Molecular Diversity, **7**: 25–36.
KLUWER/ESCOM

Full paper

Development of chemically modified glass surfaces for nucleic acid, protein and small molecule microarrays

László Hackler Jr.[1], György Dormán[2], Zoltán Kele[3], László Ürge[2], Ferenc Darvas[2,4] & László G. Puskás[1]*

[1] *Laboratory of Functional Genomics, Biological Research Center, Hungarian Academy of Sciences, Szeged, Temesvári krt. 62;* [2] *ComGenex Inc, H-1027 Budapest, Bem rkp 33-34;* [3] *Department of Medicinal Chemistry, University of Szeged, Dóm tér 8, H-6720 Szeged, Hungary;* [4] *ComGenex International, South San Francisco, California 94080, 705 Grandview Drive*

(Author for correspondence, E-mail: pusi@nucleus.szbk.u-szeged.hu)*

Received 20 March 2003; Accepted 13 June 2003

Key words: immobilization, ligand-chip, microarray, protein microarray

Summary

Microarrays have become a widely used tool to investigate the living cell at different levels. DNA microarrays enable the expression analysis of thousand of genes simultaneously, while protein arrays investigate the properties and interactions of proteins with other proteins and with non-proteinaceous molecules. One crucial step in producing such microarrays is the permanent immobilization of samples on a solid surface. Our goal was to develop diverse linker systems capable of anchoring different biological samples, especially DNA and drug-like small molecules. We developed 6 different chemical surfaces having a 3-D-like linker system for biomolecule immobilization, and compared them to previously described immobilization strategies. The attachment chemistry utilizes the amino reactive properties of acrylic and epoxy functions. The capacity of the support was increased by creating a branching structure holding the reactive functions. The method of anchoring was investigated through a model reaction. From HPLC and mass spectrometry measurements we concluded that the covalent binding of DNA occurs through nucleobases. The tested systems offer the capability to permanently immobilize several biomolecular species in an array format.

Introduction

The idea of deposition and immobilization of thousands of samples in an array format has led to the development of several microarray techniques utilizing different biological samples like cDNA, presynthesized oligodeoxyribonucleotides, proteins, tissues and small molecules from combinatorial chemical libraries. Immobilized DNA molecules, either as double-stranded segments or as short synthetic oligodeoxyribonucleotides, have been utilized for gene expression monitoring, DNA-sequencing, disease screening, diagnostics and genome analysis. Microarrays containing DNA molecules can be divided into two groups: DNA microarrays and oligonucleotide microarrays (oligo-arrays). On the surface of a DNA microarray double stranded DNA molecules – obtained from PCR amplification – are the immobilized samples, while oligo arrays utilize presynthesized or *in situ* synthesized oligonucleotides [1–11]. Protein arrays containing numerous different entities are efficient tools in applications like antibody profiling, serum screening and, ultimately, in drug discovery [12, 13]. Newly emerging platform of microarrays are chemical microarrays. These arrays contain thousands of drug-like small molecules from parallel combinatorial synthesis and they enable the identification of new, and the screening of existing, lead molecules [14]. All these techniques demand

permanent and efficient anchoring of the samples. A great number of attachment methods have been patented or published, which vary widely in chemical mechanism, chemical structure of linker system, ease of use, attachment stability and cost. Biomolecules can be attached to solid surfaces covalently or non covalently. Non-covalent anchoring methods utilize the electrostatic interactions between the surface and the biomolecules through secondary chemical bonds (e.g. ionic interactions and hydrogen bonds). Standard methods for covalent attachment of oligonucleotides onto solid surfaces include reactive aldehyde-, thiol- [15–17], epoxy-groups, where chemically modified nucleic acids are needed for their attachment. A number of different substances have been tested as the solid support for the attachment of nucleic acids [18–19], but glass slides are generally favored for DNA and oligonucleotide arrays [20–21].

The goal of our work was to develop derivatized surfaces capable of anchoring different species of biomolecules, especially DNA and small organic compounds, and to apply them in microarray experiments.

Materials and methods

Support media

As starting material, commercially available non-derivatized microscope slides were used (Spektrum 3-D, Debrecen, Hungary). Reactions were carried out in polypropylene or glass chambers. For comparison, Superaldehyde (ArrayIt, Telechem) aldehyde (Genetix), Epoxy silane (Erie Scientific Company), and poly-L-lysine slides (Erie Scientific Company) were used.

Developing the reactive surfaces

The final reactive surfaces were obtained in multiple reaction steps. The final reaction step differs in case of producing either acryl or epoxy derivatized surfaces. Epoxy functions were introduced with two different reagents.

Dendrimer-like linker systems

The non-derivatized glass slides were etched in 10% NaOH for 24 h, then washed with water. After the alkaline treatment the slides were neutralized with 1% HCl for 1 min, washed again with water than dried. The activated glass slides were treated with 3% methacryloxypropyl-trimethoxysilane (ICN Biomedicals Inc.) solution in 95% methanol for 2 h. The slides were subsequently washed with methanol, then water, dried and baked at 105 °C for 15 min. The acryl silanized slides were incubated for 48 h in a 70 ml solution of 0.033 mmol ml^{-1} tetraethylenepentamine in dimethylformamide (DMF). The slides were then washed with DMF, methanol and dried. For preparing acrylic surfaces, the slides were treated with 9.2 mmol acryloyl-chloride and 9.2 mmol diisopropylethyl-amine (ICN. Biomedicals Inc.) in anhydrous dichloroethane (ALFA) for 2 h, then washed with dichloroethane and dried. For preparing hydrophobic epoxy surfaces the slides were treated with 30 mmol epichlorohydrin and 12 mmol pyridine in chloroform, then washed with chloroform and dried. For preparing hydrophilic epoxy surfaces the slides were treated with 30 mmol 1,4-butanediol diglycidyl ether in the presence of 5 mmol NaOH in ethanol, then washed with ethanol and dried.

