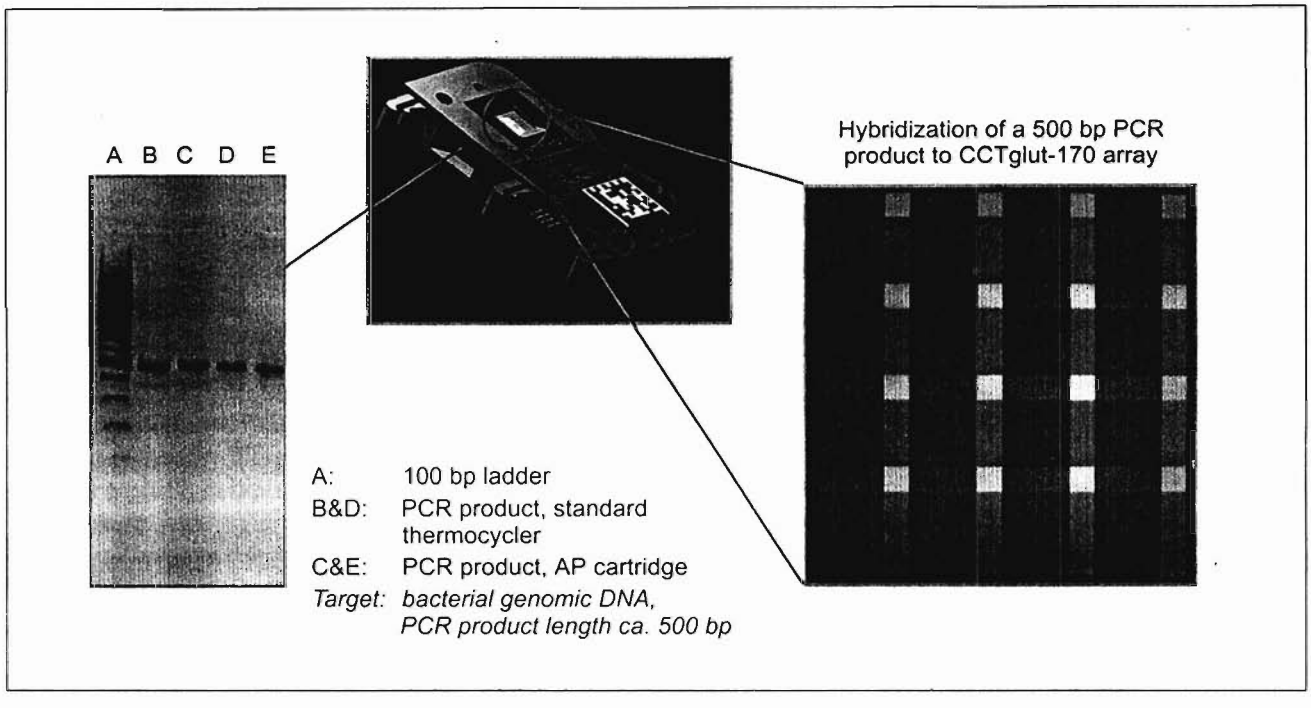
The probe array

A key unit of the AP cartridge is the integrated probe array which carries out a multitude of analytical reactions in parallel. For fabricating high quality arrays of oligonucleotides, we use in-situ synthesis with micro wet printing (µWP) technology on a four inch wafer scale. From this, hundreds of identical miniaturized arrays, each of them 3mm x 3mm in size with a density up to 10^6 probes/cm², are produced within a single manufacturing step.

Detection

In regard to assays, the system is optimized for fluorescence detection of the hybridized target. Due to the design of the reaction chamber, the AP bears the unique potential to detect the hybridization pattern online. This feature provides the opportunity to develop easy genotyping tests based on robust fluorescent protocols. Thus, the user no longer needs to perform numerous handling operations to obtain the results.

The AP-System - determination of species specific markers for point-of-care detection of infectious diseases: the amplification performance is strongly comparable to standard benchtop systems; the sequence specific determination of the amplified target is done by hybridization.



Data matrix

The integrated data matrix interfaces the AP with assay associated data. This data is accessible within the database and can be used by the management software PARTISAN arrayLIMS, permitting automatic control and processing of the whole assay process.

Computer controlled processes

The heating, cooling, filling and flushing of the AP are controlled by CLONDIAG®'s software tool IconoClust. A specific adapter connects the AP to the DNA reader or the dispensing systems. The AP injector module can handle small sample volumes of 15-40 µl. Dead volume effects are excluded.

CLONDIAG® chip technologies

AP System

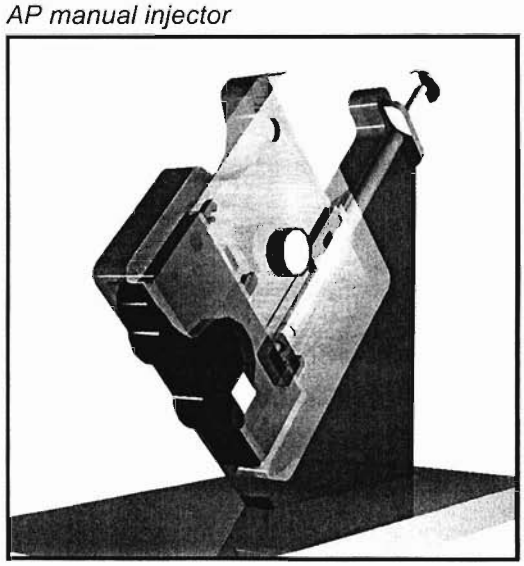
Modular design

Due to its modular design, the AP platform supports a multitude of applications. Different AP modules, like the AP injector for manual filling, can be easily plugged into the configuration, optimally fitting the user's needs. A specific module fixes the AP to the appropriate optical system for online-detection. The AP is designed to allow the integration of future devices, thus leading the way to the development of assays that meet the requirements of point-of-care diagnostics.

Fine temperature tuning

The thin film heater, temperature sensor and air channel allow tuning and rapid alterations of temperature, which is necessary for the fine modulation of the PCR and hybridziation reactions run in an interval of 0° to 100°C.

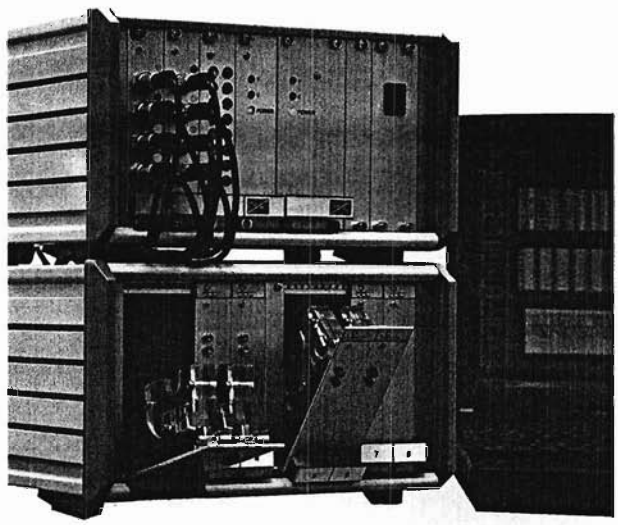
AP manual injector



Parallel assay optimization

The AP 8xController enables parallel independent processing of eight Assay Processors. Each AP unit is controlled separately, thus allowing for example the performance of eight different amplification protocols. Fast and easy assay optimization is possible.

AP 8xController for automated and independent AP processing.

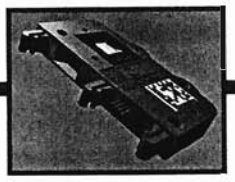


Automated image analysis

AP reader control, data readout and automated image analysis is feasible with CLONDIAG®'s software tool IconoClust. Its program structure and XML data exchange format permit easy integration with database systems like CLONDIAG®'s PARTISAN arrayLIMS.

Database connection

The AP-System is supported by an Oracle based laboratory information and management system, PARTISAN arrayLIMS. PARTISAN arrayLIMS provides management based on objects representing all data and processes of a bioarray life cycle. Thus, the entire assay is mapped in the database.



CLONDIAG® chip technologies

AP System

Future test stripe

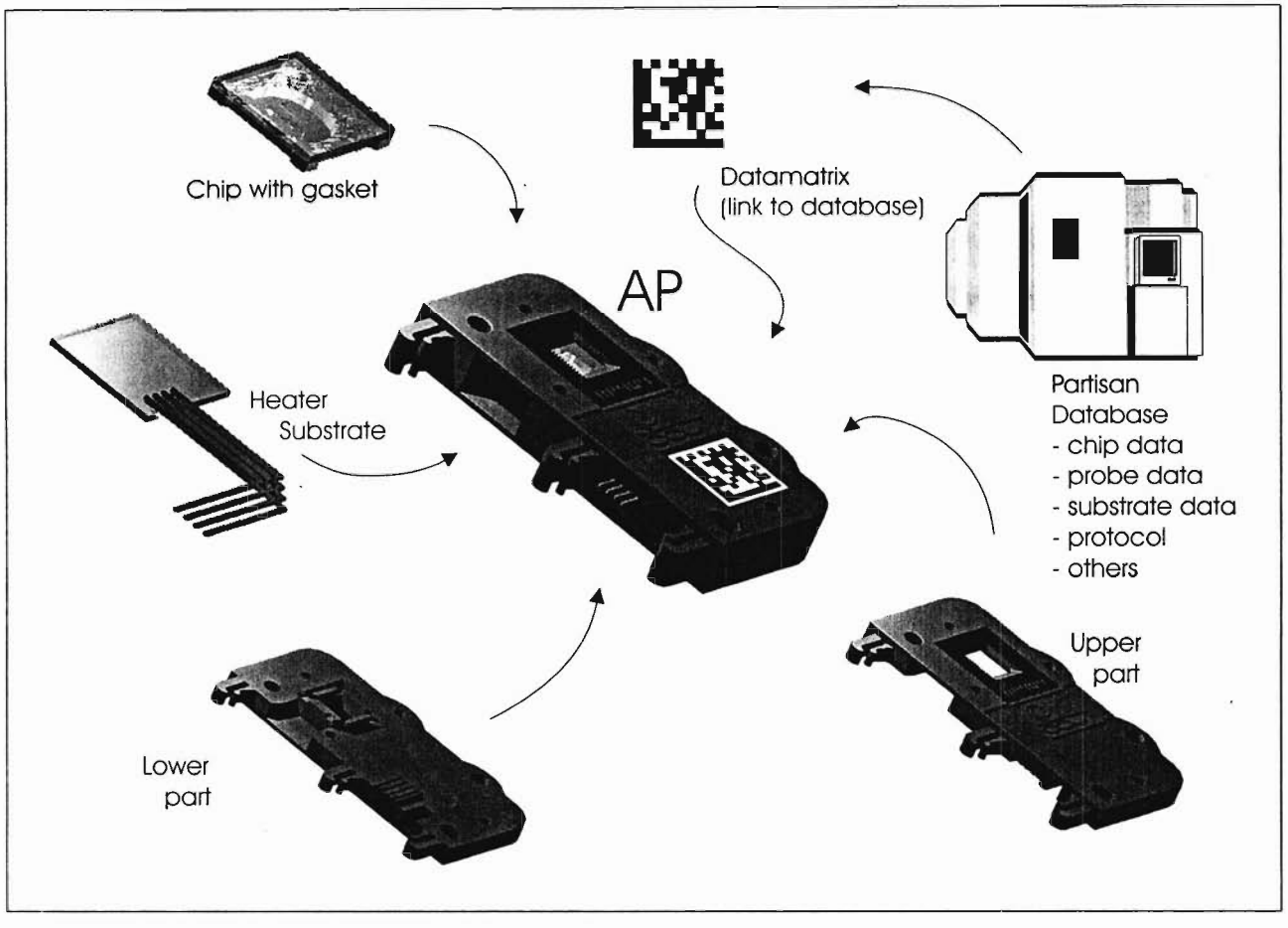
New approaches in point-of-care medical diagnostics, environmental testing and biotechnology require low cost, portable DNA-analysis instruments that perform the complete analysis within a single reactor. Since the Assay Processor satisfies these requirements, we believe that it has the unique potential to make genomics testing as easy as today's pregnancy tests.

AP access

The Assay Processor Platform is available for diagnostics applications development on a collaborative basis. For more information on AP applications development, please contact us:

clondiag@clondiag.com.

Components of the AP



CLONDIAG® chip technologies GmbH
Loebstedter Str. 103-105
07749 Jena, Germany
phone: +49/ 3641/ 59 47 0
Fax: +49/ 3641/ 59 47 20
email: clondiag@clondiag.com

CLONDIAG® chip technologies

AT-System: The Platform

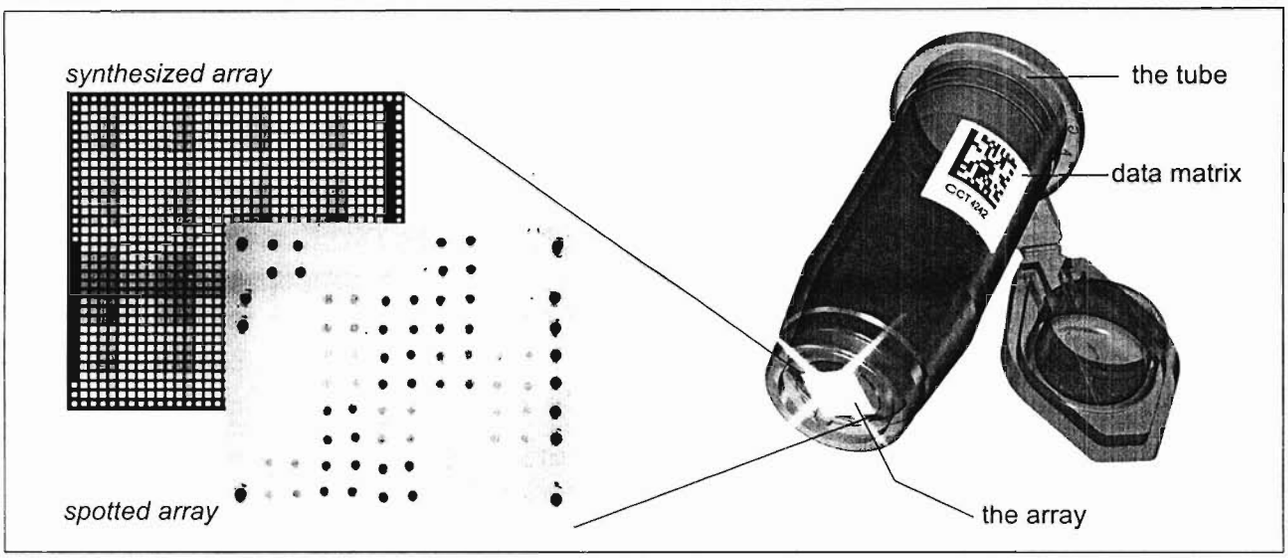
The ArrayTube® (AT) - Array Technology for Every Lab

The ArrayTube® - main features:

- New array-based platform for parallel lab-genomics
- Integration of microarray and microtube
- Easy handling
- Affordable for every lab
- Inexpensive, non-fluorescent detection

Array and reaction tube in one

CLONDIAG®'s ArrayTube® AT is a revolutionary new platform for performing easy and inexpensive experiments with microarrays. The direct implementation of high quality DNA arrays into a standard micro-reaction tube allows all hybridization and analysis procedures to be performed in an easily manageable and straightforward manner. No more highly specialized equipment is required.



*Patented **AT platform** with synthesized arrays (here: 1024 30mers / spot size 32µm / array size 2048 µm x 2048 µm / application: SNP analysis) or spotted arrays (here: array with 120 spotted oligonucleotides)*

Platform for every lab

The unique ArrayTube® design allows improved assay performance - array handling becomes much easier and more reproducible. Any solution can be easily pipetted into the AT, incubated and agitated with microtube equipment available in every lab.

High quality arrays

We develop and produce customized ArrayTubes® with both spotted and synthesized DNA arrays. The spotted arrays are fabricated by applying state-of-the-art spotting technologies and optimized substrate surfaces. For in-situ synthesizing of arrays, we employ our patented µ-wet printing technology (µWT) which results in high quality arrays of several thousand oligonucleotides.

Multiple advantages

Besides its lab compatibility and easy handling, the AT comprises many other advantages:

- Working with the AT platform reduces the amounts of all required reagents.
- The array integration within the reaction vial guarantees uniform wettability of the array during all processes.
- As a closed system, the AT ensures protection against evaporation and contamination effects.
- The specific AT design allows online detection of the signal amplification.
- With the AT platform array technology is now accessible to every lab.

CLONDIAG® chip technologies

AT-System: The Detection

Novel Robust AT Labeling Technology

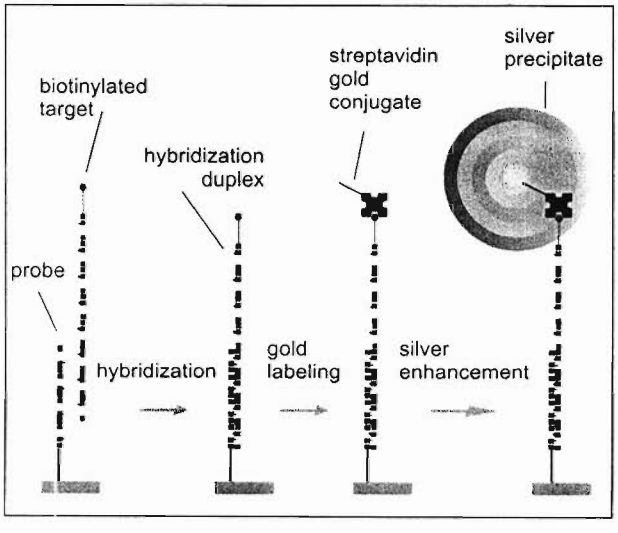
Principle

To provide a detection kit perfectly suited to our AT platform, we revived the method of precipitation staining for its application in array based analysis. Our new technology is based on gold induced silver precipitation, which directly correlates to the amount of specifically hybridized target molecules on the array. Analysis of the precipitation is done by simple transmission measurements, which leads to the effective reduction of the input in optical equipment.

Easy detection

The application of silver precipitates for hybridization detection alllows the use of a transmission imaging reader like our AT reader ATR 01, which is affordable for every lab. No highly sensitive optical equipment with cost intensive, high performance detectors is required.

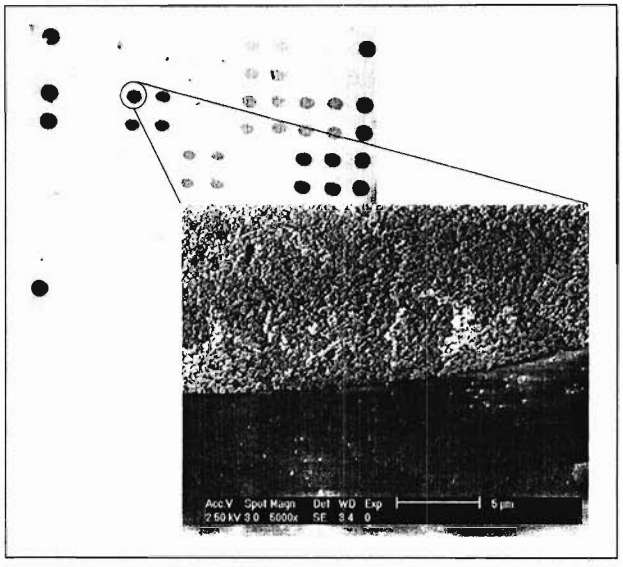
Principle of gold-silver labeling



Signal stability

As a result of detection via transmission measurements, problems such as light induced bleaching no longer arise. Stable output signals allowing comfortable data analysis are guaranteed.

Image of silver precipitate (by electron microscopy)



High sensitivity

Following hybridization, target staining is performed by applying gold-streptavidin conjugates. Gold particles catalyze the process of silver precipitation in the presence of silver solutions resulting in an enhanced detection performance. It is now possible to receive detection limits higher than those for conventional labeling methods.

High quality data

The amount of silver precipitate directly represents the target concentration. By applying the ATR 01 reader, it can be determined by endpoint detection or by monitoring precipitation dynamically over a period of time. Quantitative data analysis is feasible with the integrated reader software IconoClust-AT.

Non-hazardous dyes

All reagents required for the hybridization reaction using our AT labeling kit are non-toxic. Thus, easy and cost effective experimental handling is possible.

CLONDIAG® chip technologies

AT-System: The Reader

The Reader for Analyzing CLONDIAG® ArrayTubes®

Effective solution

Applying the ArrayTubes® in combination with colorimetric hybridization detection is the key to create affordable array reader systems that can be integrated into every lab. In analyzing the hybridization pattern by transmission measurements, cost intensive detector equipment, like confocal laser scanning systems, is not required.

The reader ATR 01 is an instrument that analyses CLONDIAG® ArrayTubes®. For reliable data acquisition, the assay temperature is kept constant during hybridization detection. Operation modes allow endpoint detection and dynamic data acquisition during the precipitation process.

ATR 01 reader for sensitive detection of AT hybridization patterns

AT Reader benefits

The system is easy to use. With its user-friendly graphics interface, working with the ArrayTubes® becomes a daily routine. The operating software IconoClust-AT is based on the successful IconoClust package. Being script based, it allows the user to customize the reporting scheme and data output, and provides instant calculation and interpretation. The system is flexible, reliable and affordable, making quality array technology accessible to any lab.

Specifications

Dimensions (W x D x H)	246 x 134 x 143 mm
Supply voltage	110-230 V AC
Pixel Resolution	752 x 582
Operating temperature	18° - 30°C (64,4 - 86°F)

Requirements

Desktop computer with 500MHz Pentium Processor, 256 MB RAM, 10GB HD;
Serial port for data matrix reader (optional).

CLONDIAG® chip technologies

AT-System: The Software

IconoClust-AT, the Tool for AT Data Acquisition and Analysis

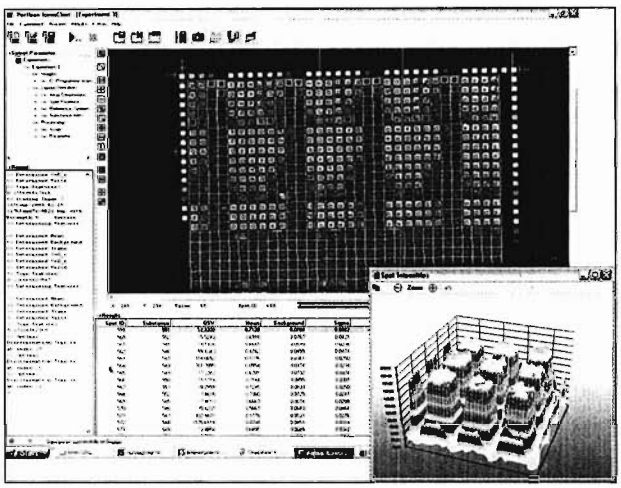
IconoClust-AT

Reader control, data readout and data analysis are performed with one software package, the IconoClust-AT. Based on our powerful stand-alone software IconoClust, the package comprises all features for automatic reader control and fast and easy image acquisition. Live previews and online data collection can be easily performed. A special submission algorithm allows immediate transfer of all image data into the module for analysis.
With the concept of experiment templates, the fast and automated image analysis of series of images for high throughput applications is easily accessible.

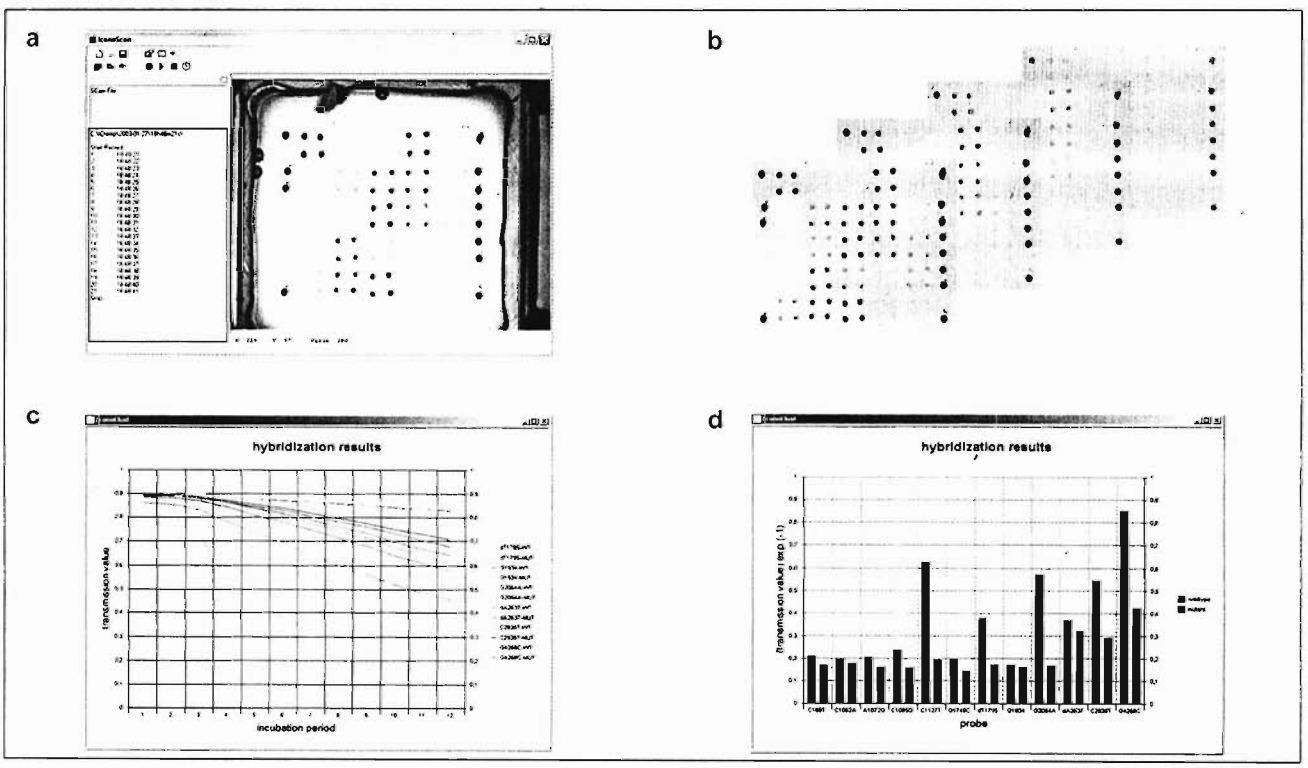
Database- and LIMS-Interface

IconoClust-AT can be optionally supported by our extensive laboratory information management system, PARTISAN arrayLIMS. Program structure and the XML data exchange format of IconoClust-AT permit its easy integration into our LIMS-system.

Image analysis of an in-situ synthesized array (1024 probe molecules) with IconoClust-AT. Related array probes and 3D details are visualized.



*Hybridization of multiplex PCR products against human cyp2D6-array (P450): **a)** IconoClust-AT screen plot of data acquisition, **b)** images of time series, **c)** plot of time series data and **d)** plot of analyzed hybridization results with IconoClust-AT.*



AT System: The Workflow

The Genomics Lab Analysis Work Platform

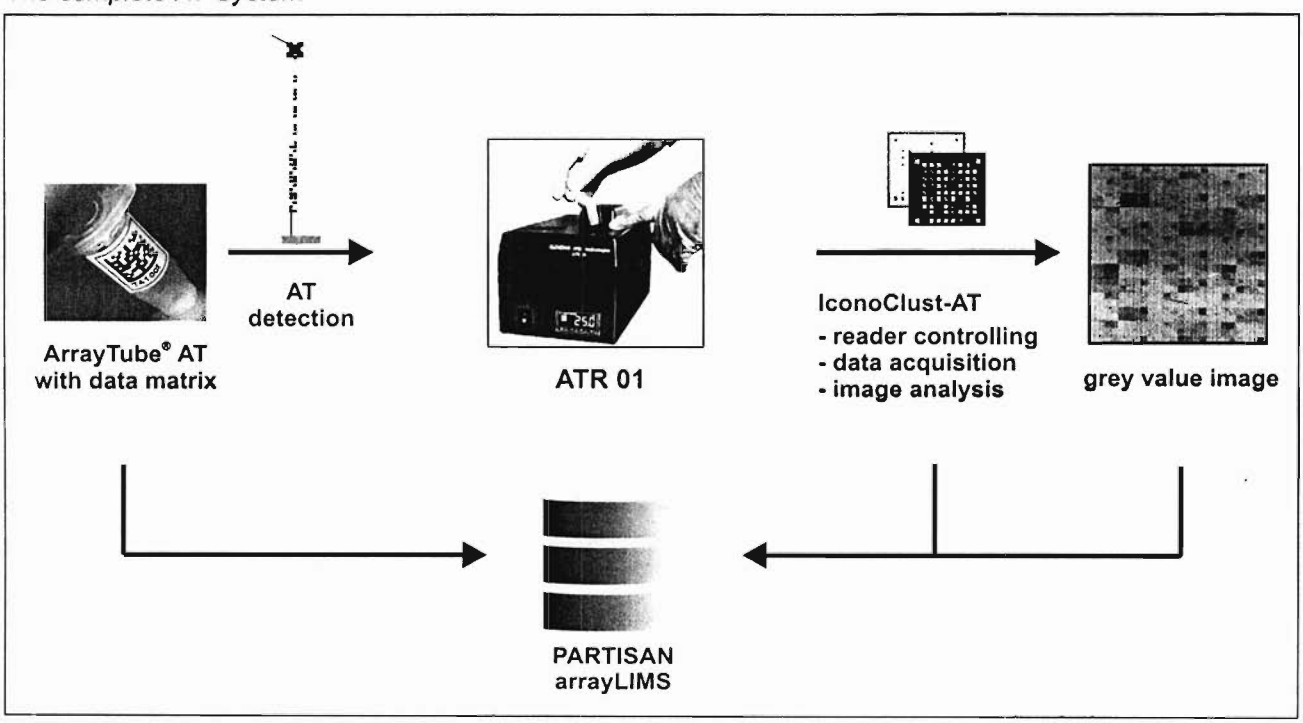
The system

CLONDIAG® has combined the benefits of all AT components into a powerful working station. The AT System comprises the new AT platform providing the array and the reaction vial as one system, the novel non-fluorescent labeling kit for easy AT detection and the smart and robust reader system ATR 01. By integrating our IconoClust-AT software package, optimal instrument control, data aquisition and custom specific image analysis are guaranteed .

Optionally, the AT System is supported by our powerful laboratory information management system PARTISAN arrayLIMS, which provides reliable data and experiment management. All ArrayTubes® include a unique data matrix accessible within the database allowing correct AT and data tracking.

With the AT System, CLONDIAG® provides an affordable work station for superior array-based analysis results, bringing the world of array technologies and multiplex assays to any lab.

The complete AT System



CLONDIAG® chip technologies GmbH
Loebstedter Str. 103-105
07749 Jena, Germany
phone: +49/ 3641/ 59 47 0
fax: +49/ 3641/ 59 47 20
email: clondiag@clondiag.com

CLONDIAG® chip technologies

AT-System: Solutions

Products and Services

Your AT starter package

We provide the AT System as the new working platform for easy and affordable array-based analysis. A typical starter package includes the reader ATR 01 and a starter kit of ArrayTubes® designed to customer needs with comprehensive user instructions. Subsequent ArrayTube® purchases will be provided with identical or modified array designs.

Services

The services we provide around the ArrayTube® System include probe and array design. The complete ArrayTube® package comes with probes generated by us. On demand, we also apply probes delivered by the customer for producing the AT arrays. Additionally, we provide assay development and data analysis.

AT Reader Solutions	**Specification**
ATR 01 Standard Package:	ATR 01 reader with PC provided with framegrabber card and preinstalled IconoClust-AT software for reader control and image analysis.
ATR 01 Modular Package:	AT reader with IconoClust-AT software for reader contol and image analysis, and with framegrabber card (for installing in your own PC)
AT PARTISAN arrayLIMS:	System for efficient organization and management of AT experimental data, for 5 users, preinstalled.
ArrayTube® Kits	**Specification**
ArrayTube® Starterkit[1]:	Set of 100 (50^2) ArrayTubes®
Additional ArrayTubes®:	Packs of 100 (50^2) ArrayTubes®

[1] ArrayTubes® specified by the customer. Standard AT arrays contain up to 144 spotted features. Probes can be provided by CLONDIAG® or by the customer. For synthesized arrays with up to 4096 features please contact us.

[2] Sets of 50 ArrayTubes® can be delivered, if the probes are provided by the customer.

For inquiries and further information, please contact us.
clondiag@clondiag.com
www.clondiag.com

CLONDIAG® chip technologies

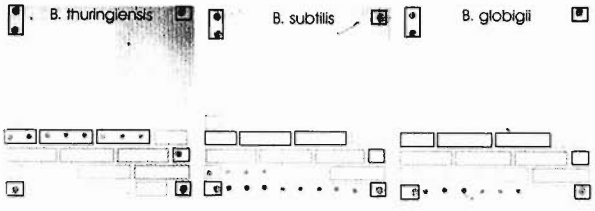
Application Fields of the ArrayTube® System

The ArrayTube® (AT) System can be used for multiple applications in array-based analysis. Its unique features make it exceptionally suitable for routine genotyping and mutation analysis, as well as for applications in expression and proteomics analysis.

Depending on the specific application, ArrayTubes® with spotted arrays (up to 144 features, various substance classes possible) or in-situ synthesized arrays (up to 4096 oligonucleotide probes) can be provided.

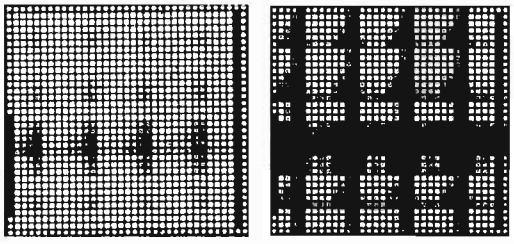
Examples of AT-Applications:

- **Genotyping of Bacterial Resistance Genes**
  The ArrayTube® System was used to setup a genotyping assay for the fast and reliable analysis of different resistance genes of the germ *Staphylococcus aureus*.
- **Mutation Analysis of the Human *cyp2D6* Gene**
  Using AT-arrays with probes containing different mutations of the *cyp2D6* gene, clinical samples were screened for *cyp2D6* mutations.
- **Species Discrimination & Identification**
  With a set of specific ArrayTubes®, different bacteria were identified by hybridization against AT-arrays carrying selected 16S rDNA sequences.
- **Genotyping of Human HVR2 mtDNA**
  A 4^5 sequence space of the HVR2 region of human mtDNA was analyzed. AT-arrays carrying a set of 4096 combinatorial oligonucleotide probes were applied.
- **Analysis of Thrombogenetic Mutations (OGHAM Diagnostics GmbH)**
  The AT-System is used for the mutation analysis of 8 different loci providing risk factors for thrombosis.



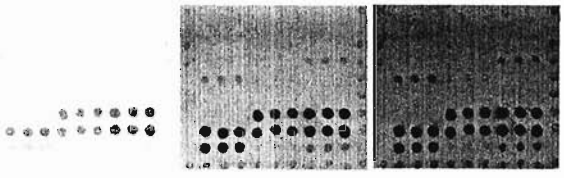
Identification of different bacteria with spotted AT-arrays (different colored frames represent bacteria specific probe molecules).

- **Mutation Analysis of Atherosclerosis associated Mutations (OGHAM Diagnostics GmbH)**
  The ArrayTube® Platform can be applied for the determination of genetic variants associated with atherosclerosis.



AT-arrays with in-situ synthesized oligo-nucleotides for optimal probe design.

- **Protein Assays**
  AT-arrays with spotted proteins can be produced for the fast detection of protein-target interactions with the ArrayTube® System.
- **Identification of Respiratory Bacteria and Viruses (Institut Dr. Viollier)**
  The fast and differential diagnosis of different respiratory bacteria and viruses was proved with the ArrayTube® System.
- **Expression Profiling**
  The AT-System can be applied for the expression analysis of selected genes.



Detection of specific protein-antibody interactions with AT-arrays (here: phosphorylated proteins).

For inquiries and further information, please contact us at clondiag@clondiag.com or www.clondiag.com.

CLONDIAG®
chip technologies

AT Application: Genotyping of a 4^5-Sequence Space of Human HVR2 mtDNA Region

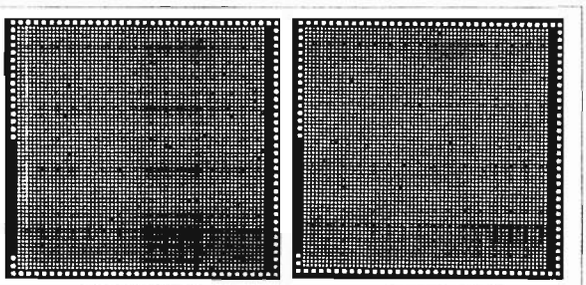
Detection and analysis

The detection of the characteristic sample hybridizations was performed with the ArrayTube® reader ATR 01. Time series of the specific silver precipitations representing the hybridization images were collected and analyzed with the software IconoClust-AT.

Results

Different hybridization patterns of the amplified sample fragments were detected. The results were compared and aligned with sequence data from public mtDNA databases. With regard to the five selected mutation sites, the detected sequence variations could be correlated to distinct geographical population origins (database MOUSE II[1]).

[1] *MOUSE II:*
Mitchondrial and Other Useful SEquences, LMU Munich, www.gen-epi.de/mouse/index.html



Characteristic hybridization patterns of two different samples.

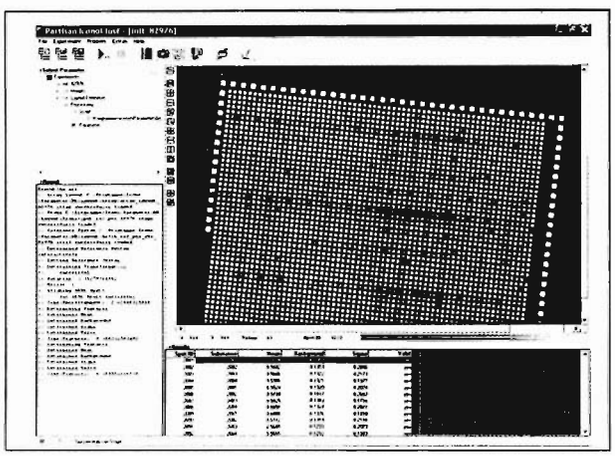
example 1:
origin: Africa
detected sequence:

CCCCATCCCATTATTTATCGCACCTAC

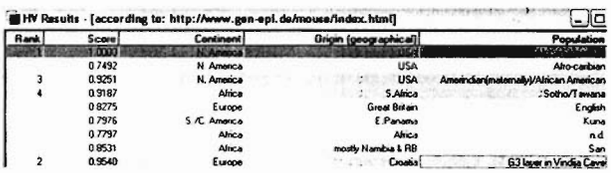
example 2:
origin: Europe
detected sequence

CCTCATCCTATTATTTATCGCACCTAC

Image analysis of 4096-spot AT-arrays with IconoClust-AT.



Comparison of the hybridization results to sequence data of MOUSE II and correlation to population origins.



Conclusion:

The ArrayTube® (AT) System is most suitable for the performance of fast and reliable genotyping assays using highly integrated microarrays. It opens up the possibility to extend the application of array technology to fields like forensic medicine.