Linker systems based on triamino-silanized surface

The non-derivatized glass slides were etched in 10% NaOH for 24 h, then washed with water, neutralized with 1% HCl, and washed again with water then dried. The activated glass slides were treated with 3% 3-[2-(2-Aminoethylamino)ethylamino)propyl-trimethoxysilane (Fluka) solution in 95% methanol for 2 h. The slides were then washed with methanol, then water, dried and baked at 105 °C for 15 min. For preparing acrylic surfaces the silanized slides were incubated for 2 h with 30 mmol acryloyl-chloride (Fluka) and 30 mmol diisopropylethyl-amine (ICN. Biomedicals Inc.) in dichloroethane. The slides were subsequently washed with dichloroethane, methanol and dried. For preparing hydrophobic epoxy surfaces the silanized slides were incubated for 2 h with 30 mmol epichlorohydrin (Fluka) and 12 mmol pyridine (Fluka) in ethanol (Molar). The slides were then washed with chloroform and dried. For preparing hydrophilic epoxy surface the silanized slides were incubated for 2 h with 30 mmol 1,4-butanediol diglycidyl ether (Fluka) and 5 mmol NaOH in ethanol (Molar). The slides were then washed with chloroform and dried.

Spotting onto activated surface

Spotting was performed using a pin-tool based spotting robot MicroGrid Total Array System (BioRobotics, UK) or pipetting manually. When using automated spotting, 1–10 nl was introduced onto the surface us-

ing split or solid pins, respectively. When applied manually 0,5 or 1 μl of solution was introduced onto the surface. After spotting, the slides were incubated in a humid chamber for 2 h. In comparison experiments the commercially available slides were treated following the manufacturer's instructions. The slides were stored at room temperature in the dark.

The pH characteristics of immobilization to the novel acryl derivatized surface

The optimal parameters of immobilization were determined using fluorescently labeled 5'-Cy5 oligonucleotides as subjects of anchoring. Spotting was carried out manually in 0.5 μl aliquots. Two dilutions of oligonucleotide solution, 0.25 pmol μl^{1} and 0.062 pmol μl^{-1} respectively, were introduced onto the surface at seven pH values (pH 4, pH5, pH6, pH 7, pH 8, pH 9, pH 10) in 0.1 M phosphate buffer. The slides were incubated for 2 h in a humid chamber and UV irradiated with a UV Crosslinker (700 mJ, Ultra Lum), then extensively washed with deionized water and dried. The slides were scanned with a ScanArray Lite (GSI Lumonics, Billerica, U.S.A.) microarray scanner at 543 nm. Results are shown in Figure 3.

Mutation detection by hybridization

Oligonucleotides having 0, 1, 2, or 3 alterations were arrayed on the surfaces at 50 pmol μl^{-1} concentration. Arrays were hybridized with complementary Cy5-labeled oligonucleotide (0.1 pmol μl^{-1}) for 2 h then washed, dried and scanned (see above). The applied oligonucleotide sequences are presented in Table I.

Hybridization detection of PCR products

Different species-specific, randomly sequenced (negative control) and Cy5-labeled control oligonucleotides (Table I) were printed on the acryl derivatized dendrimer-like surface and on a commercially available aldehyde coated support (ArrayIt), using solid pins and a BioRobotics spotter. Specific PCR products were hybridized for 2 h to the arrays in 20 μl Huntsman hybridization buffer (50% formamide, 5x SSC, 0.1% SDS, 100 μg ml^{-1} salmon sperm DNA). The arrays were washed, dried and scanned as described above. Sequences for human cytomegalovirus (HCMV) were designed according to HCMV major immediate-early protein (IE) gene, complete cds, Accession number: M21295. PCRs were carried out by using an unmodified forward and a Cy5-end-labeled reverse primer (Table I) with the following protocol: 360 sec heat start, 45 cycles of denaturation at 95 °C for 30 sec, annealing at 58 °C for 30 sec and extension at 72 °C for 40 sec. PCR products were analyzed with agarose gel-electrophoresis and ethidium-bromide staining. PCR products were denatured with incubation at 95 °C for 6 min before applying onto the arrays. Both the acrylic-based and triamino-based chemically modified surfaces were tested in hybridization experiments.

Protein immobilization

1 μl of 0.5 mg ml^{-1} Alexa Fluor 546 labeled streptavidin (Molecular Probes, Leiden, The Netherlands) solution in pH10 phosphate buffer was introduced onto the surfaces. The same protein solution without pH buffer was also spotted. The slides were incubated for 30 min in a humid chamber and washed with water then 1x SSC, 0.15% SDS for 5 min. The slides were scanned after each washing step.

Mimicking surface immobilization

In order to determine the way biomolecules are immobilized on the acryl derivatized surface, a model experiment was designed. Saturated aqueous solutions of nucleosides (cytidine, adenosine, guanosine, thymidine) in 10 x excess were incubated with acrylamide at 37 °C. After 2 h the reaction mix was UV irradiated with 2 × 700 mJ in a UV crosslinker (Ultra Lum), and analyzed by reverse phase HPLC and mass spectrometry.

HPLC analysis

HPLC was performed on a HP1100 instrument with the following conditions: Lichrospher RP select B column, 60 Å, 250 × 4 mm (Merck, Budapest, Hungary); detection at 260 nm; flow rate, 1 ml min^{-1}; eluent A, 0.1M aq. TEAC (pH 7.0); eluent B, 0.1M aq. TEAC (pH 7.0): acetonitrile 1:4 gradient, 0–30% B in A in 30 min.

Mass spectrometry

For mass spectrometric measurements a Finnigan TSQ 7000 tandem mass spectrometer equipped with a microelectro spray ion source was used with the previously described protocol [22].

Table 1. Sequences of the oligodeoxynucleotides used in this study

Oligo name	Sequence	*Description* [length]
AREV-M1	5′-CCTGTGT**T**AAATTGTTATCCGC-NH_2-3′	1 alteration [22]
AREV-M2	5′-CCTGT**T**TGAAATT**T**TTATCCGC-NH_2-3′	2 alterations [22]
AREV-M3	5′-CCTGT**T**TGAAATT**T**TTATCC**T**C-NH_2-3′	3 alterations [22]
3′A-AREV	5′-CCTGTGTGAAATTGTTATCCGC-NH_2-3′	Perfect match [22]
5′A-AREV	5′-NH_2-CCTGTGTGAAATTGTTATCCGC-3′	Perfect match [22]
AREV	5′-CCTGTGTGAAATTGTTATCCGC-3′	Perfect match [22]
Cy5-AREV	5′-Cy5-CCTGTGTGAAATTGTTATCCGC-3′	Perfect match [22]
AR17	5′-GTGAAATTGTTATCCGC-3′	Perfect match [17]
AR17-MC	5′-GTGAAATT**C**TTATCCGC-3′	1 alteration G → C [17]
AR17-MA	5′-GTGAAATT**A**TTATCCGC-3′	1 alteration G → A [17]
AR17-MT	5′-GTGAAATT**T**TTATCCGC-3′	1 alteration G → T [17]
AR17-M2	5′-GT**T**AAATT**T**TTATCCGC-3′	2 alterations [17]
Cy5-REV	5′-GCGGATAACAATTTCACACAGG -3′	complementary [22]
R2	5′- NH_2-GATCGATTAAGTTCCTCGTTCGC-3′	random [23]
M1	5′- NH_2-GGCGCCTTTAATATGATGGGAGGA-3′	CMV probe1 [24]
M2	5′- NH_2-CCTTTCGAGGAGATGAA-3′	CMV probe2 [17]
MPF	5′-CGGCATAGAATCAAGGAGCACATGC-3′	CMV forward [24]
MPR	5′-Cy5-AAGGCTGAGTTCTTGGTAAAGAAC-3′	Cy5-CMV reverse [24]

Nucleotides changed in altered sequences are denoted as bold underline characters.

Small molecule microarrays

Anchoring

Biotin and benzamidine were printed on the developed surfaces at 6 concentrations (20, 4, 2, 0.8, 0.2, and 0.1 mM) to determine the optimal immobilization concentration. Test chemical arrays were prepared containing 140 molecules from combinatorial, parallel synthesis. In order to achieve comparable immobilization among the different molecules each species carried an amino functional group active towards the derivatized surface. The molecules were printed both manually and with a printing robot. Each molecule was arrayed from a 20 mM solution in DMF.

Among the immobilized molecules were biotin and benzamidine that are substrates of avidin and trypsin, respectively. Both the acrylic-based and triamino-based chemically modified surfaces were tested in anchoring small chemical compounds.

Affinity experiments

For affinity experiments Alexa Fluor 555-labeled streptavidin (Molecular Probes, Leiden, The Netherlands) and Alexa Fluor 647-labeled trypsin were used.

Labeling of trypsin

30 μl of 1μg μl^{-1} Alexa Fluor 647 carboxylic acid succinimidyl ester (Molecular Probes, Leiden, The Netherlands) solution in DMF containing 1% diisopropylethyl-amine and 6 μl of 10 μg μl^{-1} trypsin solution containing 0.5% TFA were incubated at room temperature for 1 h. To the reaction mixture 100 μl 20 mM Tris pH 7.0 was added. The mixture was loaded onto a millipore centrifuge column (3000 and 10.000 Da cutoff) and centrifuged for 20 min at 10.000 rpm. 200 μl 20 mM Tris pH 7.0 was then added and the column was centrifuged again for 20 min at 10.000 rpm. The process was repeated until the overflow was colorless. The labeled trypsin was redissolved in 200 μl PBS.

Development and scanning of microarrays

Printed microarrays were blocked with 1x SSC, 0.2% SDS, and 1% BSA solution for 30 min at 42 °C, then thoroughly washed with deionized water and dried. The spots were developed with fluorescently labeled trypsin or streptavidin (app. 0.28 μg μl^{-1}) under glass cover slips for 2 h at room temperature in a humid chamber. The coverslips were removed in PBS solution and then the slides were washed with PBS, 1x SSC and 1x SSC and 0.1% SDS for 5 min each.