References:

- Anderson, S. et al. (1981); Nature 290, p. 457-465
- Meyer. S., Weiss, G., von Haseeler, A.(1999); Genetics 152, p. 1103-1110
- Stoneking. M. (2000); Am.J.Hum.Genet. 67, p. 1029-1031

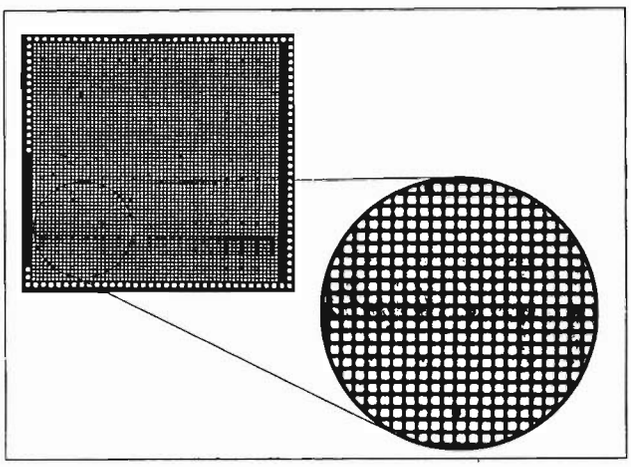
AT Application: Genotyping of a 4^5-Sequence Space of Human HVR2 mtDNA Region

Introduction

Sequences of the noncoding region of human mitochondrial DNA are potent tools to infer aspects of genetic population history. They show specific characteristics like high evolutionary rates and the maternal mode of inheritance. Especially the hypervariable sites HVR1 and HVR2 in the noncoding or D-loop region of mtDNA can be used to differentiate between non related species: HVR1 and HVR2 contain mutational hot spots with one polymorphism in 200 - 400 generations [1], [2], [3].

The analysis of mtDNA with conventional methods like sequencing, PCR, or restriction fragment length polymorphism is time-consuming and cost-intensive. Here, the ArrayTube® (AT) System was applied for the first time to set up a fast, affordable and reliable genotyping assay for a set of mutations within the HVR2 region.

AT-array carrying 4096 oligonucleotides; spot size 20 µm 20 µm, arrays size 2048 µm x 2048 µm.



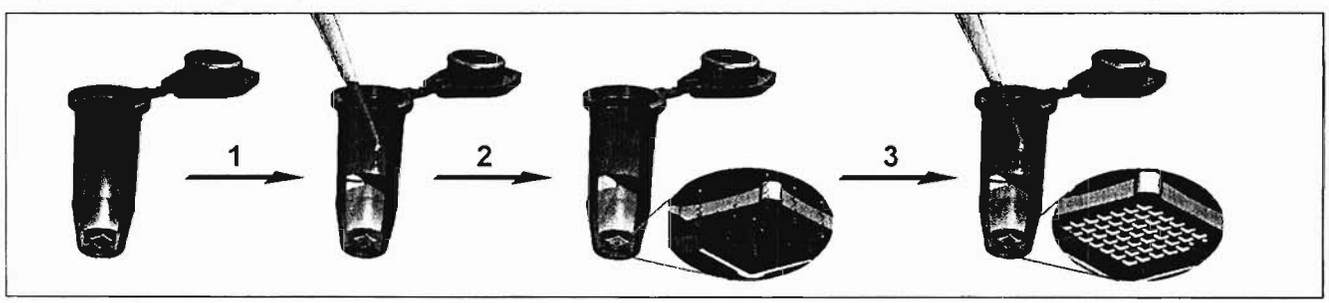
Methods

Microarray preparation

After screening public gene databases for the relevant sequences, five distinct mutation sites within the hot spot areas of the HVR2 region were selected for the genotyping analysis. ArrayTubes® with oligonucleotide arrays covering the complete sequence space of all 4^5 = 1024 variants of the five mutation sites were manufactured. The production of the 4096 probe arrays (each probe 4 times redundant) was performed by CLONDIAG®'s in situ synthesis technology (see 'technologies' in http://www.clondiag.com).

Sample preparation and hybridization:

Capillary blood samples (5-10 µl) were collected. After lysis, the complete HVR2 region was amplified using 5'-biotinylated primers for labeling. Biotinylated PCR fragments were diluted 1:20 in AT-hybridization buffer and pipetted into a pre-conditioned ArrayTube® containing the HVR2 microarray. Hybridization reaction was performed in a conventional thermomixer. After washing, the AT-conjugation and AT-silver enhancement reaction were performed according to standard ArrayTube® labeling protocols (http://www.clondiag.com).



ArrayTube® processing: After pre-conditioning the AT-array, the biotinylated sample is pipetted (1) into the ArrayTube® for hybridization (2). Detection of the hybridization pattern is performed by the silver enhancement reaction (3).

CLONDIAG® chip technologies

AT Application Sample: Genotyping of the Human *cyp2D6* Gene

Background

The human *cyp2D6* gene is coding for the Cyp2D6 enzyme, a member of the Cytochrom P450 family. The enzyme is involved in the metabolism of drugs like analgesics, beta-blockers, antidepressants and others. It was found to be responsible for more than 70 different drug oxidations. Mutations in the *cyp2D6* gene can cause severe adverse drug reactions like toxicity, inefficacy and hypersensitivity. In addition, Cyp2D6 was also found to be involved in severe disorders like Parkinson's disease and cancer. Up until now at least 18 *cyp2D6* mutations are known, which differ highly in their impact on the Cytochrom P450 function.

Experiments

DNA containing the related mutations was prepared by asymmetric multiplex PCR using chromosomal DNA of clinical samples as a template. For target DNA labeling, biotinylated primers were used.

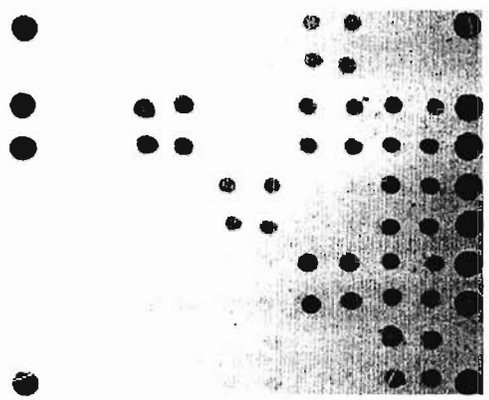
Analysis of multiplex PCR products for different clinical samples with standard agarose-gel-electrophoresis (1-3) (N=negative control).

Hybridization against a 100 probe AT array was performed by inserting the AT into a conventional microtube thermomixer. Hybridization was started by pipetting aliquots of PCR product in an appropriate buffer into the AT. After subsequent washing steps, a streptavidin-gold-conjugate was added. Detection of the hybridized target molecules was achieved by adding a silver enhancement solution: catalyzed by the gold particles attached to the hybridized target molecules, silver precipitation was initiated. The AT was inserted into the AT reader ATR 01, which collected the hybridization images as a time series. The reader is provided with heating and cooling elements guaranteeing constant hybridization conditions.

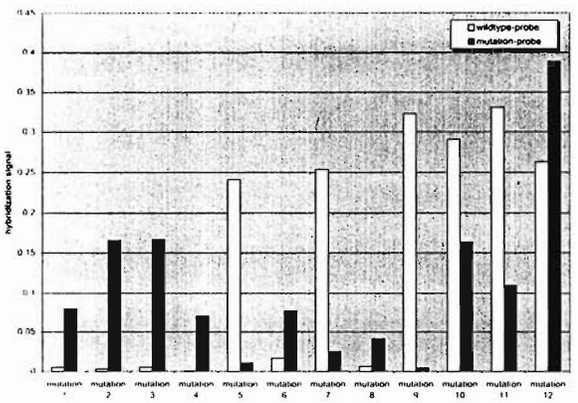
Data acquisition and image analysis were done by an IconoClust-AT software package.

Results

Hybridization pattern of a multiplex PCR product of a clinical sample carrying several mutations. All probes were immobilized 4 times redundant. Marker molecules and a control sequence were added.

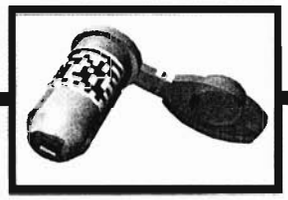


Bar graph of the hybridization pattern received with the multiplex-PCR product of a clinical sample carrying several mutations.



Conclusion

CLONDIAG®'s AT Platform is an optimal tool for genotyping *cyp2D6* because it allows fast and reliable screenings of clinical samples for mutations.



CLONDIAG® chip technologies

AT Application Sample: Fast Genotyping of *Staphylococcus aureus* Resistance Behavior

Results

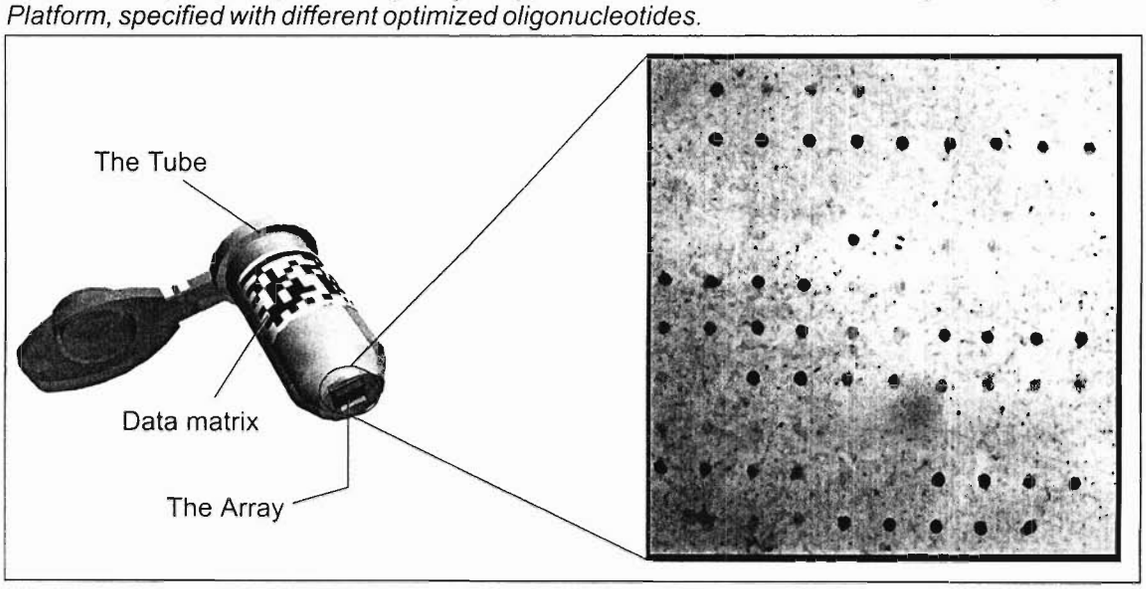
The results of growth inhibition assays and hybridization experiments of one clinical sample isolate are presented. Table 1 summarizes the resistance pattern as determined by both the phenotypical and genotypical characterization methods.

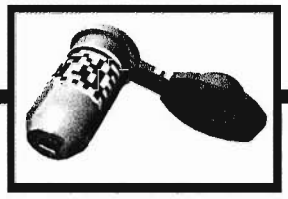
Both methods resulted in well corresponding data. AT based diagnostics have the advantage of being very fast, inexpensive and robust in handling. Compared to classical phenotypical characterization, AT based diagnostics can provide a high amount of clinically and biologically relevant data in parallel within a short period of time. Arrays with modified or additional probe molecules can be generated easily, thus making the AT Platform a highly flexible tool. The AT Platform is best suited for genotyping studies as described here.

Table 1:
Comparison of the conventional growth inhibition assays of S.aureus and the genotyping analysis applying the AT Platform.

virulence factor or resistance	Isolate 2000 V 8510 (epidemic strain "Hannover") phenotype	detected gene
Coagulase	positive	*coA*-positive
Penicillin	resistant	*blaZ*-positive
		mecA-positive
Oxacillin + all beta lactams	resistant	*mecA*-positive
Erythromycin	resistant	*ermA*-positive
Clindamycin		*ermC*-negative
		linA-negative
Levofloxacin	resistant	*norA*-positive
Aminoglycosides	resistant	*aacA/aphD*-pos.
		aphA-3-positive
Tetracyclines	resistant	*tetM*-positive
		tetK-negative
Toxic Shock Syndrome Toxin	n. available	*Tst1*-negative

Hybridization pattern representing the genotypical characterization of resistant genes using the AT Platform, specified with different optimized oligonucleotides.





AT Application Sample: Fast Genotyping of *Staphylococcus aureus* Resistance Behavior

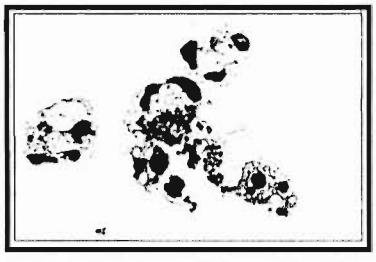
Background

Staphylococcus aureus is a common germ that causes wound infections, food poisoning, toxic shock syndrome and other effects. The appearance of distinct *S.aureus* populations showing resistance to many antibiotics is of increasing significance to hospitals because they complicate patient treatment. Choosing the appropriate treatment depends on the rapid and accurate identification of the infecting *S.aureus* strain and its specific antibiotic resistance.

As of now, about 30 *S.aureus* genes that are resistant and their complete sequence information have been identified. Screening for these genes using conventional methods like PCR is extremely time-consuming and cost-intensive. Therefore, these screenings cannot be performed routinely with clinical specimens.

CLONDIAG®'s **AT Platform** provides a solution that allows fast and reliable array-based genotyping and SNP analysis, which, for example, can be used to determine the antibiotic resistance patterns of bacterial pathogens based on their genes.

Staphylococcus aureus cells in pus

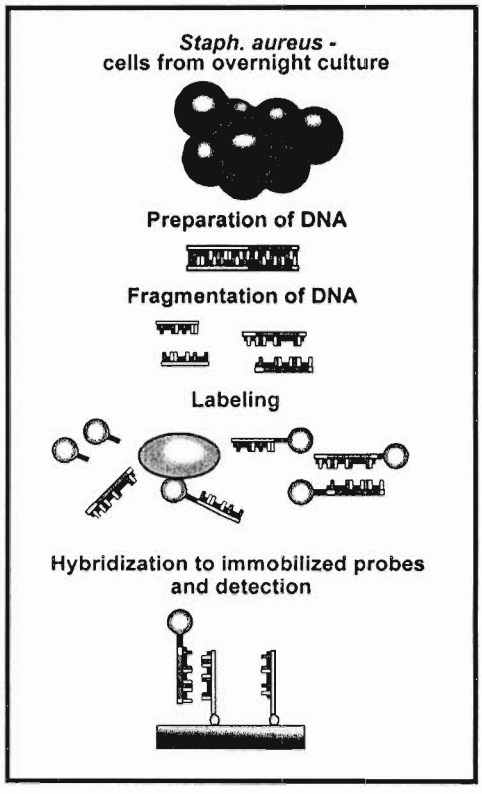


Experiments

In order to generate the appropriate AT array used in the genotyping analysis, oligonucleotide probes complementary to the resistance genes shown in table 1 were determined using CLONDIAG® software tools. Probe molecules were synthesized by standard procedures and immobilized by needle-based spotting techniques and covalent coupling to the substrate surface. For the hybridization experiments with the AT Platform, genomic DNA from clinical isolates of *S.aureus* was prepared.

An amount of 5 µg of isolated genomic DNA was labeled by applying a method developed by CLONDIAG® resulting in fragmented and biotinylated DNA that is ready for the hybridization procedures. Hybridization against the array was performed in the AT inserted into a conventional microtube thermomixer. Detection was done with the AT reader ATR 01. As a control, the antibiotic resistance patterns of the isolates were characterized phenotypically based on growth inhibition assays using the VITEK 1 system (BioMerieux).

Scheme of genotyping experiments with the AT Platform.



Review TRENDS in Genetics V

Fundamentals of cDNA microarray data analysis

Yuk Fai Leung and Duccio Cavalieri

Bauer Center For Genomics Research, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138, USA

Microarray technology is a powerful approach for genomics research. The multi-step, data-intensive nature of this technology has created an unprecedented informatics and analytical challenge. It is important to understand the crucial steps that can affect the outcome of the analysis. In this review, we provide an overview of the contemporary trend on various main analysis steps in the microarray data analysis process, which includes experimental design, data standardization, image acquisition and analysis, normalization, statistical significance inference, exploratory data analysis, class prediction and pathway analysis, as well as various considerations relevant to their implementation.

The development of microarray technology has been phenomenal in the past few years. It has become a standard tool in many genomics research laboratories. The reason for this popularity is that microarrays have revolutionized the approach to biological research. Instead of working on a gene-by-gene basis, scientists can now study tens of thousands of genes at once. Unfortunately, they are often daunted and confused by the complexity of data analyses. Although it is advisable to collaborate with statisticians and mathematicians on performing a proper data analysis, it is crucial to understand the fundamentals of data analysis. In this review, we explain these fundamentals step-by-step (Figure 1; Table 1). Instead of discussing any particular analysis software, we focus primarily on the rationale behind the analysis processes and the key factors that affect the quality of the result. For a compilation of current microarray analysis software see a recent article [1] and author's website (http://ihome.cuhk.edu.hk/~b400559/arraysoft.html; permanent link: http://genomicshome.com). We also focus on the use of the two-dye cDNA microarray data analysis, although most of our discussions are also applicable to the single-dye oligonucleotide platform (i.e. Affymetrix) (Box 1). We hope that by appreciating the fundamentals novices will become successful at microarray data analysis.

Experimental design and implementation

> 'If the experimental design is wisely chosen, a great deal of information is readily extractable, and no elaborate analysis might be necessary. In fact, in many happy situations all the important conclusions are evident from visual examination of the data'. [2]

'Well begun is half done', is an aphorism that is especially true of for microarray experiments. Good design is very important at the beginning of a microarray experiment. A typical microarray usually consists of tens of thousands of elements. On the one hand, it provides a comprehensive coverage that almost always promises some new discoveries. On the other hand, analyzing the vast amount of data being generated can be daunting to scientists. It is therefore, more important now than ever, to design a microarray project carefully to generate high-quality data and to maximize the efficiency of data analysis.

Good microarray experimental design should comprise at least four elements: (i) a clearly defined biological question and/or hypothesis; (ii) treatment, perturbation and observation of the biological materials, as well as the microarray experimental protocols, should be as little affected by systematic and experimental errors as possible; (iii) a simple, sensible and statistically sound microarray experimental arrangement that will give the maximal amount of information given the cost structure and complexity of the study [3–5]; and (iv) compliance with the standard of microarray information collection, which will be further discussed in the next section.

Standardization of information generated by microarray experimentation

The adoption of international standards have long been seen as vital in science because of the confusion generated through the use of various units. We have been experiencing a similar issue in the microarray field. The same increase or decrease in gene expression observed by two different laboratories might actually be different, especially when they are using different experimental protocols and data-analysis methods. Without a standard, it is almost impossible to judge the validity of a result just by inspecting the expression changes or even the raw data [6]. In view of this problem, the Microarray Gene Expression Data (MGED) Society (http://www.mged.org), an international initiative to develop standards for microarray data, has recently proposed a standard Minimum Information About a Microarray Experiment (MIAME) (http://www.mged.org/Workgroups/MIAME/miame.html) [7]. The research community has embraced it and many major journals now require compliance with MIAME for any new submission [8]. It is therefore advisable to ensure that the experimental design, implementation and data analysis comply with the MIAME standard

Corresponding author: Yuk Fai Leung (yfleung@cgr.harvard.edu).

Glossary

Adaptive circle segmentation: a segmentation process in which the diameter of the circle being applied to the spot is calculated case by case in order to address the variation of spot diameter. The pixels that fall within the circle are regarded as foreground.

Background estimation: the background fluorescence signal usually originates from non-specific hybridization of the labeled samples or auto-fluorescence of the glass slide. This unwanted background signal needs to be estimated and removed from foreground signal during image analysis.

Background intensity subtraction: the calculation of fluorescence signal from the background pixels of a spot identified during the segmentation process. Usually the median of the pixel intensities is used.

Dye-swapping experiment: two hybridizations of the sample pair of samples in which the labeling dye of the two samples is reversed in one of hybridizations. Averaging the two expression ratios would give one a good estimate of the true ratio.

Fixed circle segmentation: a segmentation process in which a circle with a constant diameter is applied to all spots on the image. The pixels that fall within the circle are regarded as foreground.

Intensity extraction: the process that calculates the foreground (signal) and background intensities from the pixels after the segmentation process.

Local background estimation: a commonly used background estimation method in which the immediate background pixels surrounding the spot, as identified by the segmentation process, are used for estimating the background signal.

Segmentation: a computational process which differentiates the pixels within a spot-containing region into foreground (true signal) and background.