Finally, the slides were rinsed with deionized water and dried.

The slides were scanned with a ScanArray Lite (GSI Lumonics, Billerica, U.S.A.) microarray scanner at 543 nm for Alexa Fluor 555 and at 633 nm for Alexa Fluor 647 labeling.

Results and discussion

Surface chemistry

We have developed 6 different chemical surfaces with a 3-D-like linker system for biomolecular immobilization, and compared them to previously described immobilization strategies.

The surfaces can be sorted into two groups according to their structure.

First group includes the dendrimer-like structures. The multiplication of active sites is achieved through incubation with TEPA. This reaction step theoretically increases the number of active sites by a factor of four (Figure 1A). The second group includes the triamino based surfaces. The increase of active sites was achieved through the application of 3-[2-(2-Aminoethylamino)ethylamino)propyltrimethoxysilane, which introduces three amine functions. In this case the multiplying factor was three (Figure 1B).

The created branching structures also yield good accessibility of the immobilized molecules in hybridization and affinity experiments.

Optimal binding parameters were determined with the anchoring of labeled oligonucleotides. The application of labeled oligonucleotides provided ease of direct detection and determination of efficiency of anchoring. Immobilization occurred at all investigated pH values, but more effective anchoring was carried out at higher pH values, revealed by harsh (more stringent) washing conditions. The optimal pH value for immobilization was pH10 (Figure 2). In all later experiments the pH was adjusted to 10.

The advantage of the system is that processing of the microarrays includes only temperature controlled incubation, after printing. A reductive coupling step is not necessary, as it is when using carbonyl activated surfaces [15]. Reduction often interacts with reduction-sensitive functional groups linked to nucleic acids (different fluorescent dyes linked to one or both ends of the nucleic acid). Therefore, oligonucleotides having fluorophores at either or both ends can be easily attached using the method described here.

Oligonucleotide hybridization and the mechanism of anchoring

The hybridization properties of the surfaces were assessed and compared to commercially available solid supports. 17 and 22 nucleotides long oligonucleotides were spotted. The longer oligonucleotides carried 0, 1, 2 and 3 alterations in middle positions (AREV, AREV-M1, AREV-M2, AREV-M3). Shorter oligonucleotides carried 0, G → C, G → A, G → T and 2 alterations in middle positions (AR17, AR17-MC, AR17-MA, AR17-MT, AR17-M2). Complementer Cy5 labeled oligonucleotide was hybridized to the arrays. All arrays showed high hybridization signals. The longer oligonucleotides had 3' amino modification. The shorter, native oligonucleotides gave high intensity and a reproducible signal and could differentiate between perfectly matching and altered sequences, while longer 3′ amino modified oligonuceotides were unable to differentiate between a perfect match and 1 alteration sequence, but could between oligonucleotides with a perfect match and 2 alteration sequences (Figure 3). (Spot size varied between supports depending on hydrophilic character.) Among the developed supports the dendrimer like, acryl-derivatized support performed slightly better than the others. Using this support the highest intensity differences could be obtained – by hybridizing a labeled oligonucleotide without mismatch – between a mutated sequence bearing one base mismatch in the middle position and an attached complementary sequence.

The binding efficiency of chemically modified and unmodified oligonucleotides was comparable. Applying unmodified oligonucleotides results in greater cost efficiency in microarray production.

From these findings we concluded that the 3′ or 5′ modification in oligonucleotides are not the only sites for immobilization. To determine the mechanism of anchoring, a model reaction was designed. Nucleosides (cytidine, adenosine, guanosine, thymidine) were incubated with acrylamide at 37 °C in aqueous solution. After 2 h the reaction mixtures were analyzed by HPLC and mass spectrometry. In the HPLC chromatograms a new peak represented the product of the conjugation. To determine the properties of the conjugate the fraction containing the reaction product was analyzed by mass spectrometry (Figure 4). All four bases reacted with acrylamide and the products were represented by a new peak on each spectrum with a m/z value corresponding to the mass of the acrylamide nucleoside adduct. To evaluate the relationship between acrylamide and the nucleosides, secondary

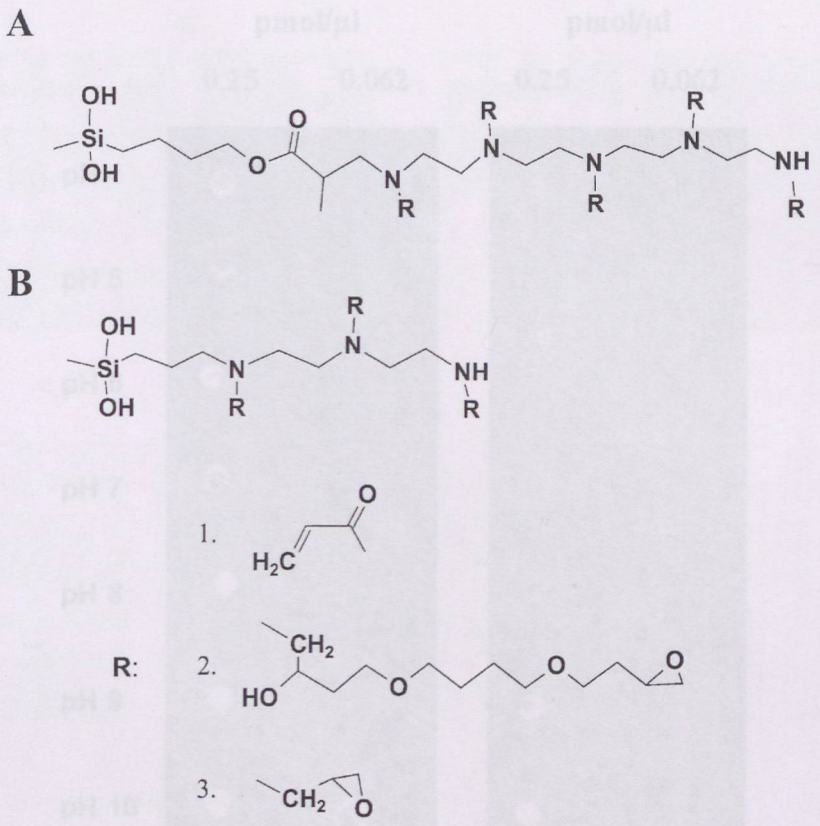


Figure 1. Chemical structure of the developed solid supports and the active linker arms (A: dendrimer-like linker system. B: triamino-based linker system).

(collision) spectra were taken (data not shown). In the spectra, two new major peaks appeared. One corresponds to the sugar part of the nucleoside, the other to the acrylamide nucleobase adduct. Secondary spectra of the reaction products proved that the acrylamide and the nucleosides formed covalent bonds.

From these findings we concluded that the immobilization occurs via the bases of the unmodified oligonucleotides in alkaline pH, supposedly through amino functions.

Detection of PCR products with immobilized oligonucleotides

To test the applicability of the developed solid supports as potential diagnostic tools in clinical molecular biology we studied the attached oligonucleotide – PCR product interactions. Sense (complementary to labeled reverse strand of the amplicon) HCMV-specific oligonucleotides of different length (17 and 24-mers), random sequences (negative control) and Cy5-labeled control oligonucleotides (Table I) were printed onto the acryl derivatized dendrimer-like surface and on a commercially available aldehyde coated support. After hybridization, washing and scanning the slide's signal intensities were compared (Figure 5). It can be seen that much higher intensities could be obtained by using the dendrimer-like surface than by using the simpler aldehyde-coated support. Moreover, the spotted control oligonucleotides bearing fluorescent dyes could be visualized in only the acryl-surface, while no signal could be detected in the case of the aldehyde-coated slide because of the reduction step after immobilization. Surfaces based on triaminosilane starting linkers gave smaller signals than acrylic supports, however they exhibited higher intensities than the aldehyde-coated support (data not shown).