Spot intensity extraction: the calculation of fluorescence signal from the foreground pixels of a spot identified during the segmentation process. Usually the mean of the pixel intensities is used.

Spot recognition or gridding: a computational process which locates each spot on the microarray image.

MIAME represents the minimal information to be recorded that enables faithful experimental replication, the verification of the validity of the reported result, and the facilitation of the comparison among similar experiments. Besides, the information should be structured with controlled vocabularies and ontology to assist in developing database and automated data analysis. Currently, the minimal information includes the six parts: (i) experimental design; (ii) array design; (iii) samples; (iv) hybridizations; (v) measurements; and (vi) normalization controls. A detailed description of each part and a convenient checklist are available on the MIAME website (http://www.mged.org/Workgroups/MIAME/miame_checklist.html).

Image acquisition and analysis

After performing all biological and hybridization experiments, the first step of data analysis is scanning the slide and extracting the raw intensity data from the images. There are four basic steps in image acquisition and analysis: (i) scanning; (ii) SPOT RECOGNITION OR GRIDDING (see Glossary); (iii) SEGMENTATION; and (iv) INTENSITY EXTRACTION and ratio calculation.

Image acquisition is a very important step in data analysis. Once an image has been scanned, all data, high or poor-quality, are essentially fixed. A poor-quality image requires further manipulations, which will lead to a decrease in the power of analysis. There are two prerequisites for obtaining a high-quality image. First, all steps in array construction, RNA extraction, labeling, and array hybridization have to be performed to the highest possible standards. These endeavors ensure that all images would be least affected by contamination (e.g. dust or dirt), and have consistent spots with high signal-to-noise ratios. Second, the choice of scanning parameters is also important. We discuss the settings for the Axon scanner, but the general principle is applicable to other platforms. A low laser power (30%) should be used whenever possible to prevent photo-bleaching. The photomultiplier tube (PMT) gain settings are adjusted during the scanning process to balance the overall intensities between the two channels (i.e. cy3 and cy5) as much as possible. This balance can be evaluated in several ways: (i) visual inspection of the scanning image. The non-differentially expressed spots should appear

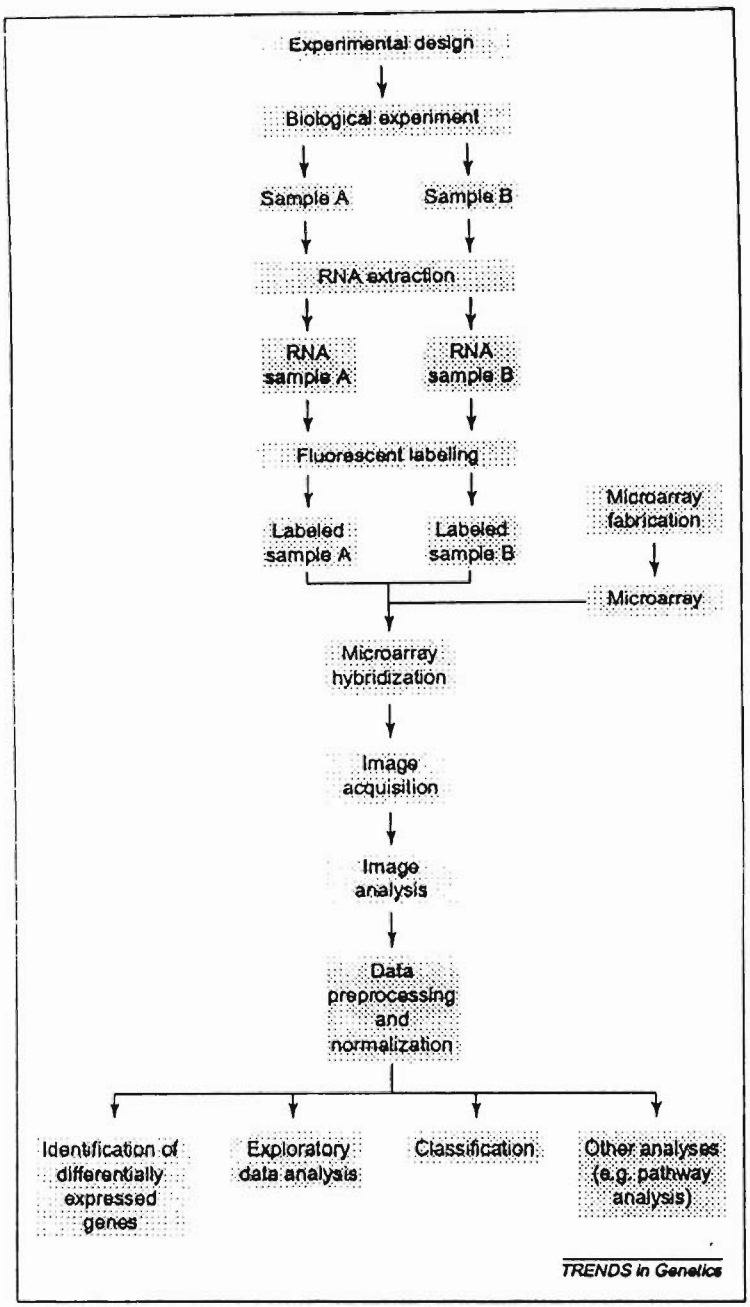


Figure 1. Flow of a typical microarray experiment. A typical microarray experiment begins with good experimental design. After carrying out the biological experiment, the samples, either tissues from patient or animal model, or cells from *in vitro* cultures, are collected. Their RNAs are then extracted and labeled with different fluorescent dyes, and co-hybridized to a microarray. The hybridized microarray is scanned to acquire the fluorescent images. Image analysis is performed to obtain the raw signal data for every spot. Poor quality data are filtered out and the remaining high quality data are normalized. Finally depending on the aim of the study, one can infer statistical significance of differential expression, perform various exploratory data analyses, classify samples according to their disease subtypes and carry out pathway analysis. Note that data from all the steps should be collected according to certain standards, minimum information about a microarray experiment (e.g. MIAME), and archived properly.

Table 1. Summary of microarray analysis steps[a]

Analysis step	Caveats
Experimental design and implementation	Define the biological question and hypothesis clearly
	Design the microarray experimental scheme carefully; include biological replication in experimental design
	Avoid experimental errors
Data collection and archival	Compliance with microarray information collection standards (e.g. MIAME)
Image acquisition	Avoid photo-bleaching
	Try to balance the overall intensities between the two dyes
	Scan image at appropriate resolution
Image analysis	Inspect the gridding result manually; adjust the mask and flag poor-quality spots if necessary
	Choose and apply an appropriate segmentation algorithm
	Apply quality measures to aid decision of spot quality
Data pre-processing	Remove poor-quality spots
	Remove spots with intensity lower the background plus two standard deviations.
	Log-transform the intensity ratios
Data normalization	Use diagnostic plots to evaluate the data
	Consider using LOWESS and its variants for normalization
Identifying differentially expressed genes	Do not use fixed threshold (i.e. two-fold increase or decrease) to infer significance
	Calculate a statistic based on replicate array data for ranking genes
	Select a cut-off value for rejecting the null-hypothesis that a gene is not differentially expressed; remember to adjust for multiple hypothesis testing
Exploratory data analysis	Use different analysis tools with different setting to 'explore' the data
	Validate the result by follow-up experiments
Class prediction and classification	Do not over-train the classifier; try to balance the accuracy and generalizability
Pathway analysis	Try to understand the microarray data in a pathway perspective and not genes in isolation

[a]Abbreviations: LOWESS, locally weighed scatterplot smoothing; MIAME, minimum information about a microarray experiment.

yellow (i.e. ratio equals to 1) on a balanced image (Figure 2a). In many cases, most of the spots on the array are non-differentially expressed; (ii) examining the extent of overlap between the pixel distribution histograms of both channels (Figure 2b); and (iii) computation of the global normalization factor for all the spots contained in the two channels, for example the sum of signals in one channel divided by the sum of signals in the other one. A well-balanced image should have a factor close to 1.

The choice of a suitable scanning resolution depends on the array specification. A rule of thumb is that the resolution setting should be at least 10% of the spot diameter. At the same time, the number of spots with saturated pixels should be kept to a minimum (e.g. <3–5 spots in a whole yeast genome array with 6240 elements) to maximize the dynamic range usage of the scanner. Excessive scanning of a slide should be avoided to prevent photo-bleaching. Images of high-quality can be acquired routinely when all these factors are taken into consideration (Figure 2a).

Spot recognition or gridding is not a difficult problem for most contemporary image analysis software, although it is often necessary to adjust the grid for some spots manually afterwards. In fact, many scientists prefer to visually inspect the images for adjusting the grid and flagging low quality spots instead of totally relying on software recognition. Segmentation is a process used to differentiate the foreground pixels (i.e. the true signal) in a spot grid from the background pixels. This is a tricky computational problem because the spot morphology in a poor-quality image can vary substantially and the background can be high. Furthermore, the image can contain other

Box 1. Different microarray technologies

In general, there are two types of microarray platforms depending on the method of nucleic acid deposition on the chip surface: robotically spotted [52] or *in situ* synthesis by photolithography, a technology that is commonly used in computer chips fabrication [53]. The latter is commercially available from Affymetrix™. Historically the robotically spotted microarrays were referred to as cDNA microarrays because the nucleic acids being spotted were PCR products amplified from cDNA libraries. And the photolithographically synthesized arrays were commonly called oligonucleotides arrays or oligoarrays because shorter oligonucleotides (~25mers) were placed on the arrays and each gene is represented by multiple oligos. It is inaccurate to use the type of probes on arrays to differentiate different platforms because researchers now also prepare oligoarrays by robotically spotting oligonucleotides (~50 to 70mers) on the slide.

Nonetheless, there is still a fundamental difference in the experimental setup between the robotically spotted arrays and photolithographically synthesized ones. In the robotically spotted array experiments, the two samples under comparison are labeled with two different fluorescent dyes and co-hybridized to the same array. This is essentially a comparative hybridization experiment. The ratio between the two dyes indicates the relative abundance of a gene in these two samples. In the photolithographically synthesized array experiments, the two samples under comparison are labeled with the same dye and individually hybridized to different arrays.

Although most downstream analyses like exploratory analysis are similar for the two-microarray platforms, the differences in sample labeling and hybridization have created different requirements in upstream data pre-processing. In particular, because the samples are individually hybridized to different arrays in the case of photolithographically synthesized array experiments, there are specific concerns on features selection [54,55], background adjustment [56], the relationship between signal intensity and transcript abundance [56,57], probe-specific biases [58] and normalization across different arrays [55,56]. This review is focused on the data analysis of the spotted cDNA microarrays, the most accessible microarray platform for general biologists.

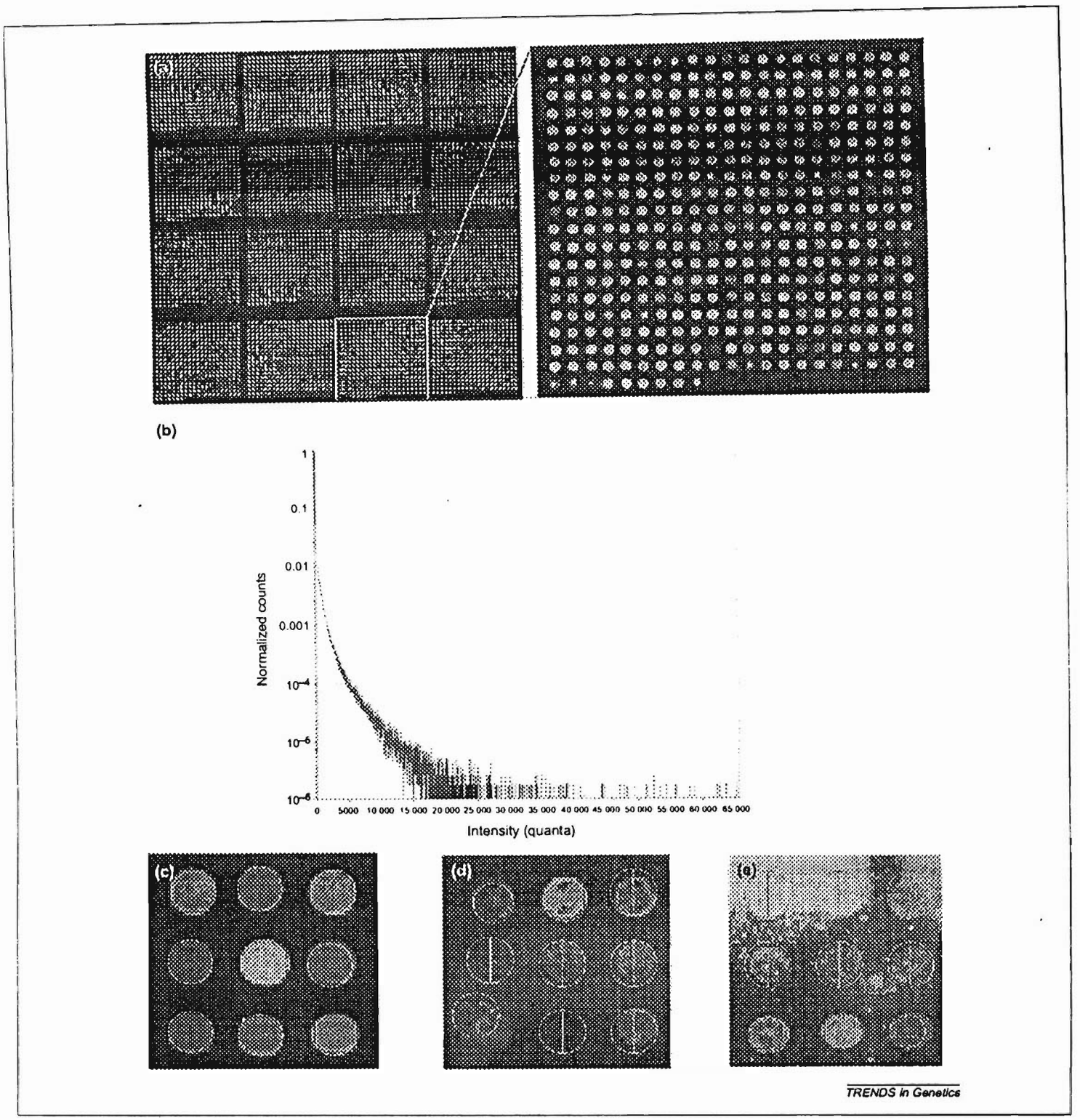


Figure 2. A typical microarray image, pixel distribution histogram for image acquisition, and the effect of image quality on spot recognition and segmentation. **(a)** In this microarray experiment yeast cells treated with a chemical that induced a subtle expression change was compared with the untreated cells by hybridization to a microarray with a complete set of yeast open reading frames (ORFs). **(b)** Pixel histogram for image acquisition. The histograms of the two channels should overlap as much as possible. (c–e) Effect of image quality on spot recognition and segmentation. **(c)** A high-quality image. **(d)** Image with dust contamination. **(e)** Image with high background. (More poor-quality images and how to trouble shoot are available at http://stress-genomics.org/stress.fls/expression/array_tech/trouble_shooting/troubles_index.htm.

imperfections. This can make a proper segmentation difficult. There are several algorithms for segmentation, including FIXED CIRCLE SEGMENTATION, ADAPTIVE CIRCLE SEGMENTATION, adaptive shape segmentation and histogram segmentation. There are also several algorithms for BACKGROUND ESTIMATION, for example constant background, LOCAL BACKGROUND and morphological opening. These algorithms are implemented in different image analysis software [9]. The adaptive circle segmentation and local background estimation algorithms work efficiently for us, but the choice of appropriate algorithms obviously depends on the quality of the raw images. For example, the adaptive circle segmentation that estimates the diameter separately for each spot, works best when all

the spots are circular. Figures 2c–e show the recognition and adaptive circle segmentation results of spots with different background contaminations. When the image quality is high, the algorithm can predict the size of the spots and segment their signal accurately (Figure 2c). If there is dust contamination (Figure 2d) or a high background signal in the image (Figure 2e), the algorithm will not only reject those poor-quality spots, but might also recognize the contamination as a spot (Figure 2d). In this case, both the true signal and background signals will be erroneously estimated. Because it is much more robust for various algorithms to perform segmentation and background estimation processes on a high-quality image than on a low-quality one, it is crucial to produce a high-quality microarray and collect a high-quality image from it in the first place.

Recently there has been an interesting experimental segmentation method reported in which the DNA spots on the microarray were counterstained by 4′, 6′ -diamidino-2-phenylindole (DAPI) and the counterstained image used to assist in the segmentation process [10]. This new experimental approach has apparently resolved many limitations of the algorithmic approach and potentially facilitated the development of a fully automated image analysis system.

After the segmentation process, the pixel intensities within the foreground and background masks (i.e. the areas in the image defined as foreground and background by the software, respectively) are averaged separately to give the foreground and background intensities, respectively. Median or other intensity extraction methods can be used when there are extreme values in the spots that skew the distribution of pixel intensities. Subtracting the BACKGROUND INTENSITY from the foreground intensity in each channel gives the SPOT INTENSITY for calculating the expression ratio between the two channels.

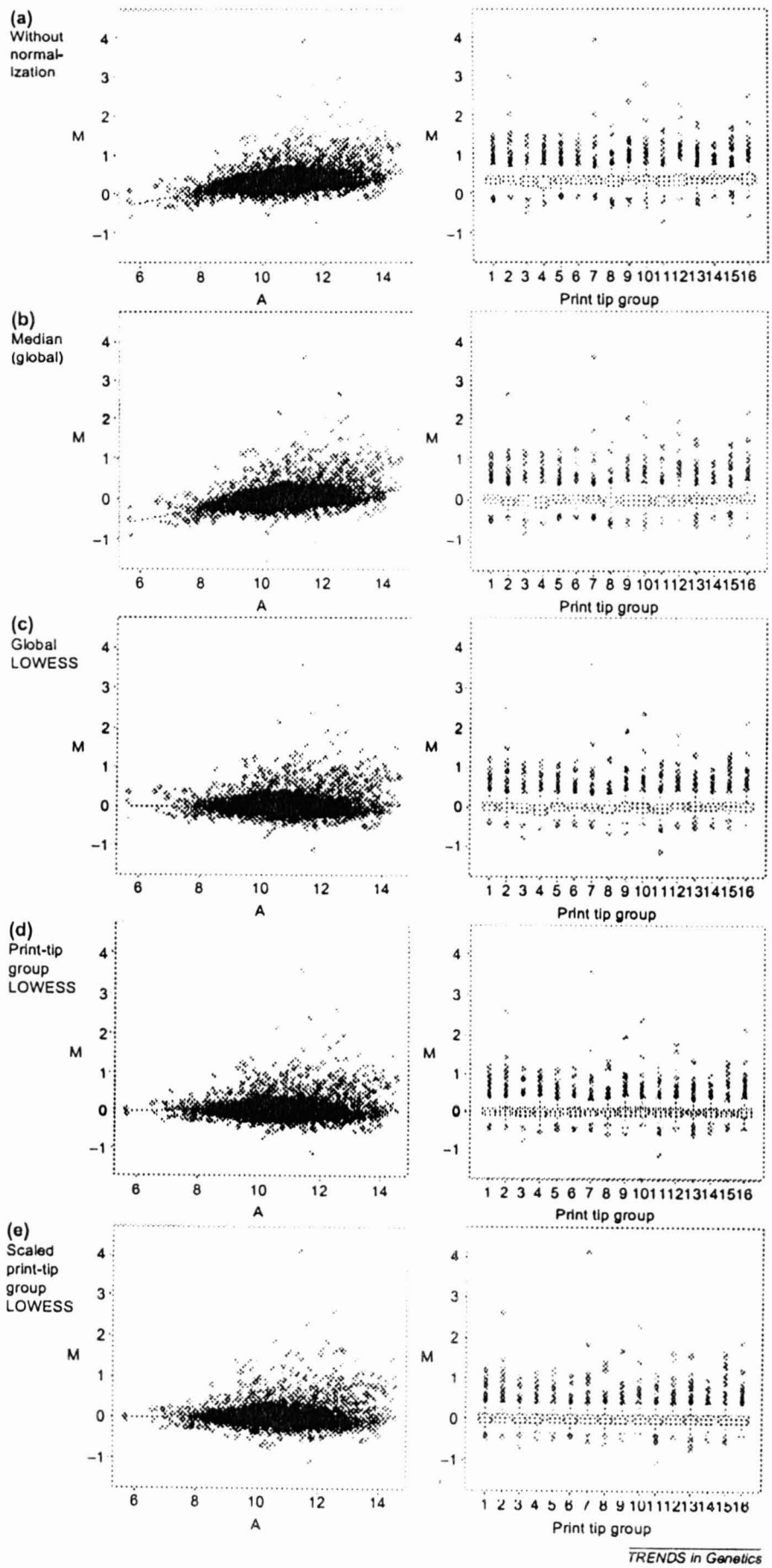
A rapidly developing area that assists in image analysis is the measurement of quality. Some software apply criteria such as diameter, spot area, circularity and replicate uniformity to judge whether a spot is of sufficiently good quality for downstream analysis. The underlying assumption of these criteria is usually a perfect spot, which can be too idealized. A working definition of a good spot is therefore necessary. There is also a need to relate these measures to more common statistical concepts in order that they can be useful for a routine image analysis [9]. A combination of the empirical counterstain segmentation method discussed above [10] and theoretical quality measures can be a practical solution. The DNA counterstain provides information about actual spot morphology and DNA distribution in the spots, which helps to formulate an improved basis for applying different theoretical measures to evaluate the spot quality.