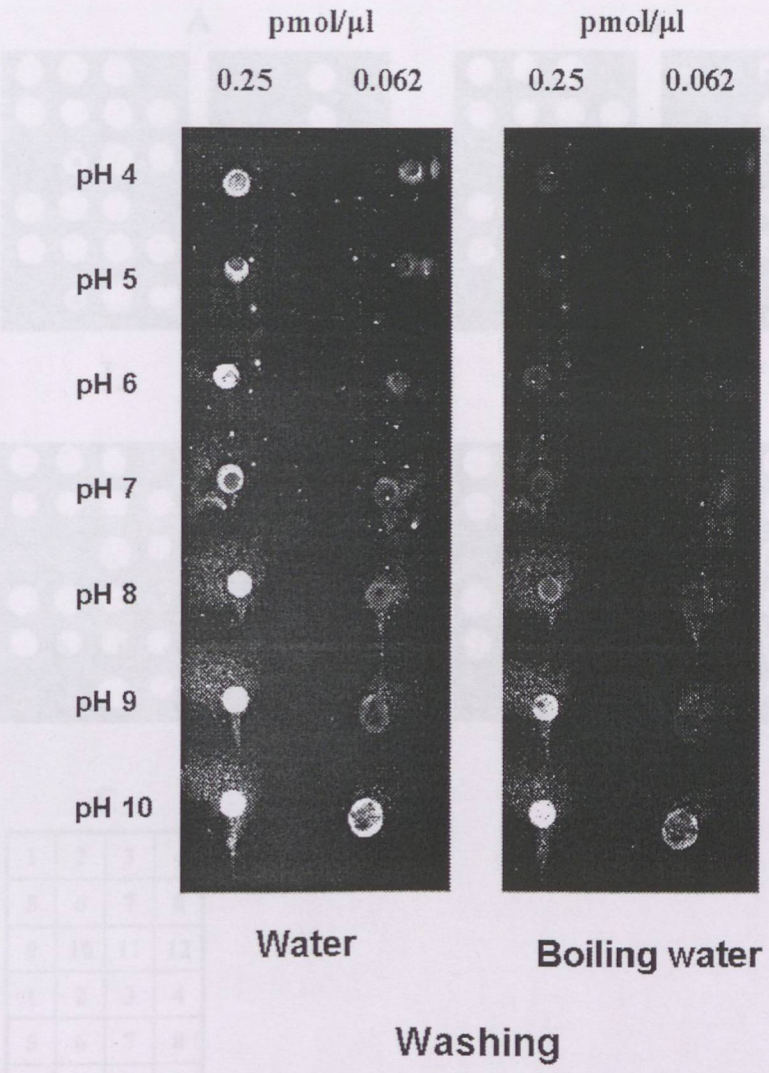


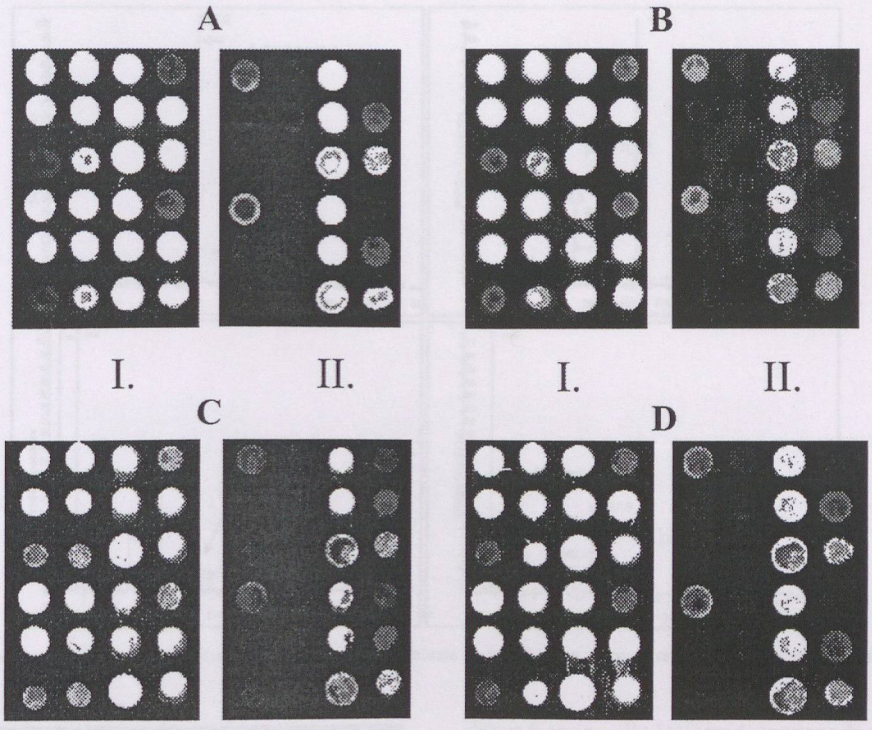
Figure 2. pH characteristics of immobilization to the acrylic surface.

Protein immobilization

The developed linker systems are capable of anchoring molecules with free amino functional groups. Because all proteins have numerous free amino groups on their side chains, a simple experiment was carried out to test protein immobilization. The developed surfaces anchored the applied fluorescent-labeled protein at both applied pH values (pH7, pH10), but according to previous findings the binding was more effective at higher pH (data not shown). The advantage of this anchoring procedure is that the hydrophobicity of the surface can be modulated by the linker sequences. Certain proteins are prone to have stabile conformation in a more hydrophobic microenvironment, while others behave in the opposite way. Therefore, testing the surfaces before application of protein microarrays in different protein-protein or protein-small molecule interaction studies is highly recommended.

Small molecule arrays

To optimize the parameters of small molecule-protein interactions on solid surfaces, two common conjugate systems, the biotin-streptavidin and benzamidine-



E

1	2	3	4
5	6	7	8
9	10	11	12
1	2	3	4
5	6	7	8
9	10	11	12

Figure 3. Fluorescently labeled oligonucleotide hybridized to arrays with oligonucleotides having perfect match and mutated sequences. A: Dendrimer surface with acryl functions, B, C, D: Triamino surface with hydrophobic epoxy, acryl and hydrophilic epoxy functions, respectively. E: layout of the printed oligonucleotides. 1.: AR17, 2.: AR17-MT, 3.: 3'A-AREV, 4.: Cy5-REV, 5.: AR17-MA, 6.: AR17-MC, 7.: AREV-M1, 8.: AREV-M3, 9.: Cy5-REV, 10.: AR17-M2, 11.: AREV-M2 12: Cy5-AREV. I.: Wash 1. (1X SSC for 5 min), II.: Wash 2. (0.1x SSC for 5 min)

trypsin were selected and tested on all the chemically modified slides described. Biotin and benzamidine were printed at 6 different concentrations on the surfaces. All slides had low background and spots corresponding to all 6 applied concentrations gave specific signals. Although spots corresponding to 20 mM and 4 mM concentrations had similar, high fluorescent intensity, the concentration for further experiments was set to 20 mM (Figure 6A).

Using the developed supports we created a test chemical array containing 140 molecules from combinatorial, parallel synthesis. The multiply-tethered, small molecules represent a diverse subset of a 240,000-member discovery library composed of pharmaceutically relevant structures [23, 24]. Figure 7