Data pre-processing and normalization

The data extracted from image analysis need to be pre-processed to exclude poor-quality spots and normalized to remove many systematic errors as possible before downstream analysis. Any spot with intensity lower than the background plus two standard deviations should be excluded. The intensity ratios should also be log-transformed so that upregulated and downregulated values are of the same scale and comparable [11].

The process of normalization aims to removing systematic errors by balancing the fluorescence intensities of the two labeling dyes. The dye bias can come from various sources including differences in dye labeling efficiencies, heat and light sensitivities, as well as scanner settings for scanning two channels. Some commonly used methods for calculating normalization factor include: (i) global normalization that uses all genes on the array (Figure 3b); (ii) housekeeping genes normalization that uses constantly expressed housekeeping/invariant genes; and (iii) internal controls normalization that uses known amount of exogenous control genes added during hybridization (http://www.dnachip.org/mged/normalization.html) [11]. Unfortunately these normalization methods are inadequate because dye bias can depend on spot intensity and spatial location on the array. Housekeeping genes are not as constantly expressed as was previously assumed [12]. As a result, using housekeeping genes normalization might introduce another potential source of error. Dye-swapping experiments are seen as a plausible solution to reduce the dye bias problem, but can be impractical because of the limited supply of certain precious samples.

Recently there have been suggestions for using a non-linear normalization method on the basis of gene intensity and spatial information [4,11], which is believed to be superior to the other methods. Figure 3 provides a comparison of various normalization methods, using the data extracted from Figure 2a. All data analyses and graph plotting were performed using statistical microarray analysis (SMA) package (http://stat-zww.berkeley.edu/users/terry/zarray/Software/smacode.html) running in R statistical environment (http://www.r-project.org/). The plots show Log_2 of the expression ratio versus average spot intensity. Ideally the center of the distribution of log-ratios should be zero, the log-ratios should be independent of spot intensity, and the fitted line should be parallel to the intensity axis. In our example, the global locally weighted scatterplot smoothing (LOWESS) normalization is a good choice because it provides a good balance on the three factors mentioned above (Figure 3c). The fluorescent images (Figure 2a) do not suffer from serious spatial effects, as indicated by a very similar log expression ratio distribution among all the print-tips in the bloxplot for the global LOWESS normalization (Figure 3c). However, when there is a significant difference in the distribution of log-ratios among the print-tips in the bloxplot, suggesting a possible spatial effect, print-tip group LOWESS (Figure 3d) or scaled print-tip group LOWESS normalization (Figure 3e) should be considered. Apart from within-a single array, the distribution of gene expression ratios from replicate experiments might have different distribution of log ratios due to the difference in experimental conditions. Therefore scaling adjustment is often necessary to standardize the distribution of log-ratios across replicate experiments to prevent any particular experiment becoming dominant and affecting downstream statistical analysis.


(a)
Without normalization
(b)
Median (global)
(c)
Global LOWESS
(d)
Print-tip group LOWESS
(e)
Scaled print-tip group LOWESS
M
A
Print tip group
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
TRENDS in Genetics

Data analysis

The next stage of analysis is to apply various statistical and data mining techniques to study the data. There are several typical approaches that are discussed in the following sections.

Significance inference – identifying significantly differentially expressed genes

Traditionally, differentially expressed genes are inferred by a fixed threshold cut off method (i.e. a two-fold increase or decrease), but this is statistically inefficient, the main reason being that there are numerous systemic and biological variations that occur during a microarray experiment. Although some of the systemic variations such as dye bias can be effectively removed by normalization, random biological variations such as sample-to-sample and physiological variations are more difficult to handle [13,14] (for a comprehensive review of various statistical issues, variations and errors of microarray experiment see Ref. [15]). Because of these underlying variations, merely using a fixed threshold to infer significance might increase the proportion of false positives or false negatives. A better framework of significance inference includes calculation of a statistic based on replicate array data for ranking genes according to their possibilities of differential expression and selection of a cut-off value for rejecting the null-hypothesis that the gene is not differentially expressed.

Replication of a microarray experiment is essential to obtain the variation in the gene expression for statistics calculation. It has been suggested that every microarray experiment should be performed in triplicate to increase data reliability [16]. There are two types of replication: biological and technical. Biological replication refers to the analysis of multiple independent biological samples (e.g. one tissue type obtained from different patients with the same disease, or individual samples of a particular cell line under the same treatment), whereas technical replication refers to the repetition of microarray experiment using the same extracted RNA samples. Biological replication is particularly important for expression profiling of disease tissues, because there might be variability of expression among the same tissue type or tissue heterogeneity. Any particular tissue might not be representative of the whole disease sample group. Technical replication provides a precise measurement of gene expression for a particular sample and eliminates many technical variations introduced during the experiment. Unfortunately, merely obtaining a precise expression measurement of a tissue by technical replication will not resolve the problem of biological variation. Therefore it is usually preferable to have biological replication rather than technical replication if there are not enough tissues or resources to perform several microarray experiments, provided the experiment procedures are carried out carefully [4,5]. Statistical methods such as Student's *t*-test and its variants [17,18], ANOVA [19,20], Bayesian method [17,20,21], or Mann–Whitney test [22], can be used to rank the genes from replicated data.

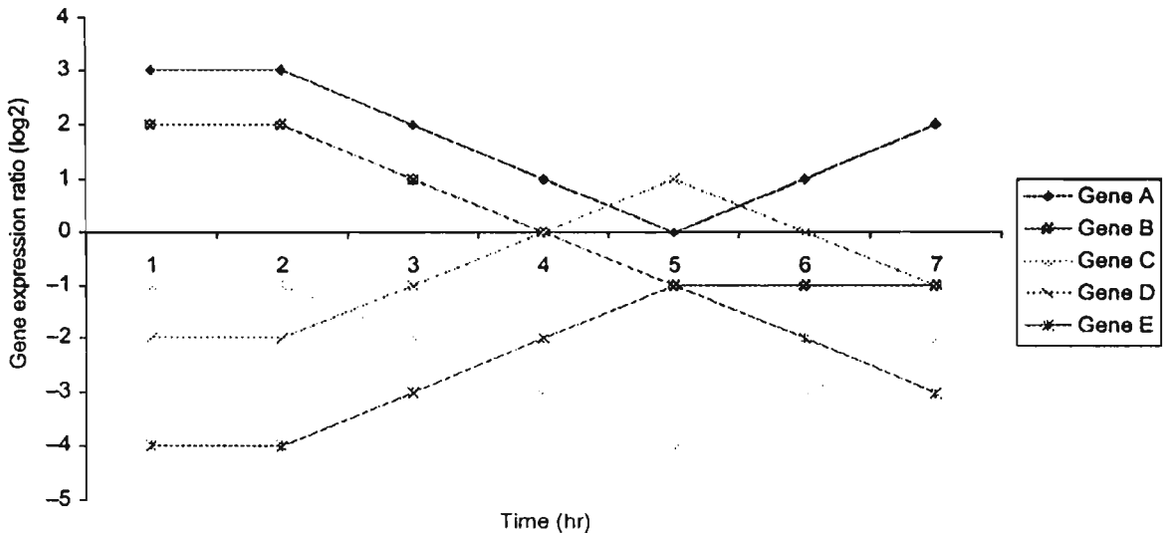
Setting a cut-off for differential expression is tricky, because one has to balance the false positives (Type I error) and the false negatives (Type II error). Furthermore, performing statistical tests for tens of thousands of genes creates a multiple hypothesis-testing problem. For example, in an experiment with a 10 000-gene array in which the significance level α is set at 0.05, $10\,000 \times 0.05 = 500$ genes would be inferred as significant even though none is differentially expressed. Therefore using a *p*-value of 0.05 is likely to exaggerate Type I errors. The multiple hypothesis testing problem is conventionally tackled by conservative approaches that control the family-wise error rate (FWER), the probability of having at least one false positive among all testing hypotheses [23]. A classical example is the Bonferroni correction. However, controlling the FWER can be too stringent and limits the power to identify significantly differentially expressed genes. In fact, differential expression is usually confirmed by RT-PCR, northern blots or *in situ* hybridization [24]. It is often acceptable to have few false positives if the majority of true positives are chosen. Therefore it might be more practical to control the false discovery rate (FDR) [25], the expected proportion of false positives among the number of rejected hypotheses. A program, statistical analysis of microarray (SAM), has been developed to utilize this FDR concept as a tool to assist in determining a cut-off after performing adjusted *t*-tests (http://www-stat.stanford.edu/~tibs/SAM/index.html) [18].

Exploratory data analysis – understanding the (dis)similarities of the gene expression levels among all samples

Also known as unsupervised data analysis, exploratory data analysis does not require the incorporation of any prior knowledge in the process. It is essentially a grouping technique that aims tot find genes with similar behaviors (i.e. expression profiles). Some commonly used examples include principal component analysis (PCA) [26] or singular value decomposition (SVD) [27] for dimensionality reduction, as well as hierarchical clustering [28], K-means clustering [29] and self organizing maps (SOMs) [30] for clustering. There are already several excellent reviews on various unsupervised analyses and their applications in microarray data mining [31–33], therefore we do not discuss their details here.

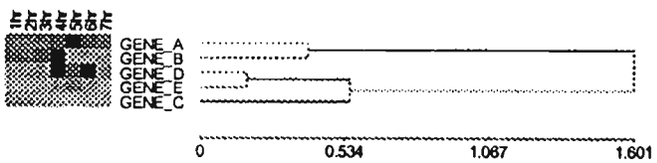
There is perhaps no unsupervised data analysis that can suit all situations. Different analyses or even different parameters of the same analysis can reveal unique aspects of the data. This idea is illustrated in Figure 4, in which five genes from a hypothetical time series data are clustered using various distance or similarity measures and unweighted pair group method with arithmetic mean (UPGMA) algorithm. Each distance or similarity measure

Figure 3. A comparison of various normalization methods. The raw data was extracted from Figure 2a. Any spot with intensity lower than the background plus two standard deviations or of poor-quality was excluded from further analysis. From top to bottom: Log_2 ratios (M) versus average intensities (A) plot and boxplot of the data without normalization (a) and with four different kinds of normalization methods: (b) median, (c) global locally weighted scatterplot smoothing (LOWESS), (d) print-tip group LOWESS, (e) scaled print-tip group LOWESS.

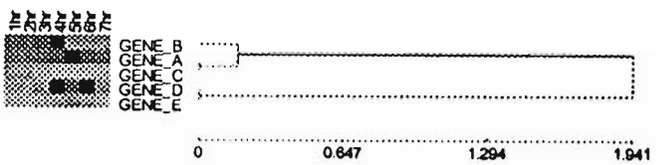


Distance similarity measure

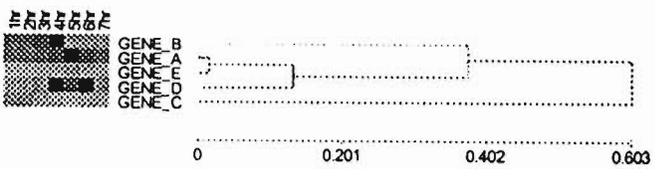
(a) Correlation coefficient without centering



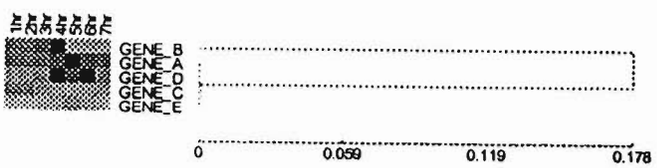
(b) Correlation coefficient with centering



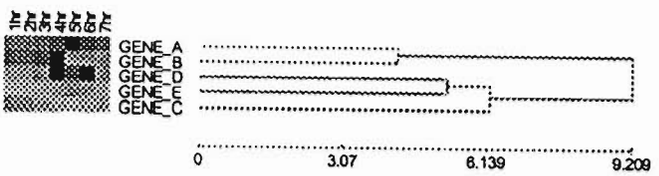
(c) Absolute correlation coefficient without centering



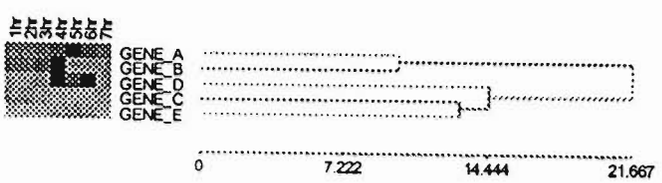
(d) Absolute correlation coefficient with centering



(e) Euclidean distance



(f) Manhatten distance



can assign the genes to different clusters. For example, Euclidean and Manhattan distances are sensitive to absolute expression levels, and are able to reveal those genes that have similar expression levels in the cluster. Two main clusters are identified in the data, one for gene *A* and *B* and the other cluster for gene *C*, *D* and *E* (Figure 4e,f). *A* and *B* are clustered with each because their overall expression ratios more similar when compared with *C*, *D* and *E*, and vice versa. The similarity between their expression profiles suggests the genes in the two clusters might be co-regulated. However, if the researchers conclude the analysis at this stage, they are likely to miss some other interesting relationship among the genes. A slightly different picture is revealed by using correlation coefficient with centering, a similarity measure that is sensitive to the expression profile shape, regardless of the expression levels (Figure 4b). Gene *A*, *B* and *C* are grouped in the same cluster whereas *D* and *E* are in another. Intriguingly, *A* and *C*, gene *D* and *E* are correlated with each other perfectly using this distance measure. An inspection of the expression profile offers a hint. Although *A* and *C* differ largely in expression level, the shape of their expression profiles is the same. This is also true for gene *D* and *E*. As a result, the correlation coefficients for both *A* and *C* and gene *D* and *E* are 1. This result suggests gene *A* and *C*, gene *D* and *E* are likely to be co-regulated, and analyzing their promoters can sometimes identify common regulatory elements. Further insight is provided using absolute correlation coefficient with centering as a similarity measure (Figure 4d). This time *A*, C, *D* and *E* are clustered perfectly together, leaving *B* separate. It is because the shape of the expression profiles of *A* and *C* are a mirror image of *D* and *E*. Although their correlation coefficient is -1, which will place them in two separate clusters as shown in Figure 4b, the absolute value of their correlation coefficient is the same and will place them in the same cluster. Therefore it is very likely that *A*, *C*, *D*, *E* are regulated by a same factor or mechanism, which represses the expression *A* and *C* while enhancing the expression of *D* and *E*, and vice versa. The same principle also applies to the choice of clustering algorithms [31].

Hence, it is always advisable to apply several unsupervised analyses and different parameters to explore the data. Nonetheless, there must be a balance between the time spent on data analysis and the time spent on subsequent experimental confirmation. Unsupervised analysis is a useful method for generating new hypotheses. The validity of the result has to be built upon both statistical significance and biological knowledge.

Class prediction – using gene expression profiles as a means to classify samples

Another intriguing type of data analysis is to train a classifier algorithm using the expression profiles of predefined sample groups, so that the classifier can best assign any new sample to the respective group. This type of analysis is also known as supervised data analysis, which has great promise in clinical diagnostics [31] and has been used successfully in several recent studies [34–36]. Examples of such analysis include support vector machines [37], artificial neural networks [38], k-nearest neighbor [39] and various discrimination methods (http://stat-www.berkeley.edu/users/terry/zarray/Html/discr.html). The ultimate goal is to generalize the trained classifier as a routine diagnostic tool for differentiating between the samples that are difficult or even impossible to classify by available methods.

The challenge for supervised data analysis is to generalize the classifier for all situations. The training samples are often limited in number that might not be sufficiently representative for their classes in general. Over-training on the same dataset would result in a situation called 'over-fitting', in which the classifier is very effective in classifying the training samples but not accurate enough for new samples. A balance between accuracy and generalizability has to be established by validation of the trained classifier. Several approaches are available for this purpose. For example, the training samples are divided into two individual sets, one for training and one for validation. The training of the classifier will be stopped when the prediction error on the validation set reaches a minimum. More sophisticated cross-validation methods divide the training dataset into several subsets. Each subset will be the validation set in turn. The overall accuracy therefore is the average accuracy across all validation trials. An extreme case of cross-validation is called leave-one-out cross-validation, in which one sample is taken away from the training set to be a validation sample each time. An investigation of several supervised analyses, their performance, and cross-validation was detailed previously [40].

An emerging approach – pathway analysis

Genes never act alone in a biological system – they are working in a cascade of networks. As a result, analyzing the microarray data in a pathway perspective could lead to a higher level of understanding of the system. There are at least three interesting approaches in this area. The first is a natural extension of the exploratory cluster analysis described above. If several genes are assigned to the same group by cluster analysis, as discussed above, they might be co-regulated or involved in the same signaling pathway. Analyzing the promoters of this group of genes can often reveal common regulatory motifs and unveil a higher level of network organization in the biological system [41]. The second is to reverse-engineer the global genetic pathways, the identification of the global regulatory network architecture from microarray data. It can be done by a systematic targeted perturbation like mutation or chemical treatment [42], and time series experiments [43]. The assumption here is that the perturbation will cause a change in expression of other proteins in the network. This

Figure 4. Different distance measures provide different views of the data. Line graphs of a hypothetical time series experiment with five genes and seven time points (upper panel). Hierarchical clustering of the data using six common distance or similarity measures (lower panel): (a) correlation coefficient without centering, (b) correlation coefficient with centering, (c) absolute correlation coefficient without centering, (d) absolute correlation coefficient with centering, (e) Euclidean distance, (f) Manhattan distance. Clustering was performed using unweighted pair group method with arithmetic mean algorithm (UPGMA).

change in the expression profiles should be able to capture the underlying architecture of the network. Various methods have been proposed for constructing a network from this kind of microarray data, such as a Boolean network that simplifies gene expression as a binary logical value to infer the induction of a gene as a deterministic function of the state of a group of other genes [44–46] and a Bayesian network that models interactions among genes, evaluates different models and assigns them probability scores [47,48] (readers are referred to two excellent reviews on these and other methods for reverse engineering of networks [49,50]). The final approach is to study the expression data on a pathway perspective. Our group has recently developed a method called Pathway Processor (http://cgr.harvard.edu/cavalieri/pp.html) that can map expression data onto metabolic pathways and evaluate which metabolic pathways are most affected by transcriptional changes in whole-genome expression experiments [51]. We used the Fisher Exact Test to score biochemical pathways according to the probability that as many or more genes in a pathway would be significantly altered in a given experiment than by chance alone. Results from multiple experiments can be compared, reducing the analysis from the full set of individual genes to a limited number of pathways of interest.

Conclusion

Microarray analysis is evolving rapidly. New and more complex analyses appear everyday, making it easy for the researcher to get lost in endless new methods and software. Collaborating with statisticians and mathematicians is often advisable for performing a proper microarray analysis. Nonetheless, this will not replace biological expertise, a good foundation for statistical methods and meticulousness in conducting experiments.

Acknowledgements

YFL is supported by a Croucher Foundation Postdoctoral Fellowship. We thank Alice Yu Ming Lee and Abel Chiu Shun Chun for their critical comments on this manuscript.

References

1 Leung, Y.F. *et al.* (2002) Microarray software review. In *A practical approach to microarray data analysis* (Berrar, D.P. *et al.*, eds), Kluwer academic
2 Box, G.E.P. *et al.* (1978) *Statistics for experimenters – an introduction to design, data analysis, and model building*, John Wiley & Sons
3 Churchill, G.A. (2002) Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.* 32 (Suppl. 2), 490–495
4 Yang, Y.H. and Speed, T. (2002) Design issues for cDNA microarray experiments. *Nat. Rev. Genet.* 3, 579–588
5 Simon, R.M. and Dobbin, K. (2003) Experimental design of DNA microarray experiments. *Biotechniques*, S16–S21
6 Perou, C.M. (2001) Show me the data!. *Nat. Genet.* 29, 373
7 Brazma, A. *et al.* (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* 29, 365–371
8 Anonymous, (2002) Microarray standards at last. *Nature* 419, 323
9 Yang, Y.H. *et al.* (2001) Analysis of cDNA microarray images. *Brief. Bioinform.* 2, 341–349
10 Jain, A.N. *et al.* (2002) Fully automatic quantification of microarray image data. *Genome Res.* 12, 325–332
11 Quackenbush, J. (2002) Microarray data normalization and transformation. *Nat. Genet.* 32 (Suppl.), 496–501
12 Lee, P.D. (2002) Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. *Genome Res.* 12, 292–297
13 Novak, J.P. *et al.* (2002) Characterization of variability in large-scale gene expression data: implications for study design. *Genomics* 79, 104–113
14 Pritchard, C.C. *et al.* (2001) Project normal: defining normal variance in mouse gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13266–13271
15 Nadon, R. and Shoemaker, J. (2002) Statistical issues with microarrays: processing and analysis. *Trends Genet.* 18, 265–271
16 Lee, M.L. *et al.* (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9834–9839
17 Lönnstedt, I. and Speed, T.P. (2002) Replicated Microarray Data. *Stat. Sinica* 12, 31–46
18 Storey, J.D. and Tibshirani, R. (2003) SAM thresholding and false discovery rates for detecting differential gene expression in DNA microarrays. In *The Analysis of Gene Expression Data: Methods and Software* (Parmigiani, G. *et al.*, eds), Springer
19 Kerr, M.K. *et al.* (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol.* 7, 819–837
20 Long, A.D. *et al.* (2001) Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in *Escherichia coli* K12. *J. Biol. Chem.* 276, 19937–19944
21 Baldi, P. and Long, A.D. (2001) A Bayesian framework for the analysis of microarray expression data: regularized *t* test and statistical inferences of gene changes. *Bioinformatics* 17, 509–519
22 Wu, T.D. (2001) Analysing gene expression data from DNA microarrays to identify candidate genes. *J. Pathol.* 195, 53–65
23 Dudoit, S. *et al.* (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Stat. Sinica* 12, 111–139
24 Chuaqui, R.F. *et al.* (2002) Post-analysis follow-up and validation of microarray experiments. *Nat. Genet.* 32 (Suppl. 2), 509–514
25 Reiner, A. *et al.* (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19, 368–375
26 Raychaudhuri, S. (2000) Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac. Symp. Biocomput.*, 455–466
27 Alter, O. *et al.* (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. U. S. A.* 97, 10101–10106
28 Eisen, M.B. *et al.* (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14863–14868
29 Tavazoie, S. *et al.* (1999) Systematic determination of genetic network architecture. *Nat. Genet.* 22, 281–285
30 Tamayo, P. *et al.* (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. U. S. A.* 96, 2907–2912
31 Quackenbush, J. (2001) Computational analysis of microarray data. *Nat. Rev. Genet.* 2, 418–427
32 Sherlock, G. (2001) Analysis of large-scale gene expression data. *Brief. Bioinform.* 2, 350–362
33 Valafar, F. (2002) Pattern recognition techniques in microarray data analysis: a survey. *Ann. N. Y. Acad. Sci.* 980, 41–64
34 Pomeroy, S.L. *et al.* (2002) Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 415, 436–442
35 Shipp, M.A. *et al.* (2002) Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.* 8, 68–74
36 Khan, J. *et al.* (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.* 7, 673–679
37 Brown, M.P. (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. U. S. A.* 97, 262–267
38 Vohradsky, J. (2001) Neural network model of gene expression. *FASEB J.* 15, 846–854
39 Theilhaber, J. *et al.* (2002) Finding genes in the C2C12 osteogenic

pathway by k-nearest-neighbor classification of expression data. *Genome Res.* 12, 165–176
40 Ben-Dor, A. *et al.* (2000) Tissue classification with gene expression profiles. *J. Comput. Biol.* 7, 559–583
41 Pilpel, Y. *et al.* (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.* 29, 153–159
42 Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126
43 Tavazoie, S. *et al.* (1999) Systematic determination of genetic network architecture. *Nat. Genet.* 22, 281–285
44 Liang, S. *et al.* (1998) Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac. Symp. Biocomput.*, 18–29
45 Akutsu, T. *et al.* (2000) Algorithms for identifying Boolean networks and related biological networks based on matrix multiplication and fingerprint function. *J. Comput. Biol.* 7, 331–343
46 Maki, Y. *et al.* (2001) Development of a system for the inference of large scale genetic networks. *Pac. Symp. Biocomput.*, 446–458
47 Friedman, N. *et al.* (2000) Using Bayesian networks to analyze expression data. *J. Comput. Biol.* 7, 601–620
48 Hartemink, A.J. *et al.* (2001) Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. *Pac. Symp. Biocomput.*, 422–433
49 de Jong, H. (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.* 9, 67–103
50 D'haeseleer, P. (2000) Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* 16, 707–726
51 Grosu, P. *et al.* (2002) Pathway Processor: a tool for integrating whole-genome expression results into metabolic networks. *Genome Res.* 12, 1121–1126
52 Schena, M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470
53 Lipshutz, R.J. *et al.* (1999) High density synthetic oligonucleotide arrays. *Nat. Genet.* 21 (Suppl. 1), 20–24
54 Zhou, Y. and Abagyan, R. (2003) Algorithms for high-density oligonucleotide array. *Curr. Opin. Drug Discov. Devel.* 6, 339–345
55 Schadt, E.E. *et al.* (2001) Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *J. Cell. Biochem.* 37 (Suppl.), 120–125
56 Schadt, E.E. *et al.* (2000) Analyzing high-density oligonucleotide gene expression array data. *J. Cell. Biochem.* 80, 192–202
57 Sasik, R. *et al.* (2002) Statistical analysis of high-density oligonucleotide arrays: a multiplicative noise model. *Bioinformatics* 18, 1633–1640
58 Li, C. and Wong, W.H. (2001) Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. U. S. A.* 98, 31–36

COMMENTARY

Questions and Answers on Design of Dual-Label Microarrays for Identifying Differentially Expressed Genes

Kevin Dobbin, Joanna H. Shih, Richard Simon

The rapid growth in the use of microarrays has generated many questions about how to design experiments that use this technology effectively. Investigators need answers to questions about RNA sample selection, allocation of samples to arrays, robustness of design, dye bias, sample size, and statistical power to ensure that the experimental objectives are achieved. We address some common questions that arise in designing dual-label microarray experiments and provide statistical answers to these questions, focusing specifically on how to select optimal designs for the identification of differentially expressed genes.

BACKGROUND

The dual-label microarray measures the expression level of thousands of genes for a sample of cells. A common goal of microarray experiments is to determine which genes are differentially expressed among two or more predefined classes of biologic specimens. These types of study goals are referred to as "class comparisons" *(1)*. Some examples of class comparisons are 1) identifying the differentially expressed genes in BRCA1 mutation–positive, BRCA1 mutation–negative, and sporadic cases of primary breast cancer *(2)*; 2) identifying the differentially expressed genes in colon cancer cells treated with high versus low doses of camptothecin *(3)*; and 3) identifying the differentially expressed genes in the prostate cancer cell line LNCaP before and after treatment with the tumor growth inhibitor, PC-SPES *(4)*. Because of their widespread use, class comparison experiments will be the focus of this commentary.

A microarray generally consists of either cDNA or externally synthesized oligonucleotides that are printed or coated on glass slides. A dual-label microarray uses competitive hybridization in which nucleic acids (i.e., cDNA, cRNA, or RNA) derived from two RNA sources are hybridized to the same microarray *(5,6)*. The cDNA from one source is labeled with green (Cy3) dye, and the cDNA from the other source is labeled with red (Cy5) dye, either directly or indirectly *(7)*. The cDNA or oligonucleotides representing different genes are immobilized on the glass slide and are often referred to as spots. For each spot there are two corresponding measurements, one for each dye, often referred to as the two channels. The advantages of competitive hybridization for cDNA experiments have been well established *(8)*. The relative intensities of two labeled specimens measured at a single spot are much less variable than the relative intensities if measured at corresponding spots on different arrays. Relative expression measurements provide a means of controlling the variability in the size and shape of corresponding spots and the effects of variation in sample concentration on the surface of the array.

The relative expression measurements compare the expression levels of labeled cDNA that have originated from two different RNA sources. cDNA from a single source is often applied to every microarray slide and is labeled with the same dye (either Cy3 or Cy5). These labeled cDNAs are referred to as the reference. If the reference is labeled with Cy3 dye, then the nonreference samples are all labeled with Cy5 dye. Comparisons between the nonreference samples are based on log-ratios of the intensity of the Cy5 dye to the intensity of the Cy3 dye for corresponding spots on different arrays. Basing comparisons between the nonreference samples on the log-ratios eliminates the sources of variability attributable to aspects of the spot that affect both channels similarly. The gene expression data from such a design, called a reference design, is easy to analyze because simple t tests or similar statistical methods can be applied directly to the log-ratios, and there is much existing software available for performing such tests. In addition, it is also possible to control for spot variability from designs that do not use a reference by statistical modeling. Hence, the reference design may or may not be the best choice for a particular situation.

The ability to measure expression levels for two samples on each cDNA array permits a number of design options for assigning specimens to labels and arrays. When choosing among these design options, one should consider the objectives of the experiment, the sources of variability, and the differences between dyes with regard to labeling and detection characteristics. The purpose of this commentary is to provide statistically sound advice about the design of investigations for finding differentially expressed genes using dual-label microarray platforms. We present a number of results comparing the statistical properties of different designs that we have established elsewhere. However, to keep the presentation nonmathematical, we have replaced equations presented in our earlier articles *(9,10)* with graphical displays where appropriate.

Affiliation of authors: Biometric Research Branch, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD.

Correspondence to: Kevin Dobbin, PhD, National Cancer Institute, 6130 Executive Blvd., EPN 8124, Bethesda, MD 20892–7434 (e-mail: dobbinke@mail.nih.gov).

See "Notes" following "References."

DOI: 10.1093/jnci/djg049

Journal of the National Cancer Institute, Vol. 95, No. 18, © Oxford University Press 2003, all rights reserved.

Sample Selection

Is It Sufficient to Sample One Individual From Each Class?

The answer is no, because the goals of class comparison are to determine whether the gene expression profiles are different between the classes and to identify differentially expressed genes. Different individuals in the same class are not expected to have exactly the same gene expression level measurements. Biologic variation and measurement error will produce some differences in the gene expression profiles. If we sample only one individual from each class, then there is no way to distinguish expression differences associated with class from those associated with biologic variation or measurement error. Some genes may vary widely in their expression level from individual to individual in the same class, whereas others may display differential expression that is relatively small but is nonetheless critical for class distinction. Therefore, it is important to have multiple (and distinct) individuals from each class to obtain an estimate of biologic variation. Similarly, in studying gene expression in model organisms under different biologic conditions, it is important to have distinct applications of the conditions and harvesting of cells.

How Many Replicates of Each RNA Sample Should Be Hybridized?

Some investigators *(11)* have promoted using three or more replicate measurements for each RNA sample, and others *(12)* have suggested that at least two replicate measurements are required for each sample. These guidelines may be correct in some situations; however, they will probably not be correct for class comparison problems. When one is interested in class comparison, then replication measurements should generally be at the population level, so that each replicate represents RNA from a different individual. Intuitively, the reason that this level of replication produces the best comparisons is that, by replicating at the population level, one simultaneously reduces variability from both population heterogeneity and experimental error. When multiple aliquots are replicated from the same RNA source, one only reduces variability from experimental error. Therefore, replication of individual samples is inefficient for class comparisons.

Hybridization replicates increase the accuracy of the individual sample measurements *(11)*. However, if the number of arrays is fixed (e.g., when one only has time or resources available to run a prespecified number of arrays), then increasing the hybridization replicates requires decreasing the number of distinct RNA samples assayed. The result of this approach is a reduction in the accuracy of the class mean estimates. The relationship between sample measurement accuracy and class mean estimate accuracy as the number of hybridization replicates per sample increases for an experiment with a fixed number of 24 arrays is shown in Fig. 1. (*see* supplemental information at http://jncicancerspectrum.oupjournals.org/jnci/content/vol95/issue18/index.shtml for details and proof). Accuracy is defined as the inverse of the variance of the mean estimate. Population parameter estimates are most accurate when hybridization replication (i.e., subsampling) is avoided, even though the accuracy of individual sample estimates is at a minimum when there is no subsampling. With less subsampling, one is better able to detect differentially expressed genes in the classes when the total number of arrays is fixed. An obvious exception to this rule is when only a limited number of valuable RNA samples are available and when one does not have access to more. Assaying each sample multiple times will clearly be preferable to assaying each sample only once.

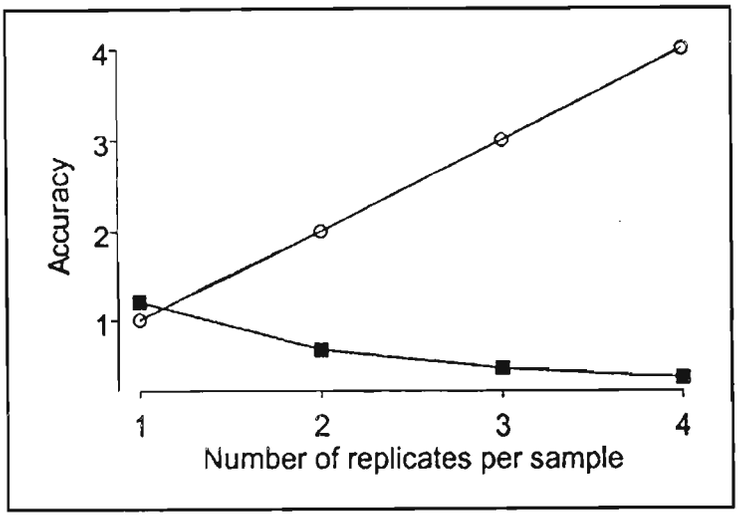


Fig. 1. Accuracy of sample and class mean estimates as a function of the number of replicates per sample. The number of arrays was fixed at 24. Accuracy is defined as the inverse of the variance of the estimate. The estimates are the difference in the class averages for class mean estimates (**solid squares**) and the average of repeated measurements on the same sample for sample estimates (**open circles**). Some parameters, such as the biologic and experimental variation, were fixed to construct the display. For further details about mathematical equations, refer to online supplemental information (*see* http://jncicancerspectrum.oupjournals.org/jnci/content/vol95/issue18/index.shtml).

One might think that replicate hybridizations would help offset high measurement variability in low-quality microarray experiments that display high variation in repeated assays on the same sample. The power to detect a differentially expressed gene as a function of the number of subsamples per sample used, for example, of both a high-quality (i.e., displays low variation in repeated assays on the same sample) and a low-quality experiment, is shown in Fig. 2 (*see* supplemental information at http://jncicancerspectrum.oupjournals.org/jnci/content/vol95/issue18/index.shtml for details and proof). The high-quality experiment is assumed to have an experimental error variance of half the biologic variance, and the low-quality experiment is assumed to have an experimental error variance twice that of the biologic variance. Although the loss of power is more dramatic for the high-quality experiment than for the low-quality experiment, the low-quality experiment also loses power when one replicates hybridizations for a fixed number of arrays.

What Are the Advantages and Disadvantages of Pooling Samples?

Pooling samples involves mixing together RNA from several sources before labeling and hybridization. Two motivations for pooling samples are 1) not enough RNA available from each individual to perform the assay, and 2) wanting to reduce the number of arrays used. Investigators sometimes hope to cut down on the number of arrays needed by comparing a single pooled sample from each class. The reasoning behind this approach is that the concentration of an mRNA molecule in a pooled sample is likely to be closer to the average concentration for the class than the concentration in a sample from a single

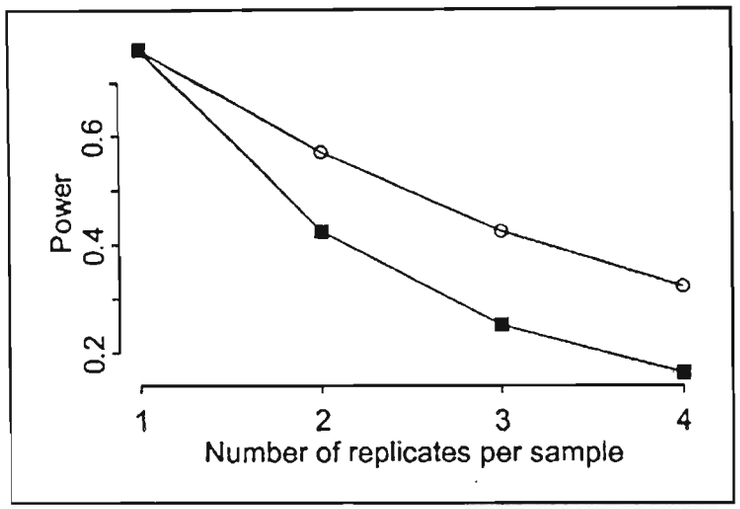


Fig. 2. Statistical power to detect differentially expressed genes as a function of the number of replicates per sample. The number of arrays was fixed at 24. The high-quality (i.e., displays low variation in repeated assays on the same sample) experiment (**solid squares**) has experimental error variance half that of the biologic variance. The low-quality (i.e., displays high variation in repeated assays on the same sample) experiment (**open circles**) has experimental error variance twice that of the biologic variance. The power is the probability of detecting a twofold change in gene expression levels for the high-quality experiment and a $2\sqrt{2}$-fold change in gene expression levels for the low-quality experiment (i.e., to make the powers comparable). For further details about mathematical equations, refer to online supplemental information (*see* http// jncicancerspectrum.oupjournals.org/jnci/content/vol95/issue18/index.shtml).

individual. Unfortunately, a single pooled sample from each class will not be adequate for statistical inference, because one has no estimate of the biologic or experimental variability in the gene expression levels for pooled samples constructed from samples of the same class. Taking multiple subsamples from each pool and repeating them on multiple microarrays does not solve this problem, because variation among the subsamples will reflect only measurement error and will not include biologic variation.

It is possible to perform valid statistical comparisons between the classes with pooled data, but this approach requires multiple pooled samples from each class. Different pools of RNA should be constructed from different sets of individuals so that the pooled samples are independent and represent true replication. Comparisons of gene expression levels between classes are then straightforward. However, there are still some disadvantages to this approach. 1) It does not allow one to understand the contribution of individual RNA samples to the observed gene expression levels, which makes it impossible to identify outlier or poor-quality RNA samples. 2) A pool average is potentially biased for the class average—that is, the average expression level of a gene in the pool may differ from the average of the expression levels of the gene in the contributing samples, which can happen because of inequalities in the amounts of RNA contributed by different samples or because mixing of the RNA causes unanticipated alteration of gene expression. 3) It may be difficult or impossible to understand how gene expression is distributed in the population from pooled data and, hence, to make valid statistical inferences or predictions for individuals. In summary, pooling of samples is recommended when there is not enough RNA from individual samples to run a microarray. The use of several independent pools from each class will allow for valid statistical inference about the classes.

Pairing Samples For Co-Hybridization

What Types of Designs Should Be Considered?