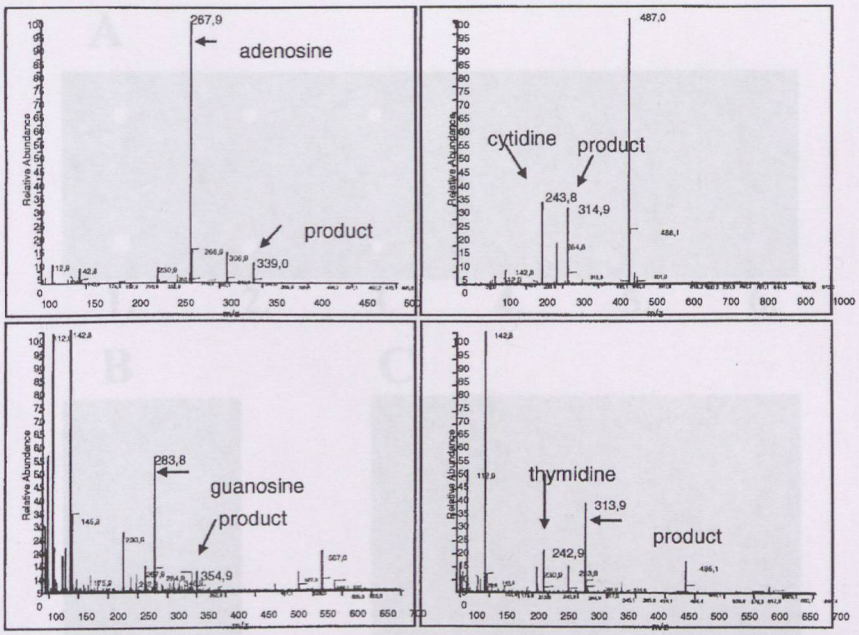


Figure 4. Mass spectra of the samples from four test reactions to evaluate immobilization. Product represents nucleoside-acrylamide adduct.

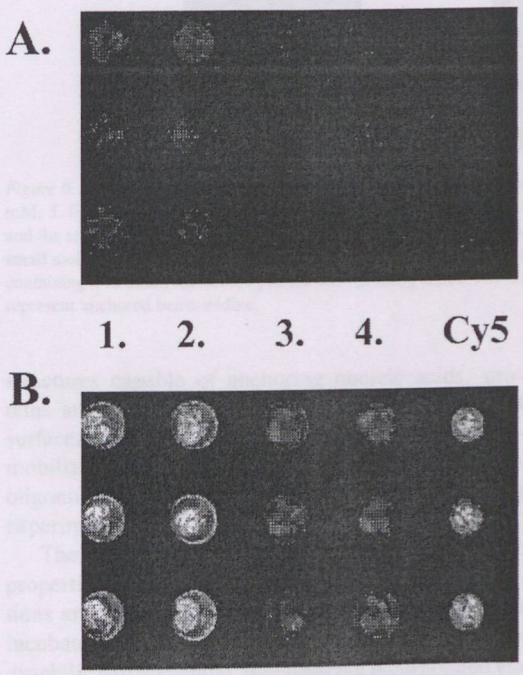


Figure 5. Detection of PCR products with species-specific probe oligonucleotides. A: aldehyde coated support, B: acryl derivatized dendrimer like surface 1, 2: HCMV specific probe, 3, 4: shorter HCMV specific probe, Cy5: Cy5 labeled oligonucleotide.

shows 6 amino-alkyl tethered compounds as examples for the immobilized small molecules. In order to achieve comparable immobilization among the different molecules, each species carried an amino functional group active towards the derivatized surfaces. The amino groups were situated on a linker arm to provide greater accessibility once anchored to the surface. Among the immobilized molecules were biotin and benzamidine, which are substrates of avidin and trypsin, respectively. For affinity experiments, Alexa Fluor 555-labeled streptavidin and Alexa Fluor 647-labeled trypsin were used. Results show that arrays developed by the triamino-based chemistry showed low background after small chemical compounds were printed manually or with a printing robot after incubation with interacting fluorescent labeled proteins. The anchored molecules were accessible for interaction and protein-specific signals were obtained (Figure 6B,C).

Conclusion

In this study we describe diverse chemical modifications of glass surfaces based on two different approaches to achieve reactive 3-D-like branching