Three designs have been proposed for cDNA microarray class comparison experiments (Fig. 3). The reference design is by far the most widely used because spot-to-spot variation can be eliminated in a simple way by using ratios or log-ratios. There are many other advantages to the reference design, which are explored later in this section; however, its widespread use should not preclude consideration of other alternatives. The distinctive feature of a reference design is that expression of a gene for a sample is measured relative to the expression of that gene at the same spot on the same array for a reference sample.

The ability to co-hybridize two differentially labeled samples to each array may appear to open a Pandora's box of experimental design possibilities. However, do we really need to sort through every possible design? The fact that the difference in gene expression levels between corresponding spots on different microarrays is a major source of variability makes the arrays analogous to a blocking factor in agricultural experiments. There is extensive statistical literature on the design of such experiments *(13,14)*, but it cannot be applied directly to dual-labeled microarray experiments, because the error structure for microarray data is somewhat different than the agricultural analog. We have adapted the method for deriving optimal designs in the

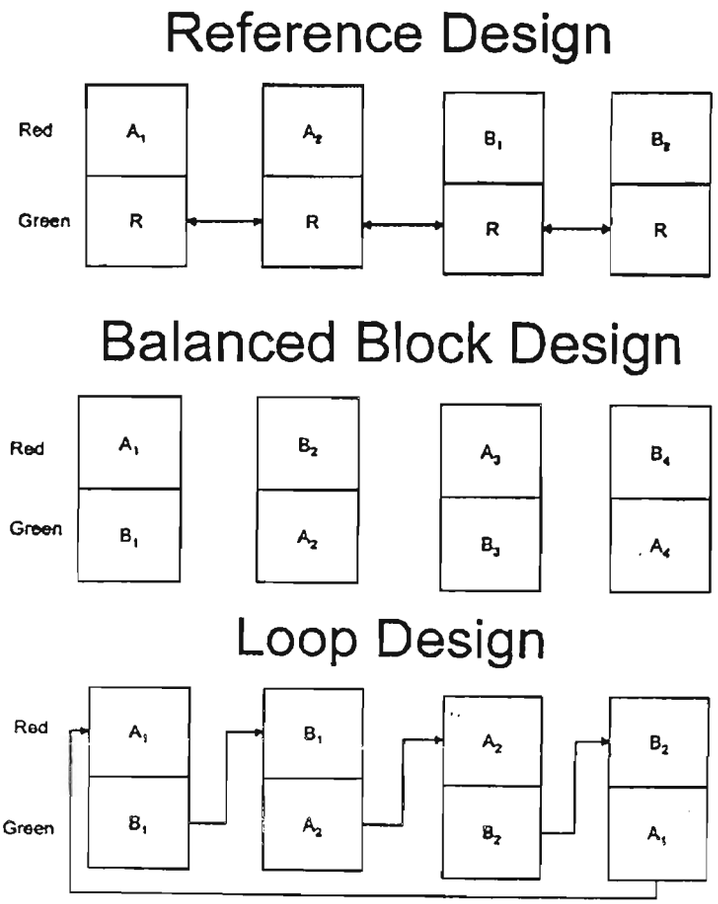


Fig. 3. Design diagrams for cDNA microarray class comparison experiments. **Rectangles** represent the arrays. A_1 is sample 1 from class A, B_1 is sample 1 from class B, A_2 is sample 2 from class A, and so on. R is the reference sample. **Arrows** connect samples repeated on multiple arrays. **Red** is the Cy5 dye and **Green** is the Cy3 dye used to label the reference and nonreference samples.

presence of a blocking factor to microarray experiments *(9)* and have established that, for many class comparison studies, the balanced block design shown in Fig. 3 is optimal. The effect of spot-to-spot variation in gene expression levels is eliminated in the balanced block design because each gene's expression level is measured at the same spot on the same array for samples from each of the two classes being compared.

The third type of design that might be considered for cDNA microarrays is one proposed by Kerr and Churchill *(15)*, which they called a loop design (Fig. 3). Unlike the two other designs, the loop design requires two aliquots from each RNA sample. These aliquot pairs connect the arrays and are arranged so that the connected arrays form a loop pattern.

Class comparisons for the balanced block design and the loop design are accomplished by fitting an analysis-of-variance (ANOVA) model to the logarithm of the background-corrected channel-specific intensities *(9)* and fitting a separate model for each gene. This approach can also be used for analysis of the reference design, but the results are equal to or very similar to applying simple Student's *t* tests to the log-ratio measurements.

More elaborate designs have been proposed to achieve different experimental objectives *(15,16)*; however, we will focus on the three types of designs presented in Fig. 3 because they are the most obvious choices for class comparisons. Other types of designs to consider are presented in the dye bias section.

Which Design Will Provide the Best Class Comparisons?

Balanced block, loop, and reference design experiments can all provide unbiased estimates of the differences in gene expression levels between class means, i.e., differences between the average gene expression levels. However, the three designs are not equally efficient. The efficiency of a design is based on the precision of the statistical estimates of the differences in the class means for "equivalent experiments." We define two notions of equivalent experiments that we think are appropriate to many microarray studies: 1) Two experiments are equivalent if they use the same number of microarrays, and 2) two experiments are equivalent if they use the same (nonreference) samples and subsamples.

Definition 1 is appropriate when nonreference RNA samples are abundant and the limiting factor is the amount of time or resources required to actually run the arrays. The question then might be "If I can afford to run only 20 arrays, how should I design the experiment?" Definition 2 is appropriate when the nonreference RNA resources are scarce and the cost of running the arrays is less critical. The question then might be "Given that I have only these 12 RNA samples, how should I design the experiment?"

Efficiency comparisons of the three designs for a typical experiment (with biologic variation twice that of the experimental error variation) calculated from equations presented in Dobbin and Simon *(9)* are shown in Fig. 4, A. When the number of microarrays is limited (equivalence definition 1), then the balanced block design is substantially more efficient than the reference or the loop designs. However, the efficiency gain with the balanced block design comes with some sacrifice, including robustness and difficulty in clustering samples.

When the nonreference RNA samples are limited (equivalence definition 2), then the efficiencies of the reference and balanced block design are similar (Fig. 4, B). The loop design is less efficient than the balanced block design and also suffers from the same lack of robustness. The more robust reference design appears to be better overall than the other two designs when nonreference RNA samples are limited.

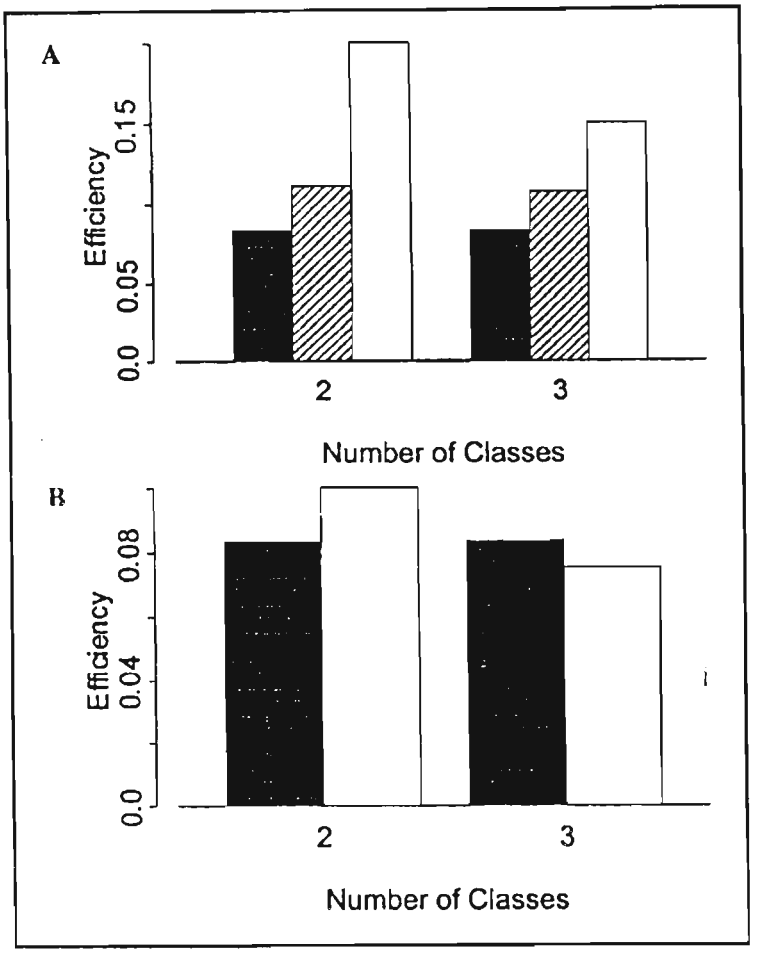


Fig. 4. Comparison of design efficiencies. A) Comparison of design efficiencies for the reference (**solid bars**), loop (**hatched bars**), and balanced block (**open bars**) designs when the number of arrays is fixed. B) Comparison of the reference (**solid bars**) and balanced block (**open bars**) designs when the nonreference RNA samples are fixed. Efficiency is the inverse of the variance of the estimated difference between the class averages. Some parameters, such as the biologic and experimental variation, were fixed to construct the display. Results are general in that the specific number of arrays or samples used does not affect the relationship between the heights of the histogram bars. The loop design was not included in the histogram because it uses a different sampling scheme. For further details about mathematical equations, refer to online supplemental information (*see* http//jncicancerspectrum.oupjournals.org/jnci/content/vol95/issue18/index.shtml).

What Happens If the Class Definitions Change?

It is not unusual to have different classifications of the samples or to have corrections in the class of specific samples. The reference design is more robust to changes in the classification scheme than either the balanced block or loop designs. The reason for this increased robustness is that the reference design will remain a reference design with a new classification. In contrast, the balanced block design will probably lose its structure (i.e., it will no longer be a balanced block design). With regard to a new classification, many arrays may contain two samples from the same class, which can result in a severe loss of efficiency. It is also possible that, with a new classification, the classes cannot be compared with the balanced block design because they never appear together on any arrays. The loop design

is also subject to large efficiency loss, because under a new classification, the classes may appear together only on a small proportion of arrays.

What If We Also Plan to Perform Class Discovery on the Samples?

Class discovery is the process of finding a new classification system for a set of biologic samples on the basis of gene expression profiles when the class labels are unknown ahead of time. Cluster analysis is the most appropriate approach to use in class discovery. Of the three designs presented in Fig. 3, effective class discovery can only be performed for the reference and loop designs. Individual samples must be compared in class discovery. The balanced block design confounds spot variability with comparison of samples on different arrays because no RNA sample appears on more than one array. The arrows connecting the samples repeated on different slides in the reference and loop designs indicate why this type of confounding is not a problem in these designs—that is, connections can be made between any two samples on different arrays using the arrows.

The reference design is recommended for class discovery because cluster analysis can perform substantially better with a reference design than a loop design *(9)*, particularly as the number of samples increases. An example of a cluster analysis for 10 and 20 samples, which was originally presented in Dobbin and Simon *(9)*, is presented in Fig. 5. The data in that figure were generated from two true clusters (i.e., the data in each cluster were generated from a different mean gene expression vector). The number of discrepancies between the clusters found by a common cluster analysis algorithm and the true clusters for the reference and loop designs appear on the *x*-axis. The reference design finds the true clusters almost every time, whereas the loop design performs poorly for 10 samples and much worse for 20 samples. Moreover, the loop design performs even worse when there are more than 20 samples *(9)*. The difference in cluster analysis performance is so dramatic that it will usually offset any relatively moderate differences in efficiency and power between the loop and reference designs. For this reason, we recommend using the reference design for class discovery experiments.

What Is Sacrificed If a Reference Design Is Not Used?

Most investigators are familiar with the reference design, and they may want to know what will be sacrificed if an alternative design such as the balanced block design is used. In addition to the issues discussed in the last two questions, there are other considerations worth mentioning. 1) The data from a balanced block or loop design may be more difficult to analyze than data from a reference design. Most microarray analysis packages assume a reference design has been used, so analyzing the experiment may require switching to different software. 2) The balanced block or loop design may be more difficult to devise than the reference design. If there are many groups being compared or many possible ways to group the samples, designing the study so that all comparisons of interest can be made may be nontrivial. 3) It may not be possible to compare data from different microarray experiments or prospective data that is analyzed by microarrays at different times. If a common reference sample is used for all experiments, then there is some foundation for the comparison of samples collected over time or samples analyzed in different experiments, a situation that is generally not possible for balanced block or loop designs.

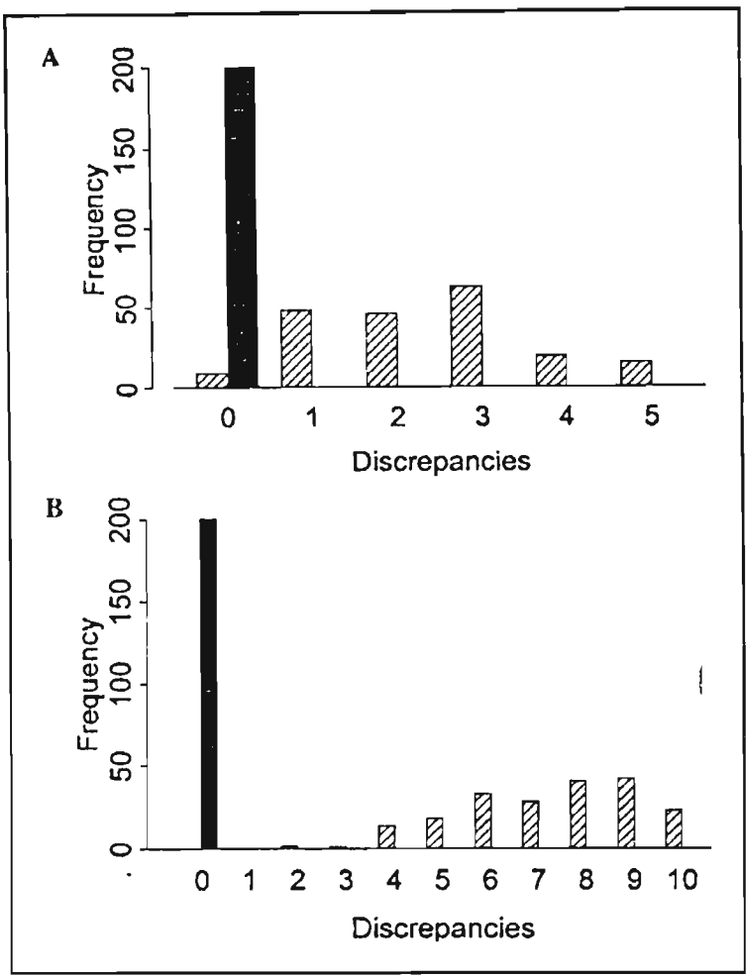


Fig. 5. Comparison of cluster analysis performance. Comparison of cluster analysis performance for the reference (**solid bars**) and loop (**hatched bars**) designs on A) 10 samples and B) 20 samples. Simulated data comes from two true (each with a different mean gene expression) clusters. One thousand genes were present in the clusters, 20 of which were differentially expressed. *x*-axis is the number of discrepancies between true clusters and closest matches. *y*-axis is the frequency of the number of discrepancies observed in 200 simulations. Simulation was based on a prostate cancer dataset [*see* Dobbin and Simon *(9)* for details].

Dye Bias

What Is the Source of Dye Bias?

Cy3 and Cy5 have different efficiencies for their labeling ability and detection characteristics. Background correction and normalization adjust for consistent dye-related differences that are not gene-specific. For example, median centering of arrays is meant to eliminate bias that is common across all genes, and intensity-dependent normalization, such as loess smoothers, adjust for bias related to overall spot intensity *(15)*. Gene-specific dye bias is displayed by genes that do not fall into the overall pattern of the dye effect that characterizes the majority of genes. This bias may persist even after normalization.

Does Gene-Specific Dye Bias Exist?

To our knowledge, there has been no definitive study characterizing the nature or magnitude of gene-specific dye bias. In addition, it is not clear that gene-specific dye bias is the same

from one experiment to another or from one laboratory to another, but it is of general concern among microarray investigators. Many studies *(3,18–22)* have been designed to guard against gene-specific dye bias, whereas others *(8,17,23–25)* have made gene-specific dye bias adjustments to their statistical analysis. Some studies have attempted to eliminate gene-specific dye bias through technical innovations in labeling *(7,23,26,27)*. Although novel labeling procedures such as indirect labeling appear to reduce gene-specific dye bias, it is not clear that they eliminate dye bias.

We have observed gene-specific dye bias and provide, as an example, one reference design experiment involving transgenic mice (Green J: unpublished data). Nine distinct RNA samples from nine mice were examined, and three of these samples were run twice, once with each dye label (i.e., once with the reference labeled with Cy3 and once with the reference labeled with Cy5), for a total of 12 arrays. The intensity data were first background-adjusted to eliminate stray fluorescence signals from the slide and normalized to make the measurements on different arrays comparable. We then performed an ANOVA on the individual channel log intensities. An ANOVA model was fit separately to data for each of 8832 genes. In the ANOVA approach, the dye bias effects are called dye-by-gene interactions. Overall, we observed that there were many genes with a statistically significant dye-by-gene interaction ($P<.001$), but these effects tended to be small. The size of these effects on the base 2 log-scale is shown in Fig. 6. The average absolute value of the gene-by-dye interactions was 0.18 (standard deviation = 0.16), corresponding to a 1.13-fold change in gene expression levels. Only 10 of the 8832 genes had dye bias that corresponded to a twofold or greater change in gene expression levels. Tseng et al. *(8)* have presented similar results. Although dye bias appears to be common in these direct-labeled cDNA experiments, it appears to be fairly small in magnitude.

When Is Gene-Specific Dye Bias an Issue?

Gene-specific dye bias is a potential issue when comparisons are made between samples labeled with different dyes. Hence, it is not generally a problem in reference design experiments because they compare classes of nonreference RNA samples. Because all of the nonreference RNA samples are labeled with the same dye, the dye bias between the nonreference and reference intensities does not become a bias in comparing classes. Gene-specific dye bias is a potential problem, however, if nonreference RNA samples are compared with a common reference RNA sample. Gene-specific dye bias is also an issue for balanced block and loop designs. When gene-specific dye bias is an issue, its magnitude must be estimated for each gene, and an explicit adjustment to the statistical analysis must be made to ensure that class comparisons are unbiased. For example, in ANOVA analysis, the adjustment involves adding terms representing gene-specific dye bias to the statistical model.

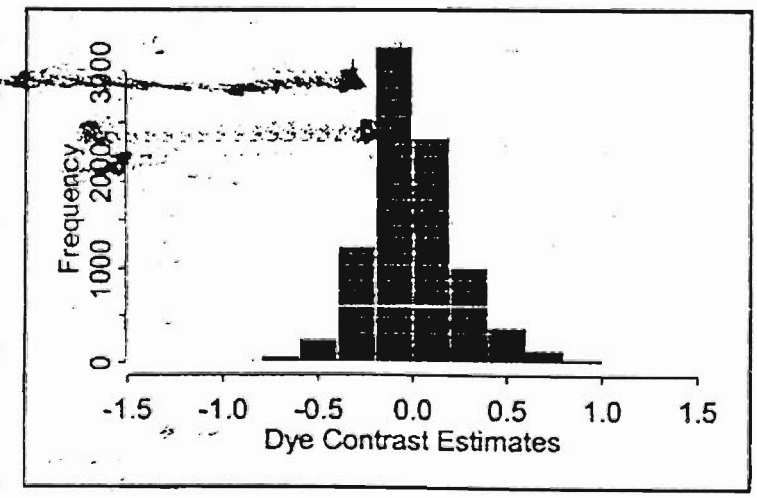


Fig. 6. Estimated dye bias contrast that was not corrected for in normalization. Estimates for dye bias were based on 8832 genes from a transgenic mouse experiment. Data were transformed to base 2 logarithms so that an estimated dye bias contrast of size 1 corresponds to a twofold change in gene expression.

How Should I Design an Experiment to Eliminate Dye Bias From the Class Comparisons?

Dye bias can be eliminated from the class comparisons in two ways: 1) by labeling all samples from all classes being compared with the same dye, and 2) by labeling half the samples with one dye and half the samples with the other dye for each class being compared.

Reference designs usually use strategy 1 to eliminate dye bias. Other designs, such as balanced block designs, often use strategy 2. Labeling exactly half the samples of a class with a dye is preferable to labeling some other fraction because it produces more accurate class comparisons and is simpler to analyze. If there is an odd number of nonreference RNA samples from each class (e.g., seven), then strategy 2 would not be able to be followed exactly (e.g., three samples labeled with red dye [Cy5] and four samples labeled with green dye [Cy3]). Dye bias can still be eliminated from such a design, but it requires a more complex weighted analysis to adjust for the dye asymmetry.

Another approach that is sometimes used to eliminate dye bias is to run a set of arrays with the reference in both channels to identify the genes that display dye bias. These genes could then be flagged as suspect if they show up as statistically significant in the class comparisons.

Some investigators *(12)* have used the existence of dye bias as a reason to run all sample pairs twice, once with each dye, to eliminate the bias. However, we *(10)* have shown that complete dye swapping is an inefficient way to adjust for the dye bias correction. If each sample is run twice in a fixed number of arrays, then the effective sample size is cut in half. The reference design or balanced block design will provide unbiased estimates of the class comparison without running any sample pairs twice. Hence, the complete dye-swapping strategy effectively halves the sample size and reduces the efficiency with no real gain as far as class comparisons are concerned. Balancing the classes with respect to the dyes is more efficient than dye swapping of individual samples for eliminating dye bias.

How Will Class Discovery Results Be Affected by Dye Bias?

Dye bias generally will not have a substantial impact on class discovery, although it may be necessary to make an explicit dye bias adjustment. In this commentary, we have focused on class comparison experiments in which we already have class labels for the samples. Class discovery can be performed on all the samples or on only the samples within a particular class. Class discovery using cluster analysis on all of the samples is sometimes performed to verify that the resulting clusters recapitulate the known classes *(28,29)*. In addition, cluster analysis within a

presence of a blocking factor to microarray experiments *(9)* and have established that, for many class comparison studies, the balanced block design shown in Fig. 3 is optimal. The effect of spot-to-spot variation in gene expression levels is eliminated in the balanced block design because each gene's expression level is measured at the same spot on the same array for samples from each of the two classes being compared.

The third type of design that might be considered for cDNA microarrays is one proposed by Kerr and Churchill *(15)*, which they called a loop design (Fig. 3). Unlike the two other designs, the loop design requires two aliquots from each RNA sample. These aliquot pairs connect the arrays and are arranged so that the connected arrays form a loop pattern.

Class comparisons for the balanced block design and the loop design are accomplished by fitting an analysis-of-variance (ANOVA) model to the logarithm of the background-corrected channel-specific intensities *(9)* and fitting a separate model for each gene. This approach can also be used for analysis of the reference design, but the results are equal to or very similar to applying simple Student's *t* tests to the log-ratio measurements.

More elaborate designs have been proposed to achieve different experimental objectives *(15,16)*; however, we will focus on the three types of designs presented in Fig. 3 because they are the most obvious choices for class comparisons. Other types of designs to consider are presented in the dye bias section.

Which Design Will Provide the Best Class Comparisons?

Balanced block, loop, and reference design experiments can all provide unbiased estimates of the differences in gene expression levels between class means, i.e., differences between the average gene expression levels. However, the three designs are not equally efficient. The efficiency of a design is based on the precision of the statistical estimates of the differences in the class means for "equivalent experiments." We define two notions of equivalent experiments that we think are appropriate to many microarray studies: 1) Two experiments are equivalent if they use the same number of microarrays, and 2) two experiments are equivalent if they use the same (nonreference) samples and subsamples.

Definition 1 is appropriate when nonreference RNA samples are abundant and the limiting factor is the amount of time or resources required to actually run the arrays. The question then might be "If I can afford to run only 20 arrays, how should I design the experiment?" Definition 2 is appropriate when the nonreference RNA resources are scarce and the cost of running the arrays is less critical. The question then might be "Given that I have only these 12 RNA samples, how should I design the experiment?"

Efficiency comparisons of the three designs for a typical experiment (with biologic variation twice that of the experimental error variation) calculated from equations presented in Dobbin and Simon *(9)* are shown in Fig. 4, A. When the number of microarrays is limited (equivalence definition 1), then the balanced block design is substantially more efficient than the reference or the loop designs. However, the efficiency gain with the balanced block design comes with some sacrifice, including robustness and difficulty in clustering samples.

When the nonreference RNA samples are limited (equivalence definition 2), then the efficiencies of the reference and balanced block design are similar (Fig. 4, B). The loop design is less efficient than the balanced block design and also suffers from the same lack of robustness. The more robust reference design appears to be better overall than the other two designs when nonreference RNA samples are limited.

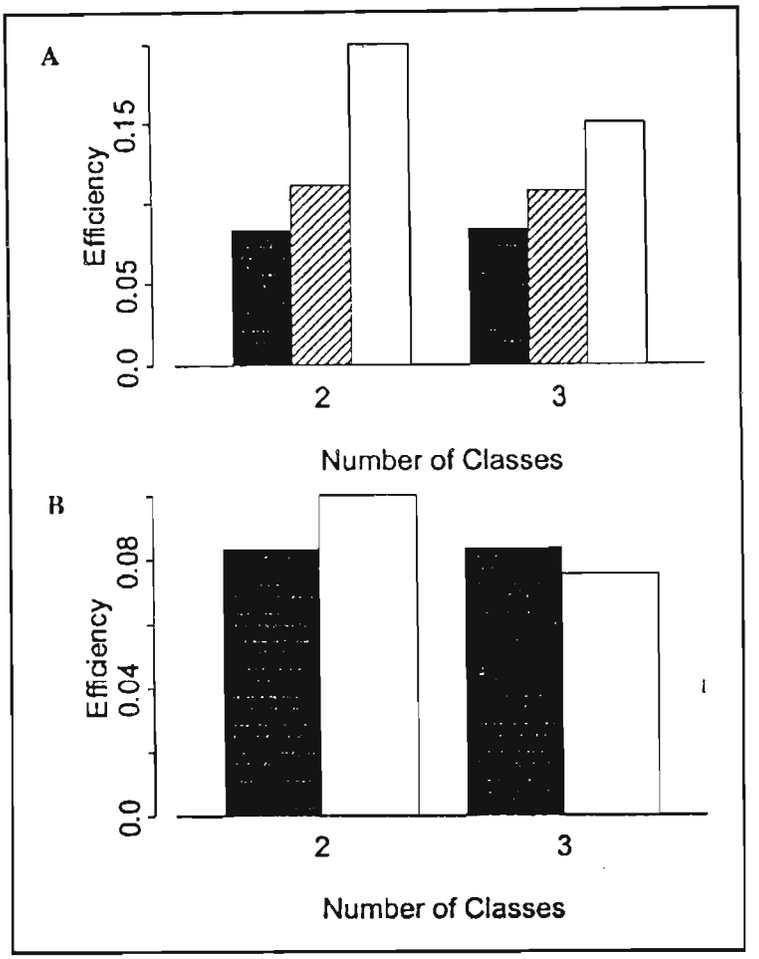


Fig. 4. Comparison of design efficiencies. **A)** Comparison of design efficiencies for the reference (**solid bars**), loop (**hatched bars**), and balanced block (**open bars**) designs when the number of arrays is fixed. **B)** Comparison of the reference (**solid bars**) and balanced block (**open bars**) designs when the nonreference RNA samples are fixed. Efficiency is the inverse of the variance of the estimated difference between the class averages. Some parameters, such as the biologic and experimental variation, were fixed to construct the display. Results are general in that the specific number of arrays or samples used does not affect the relationship between the heights of the histogram bars. The loop design was not included in the histogram because it uses a different sampling scheme. For further details about mathematical equations, refer to online supplemental information (*see* http//jncicancerspectrum.oupjournals.org/jnci/content/vol95/issue18/index.shtml).

What Happens If the Class Definitions Change?

It is not unusual to have different classifications of the samples or to have corrections in the class of specific samples. The reference design is more robust to changes in the classification scheme than either the balanced block or loop designs. The reason for this increased robustness is that the reference design will remain a reference design with a new classification. In contrast, the balanced block design will probably lose its structure (i.e., it will no longer be a balanced block design). With regard to a new classification, many arrays may contain two samples from the same class, which can result in a severe loss of efficiency. It is also possible that, with a new classification, the classes cannot be compared with the balanced block design because they never appear together on any arrays. The loop design

particular class is sometimes used to identify novel subclasses *(28,30,31)*.

In the previous section we discussed two designs for class discovery—the reference design and the loop design. Dye bias generally will not affect class discovery for the reference design because all the samples being clustered are labeled with the same dye. The effect of dye bias on cluster analysis results can also be eliminated from the loop design by making a dye bias adjustment; however, we do not recommend this design because of its poor cluster analysis performance, as discussed in the previous section.

How Can Dye Bias Be Eliminated From Comparisons Between the Reference and the Nonreference Samples in a Reference Design?

One can eliminate dye bias from the comparisons between the reference and nonreference samples by including dye-swapping arrays in the design of the experiment. Consider a reference design experiment used to study a collection of tumor samples, where the reference sample consists of a mixture of normal tissue. A fairly common experimental situation is one in which the primary goal is to perform class discovery on the tumors and the secondary goal is to compare the tumors with the normal reference to identify potential tumor markers *(32,33)*. Because the normal reference sample is labeled with a different dye than the tumor samples, there is potential for dye bias in the comparisons. In this case, we recommend appending the basic reference design with just enough dye-swapping arrays to allow for good statistical inference for the comparison with the reference sample. This comparison is made by ANOVA and is adjusted for dye bias; an example of such a design is shown in Fig. 7. Note, we do not recommend reversing all the arrays in this situation, because running all samples both forward and backward with the reference sample substantially reduces the efficiency of the tumor versus normal comparison (for a fixed number of arrays) and hinders the ability of the cluster analysis to identify true groupings in the gene expression data. Running dye-swapping arrays on all samples essentially sacrifices the primary goal of discovering a new taxonomy for the secondary goal of identifying potential markers; even for the secondary goal, complete dye swapping is inefficient in most cases.

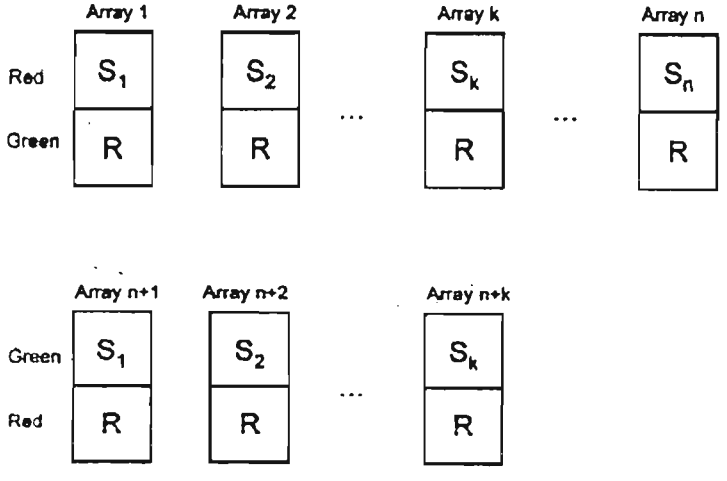


Fig. 7. Dye-swapping reference design for clustering and comparison of nonreference with reference RNA samples. Rectangles represent the arrays. S_1 is sample 1 from the nonreference samples, S_2 is sample 2 from the nonreference samples, and so on up to some numbered sample n (S_n). R is the reference sample. Of the n + k arrays, k is run as a dye swap on repeated samples. The first row of arrays represents the forward arrays and second row of arrays represents the reverse arrays. The reference sample is dyed green (Cy3) on the forward arrays and red (Cy5) on the reverse arrays. The resulting fixed-effects analysis of variance table has k −1 degrees of freedom for error.

Sample Size

How Many Biologic Samples Are Needed for a Reference Design?

Suppose we want to test whether a particular gene is differentially expressed in two classes. To test the null hypothesis that there is no difference in gene expression levels at the α significance level, we want to have $1 - \beta$ power to detect a difference of δ in the class mean log-ratios. Let σ be the standard deviation of the log-ratios within each class and n be the total number of arrays used, i.e., $n/2$ arrays for each class. Then the usual sample size formula *(34)*, based on an assumption of normal distributions within the classes, would be:

$$n = \frac{4(z_{1-\alpha/2} + z_{1-\beta})^2}{(\delta/\sigma)^2}$$

The notation $z_{1-\alpha/2}$ indicates the $100(1 - \alpha/2)^{th}$ percentile of the standard normal distribution. When the samples sizes are small, the normal approximation of the test statistic may be poor, and an iterative computational procedure based on the t distribution can be used to compute the sample size. For example, we have observed an $\sigma \approx .50$ for human cancer data using log base 2 intensities on cDNA microarrays and a reference design, and we have observed $\sigma \approx .25$ with data from inbred strains on transgenic mice *(9)*. A $\delta = 1$ corresponds to a twofold difference in gene expression. Setting $\alpha = .001$ guards against an excessive number of false-positive genes. For example, with 10 000 genes, $\alpha = .001$ results in an average of 10 false-positive genes. Setting $\beta = .05$ provides 95% probability of detecting a twofold change in gene expression. The resulting sample size is then 30 total samples for $\sigma = .50$ and 12 total samples for $\sigma = .25$. Because of the small sample sizes, we have used t distribution percentiles in both cases.

What Sample Size Should Be Used for a Balanced Block Design?

Suppose that two classes will be compared and that the samples from each class are independent. Again, we want to test the null hypothesis that there is no difference in gene expression levels between the classes at the α significance level and to have $1 - \beta$ power to detect a difference of δ in the class means. Let τ be the standard deviation of the log-ratios. In the balanced block design, each log-ratio involves two independent samples, one from each class. The τ parameter will tend to be larger than the σ parameter in the reference design because additional biologic variation is displayed in the log-ratios. Let n be the total number of arrays used, i.e., n arrays with n samples from each class. The sample size formula would now be:

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{(\delta/\tau)^2}$$

Because the reference sample appears on each array in the reference design, variability among the log-ratios will be smaller for a reference design than for a balanced block design. We provide

on-line supplemental material (*see* http//jncicancerspectrum.oupjournals.org/jnci/content/vol95/issue18/index.shtml for details) that shows how prior data from a reference design experiment can be used to estimate τ. For example, using our estimated standard deviation of the log-ratios from the reference design that used human samples ($\sigma = .50$) and the same parameter settings that we used for the reference design sample size calculation ($\delta = 1$, $\alpha = .001$, $\beta = .05$) results for $\tau \approx .67$, the sample size required changes from 30 arrays under the reference design to 17 arrays under the balanced block design. The reference design uses 30 arrays from 30 total samples, 15 from each class, whereas the balanced block design uses 13 fewer arrays but requires 17 samples from each class, or a total of 34 samples.

REFERENCES

(1) Golub T, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531–7.

(2) Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, et al. Gene-expression profiles in hereditary breast cancer. N Engl J Med 2001; 344:539–48.

(3) Zhou Y, Gwadry FG, Reinhold WC, Miller LD, Smith LH, Scherf U, et al. Transcriptional regulation of mitotic genes by camptothecin-induced DNA damage: microarray analysis of dose- and time-dependent effects. Cancer Res 2002;62:1688–95.

(4) Bonham MJ, Galkin A, Montgomery B, Stahl WL, Agus D, Nelson PS. Effects of the herbal extract PC-SPES on microtubule dynamics and paclitaxel-mediated prostate tumor growth inhibition. J Natl Cancer Inst 2002; 94:1641–7.

(5) Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. Nat Genet 1999;21(1 Suppl):10–4.

(6) Lockhart DJ, Winzeler EA. Genomics, gene expression and DNA arrays. Nature 2000;405:827–36.

(7) Manduchi E, Scearce LM, Brestelli JE, Grant GR, Kaestner KH, Stoeckert CJ Jr. Comparison of different labeling methods for two-channel high-density microarray experiments. Physiol Genomics 2002;10:169–79.

(8) Tseng GC, Oh M, Rohlin L, Liao JC, Wong WH. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. Nucleic Acids Res 2001;29:2549–57.

(9) Dobbin K, Simon R. Comparison of microarray designs for class comparison and class discovery. Bioinformatics 2002;18:1438–45.

(10) Dobbin K, Shih JH, Simon R. Statistical design of reverse dye microarrays. Bioinformatics 2003;19:803–10.

(11) Lee MT, Kuo FC, Whitmore GA, Sklar J. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci U S A 2000;97: 9834–9.

(12) Goryachev AB, MacGregor PF, Edwards AM. Unfolding of microarray data. J Comput Biol 2001;8:443–61.

(13) Cochran WG, Cox GM. Experimental designs. 2nd ed. New York (NY): Wiley; 1992. p. 95–182.

(14) Scheffé H. The analysis of variance. New York (NY): Wiley; 1999. p. 55–146.

(15) Kerr MK, Churchill GA. Experimental design for gene expression microarrays. Biostatistics 2001;2:183–201.

(16) Yang YH, Speed T. Design issues for cDNA microarray experiments. Nat Rev Genet 2002;3:579–88.

(17) Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, Nucleic Acids Res 2002;30:e15.

(18) Bayani J, Brenton JD, Macgregor PF, Beheshti B, Albert M, Nallainathan D, et al. Parallel analysis of sporadic primary ovarian carcinomas by spectral karyotyping, comparative genomic hybridization, and expression microarrays. Cancer Res 2002;62:3466–76.

(19) Klebes A, Biehs B, Cifuentes F, Komberg TB. Expression profiling of Drosophila imaginal discs. Genome Biol 2002;3:RESEARCH0038.

(20) Aharoni A, Keizer LC, Bouwmeester HJ, Sun Z, Alvarez-Huerta M, Verhoeven HA, et al. Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell 2000;12:647–62.

(21) Barrans JD, Allen PD, Stamatiou D, Dzau VJ, Liew C. Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular-based cDNA microarray. Am J Pathol 2002;160:2035–43.

(22) Desai KV, Xiao N, Wang W, Gangi L, Greene J, Powell JI, et al. Initiating oncogenic event determines gene-expression patterns of human breast cancer models. Proc Natl Acad Sci U S A 2002;99:6967–72.

(23) Yu J, Othman MI, Farjo R, Zareparsi S, MacNee SP, Yoshida S, et al. Evaluation and optimization of procedures for target labeling and hybridization of cDNA microarrays. Mol Vis 2002;8:130–7.

(24) Kerr MK, Churchill GA. Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. Proc Natl Acad Sci U S A 2001;98:8961–5.

(25) Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, et al. Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol 2001;8:625–37.

(26) Wilson AS, Hobbs BG, Speed TP, Rakoczy PE. The microarray: potential applications for ophthalmic research. Mol Vis 2002;8:259–70.

(27) Stears RL, Getts RC, Gullans SR. A novel, sensitive detection system for high-density microarrays using dendrimer technology. Physiol Genomics 2000;3:93–9.

(28) Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, et al. Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci U S A 2001;98:13784–9.

(29) Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A 2001;98:13790–5.

(30) Beer DG, Kardia SL, Huang C, Giordano TJ, Levin AM, Misek DE, et al. Gene-expression profiles predict survival in patients with lung adenocarcinoma. Nat Med 2002;8:816–24.

(31) Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503–11.

(32) Lin Y, Furukawa Y, Tsunoda T, Yue C, Yang K, Nakamura Y. Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 2002;21:4120–8.

(33) Jazaeri AA, Yee CJ, Sotiriou C, Brantley KR, Boyd J, Liu ET. Gene expression profiles of BRCA1-linked, BRCA2-linked, and sporadic ovarian cancers. J Natl Cancer Inst 2002;94:990–1000.

(34) Desu MM, Raghavarao D. Sample size methodology. Boston (MA): Academic Press; 1990. p. 30.

NOTES

We thank Jeff Green and Nianqing Xiao for the transgenic mouse data used for our analysis of gene-specific dye bias.

Manuscript received February 10, 2003; revised July 9, 2003; accepted July 16, 2003.