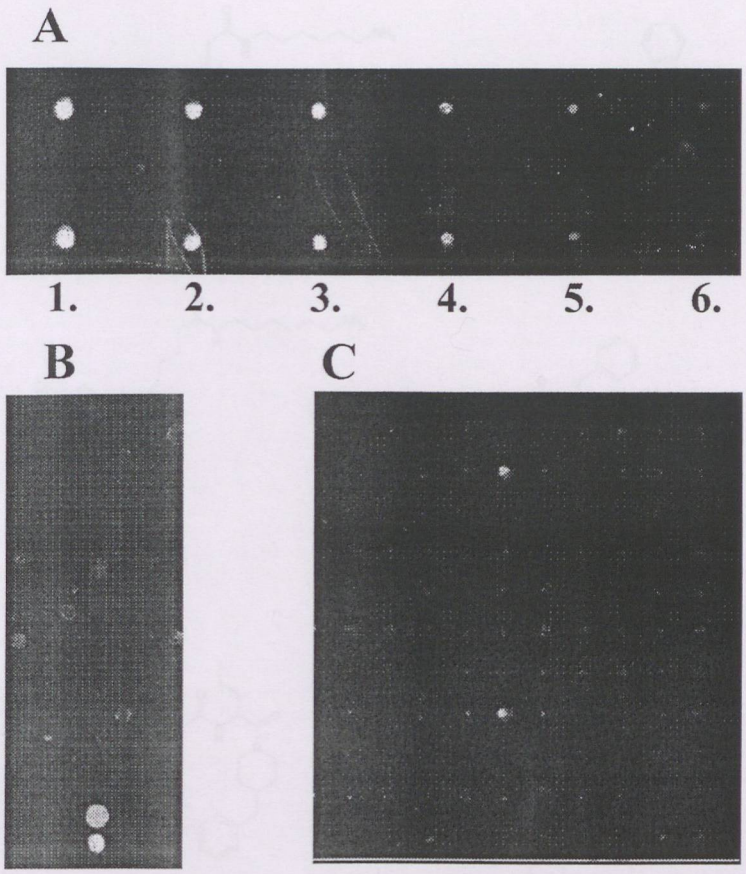


Figure 6. Small molecule arrays A: Benzamidine was spotted onto the array at different concentrations (1. 10 mM; 2. 2 mM; 3. 1 mM; 4. 0.4 mM; 5. 0.1 mM; 6. 0.05 mM – meaning 1. 10 pmol; 2. 2 pmol; 3. 1 pmol; 4. 0.4 pmol; 5. 0.1 pmol; 6. 50 fmol amounts in the spotted area) and the array was incubated with Alexa Fluor 647-labeled trypsine and visualized after scanning. B: Manually printed test array containing 140 small molecules. The array was developed with Alexa Fluor 555-labeled streptavidin. Illuminated spots represent anchored biotin. C: Test array containing 140 small molecules printed with printing robot. The array was developed with Alexa Fluor 647-labeled trypsin. Illuminated spots represent anchored benzamidine.

structures capable of anchoring nucleic acids, proteins and small chemical molecules. The developed surfaces were compared to previously described immobilization strategies. Our system performed well in oligonucleotide, protein and small molecule anchoring experiments.

The developed surfaces utilize either the reactive properties of epoxy or acrylic functions. Epoxy functions are introduced by two methods. One way is to incubate with epichlorohydrin. To create a more hydrophilic surface, epoxy functions are incorporated via a bifunctional linker, 1,4-butanediol diglycidyl ether. The latter method provides a longer linker arm holding the active functions. The binding capacity of the surface is enhanced through multiplying the active sites by creating a branching structure on the surface.

The increased capacity of the supports was confirmed by comparing the developed surfaces with commercially available, chemically modified glass slides in attaching oligonucleotides in hybridization studies. Higher signals could be obtained using surfaces with the branching structure than in the case of commercially available aldehyde-coated surfaces. The increased signals could be the results of higher sample density in the spotted feature and the better accessibility of the probes to the immobilized samples.

Furthermore, the methods described provide the possibility to modulate the hydrophobic/hydrophobic

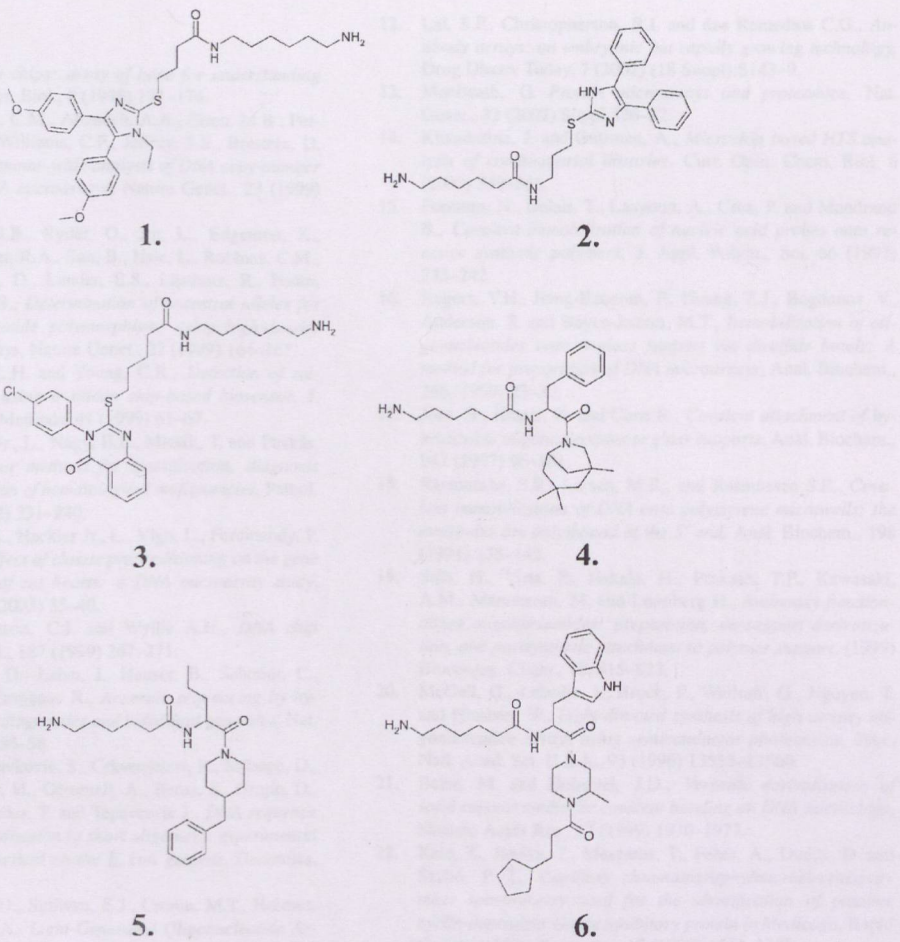


Figure 7. 6 amino-alkyl tethered compounds as examples for the immobilized small molecules.

properties of the surfaces, which can have an effect on specificity in experiments where protein-protein or protein-small molecule interactions are investigated (protein arrays and chemical arrays).

From HPLC and mass spectrometry experiments we concluded that the immobilization occurs via the bases of the unmodified oligonucleotides in alkaline pH, supposedly through amino functions. Hence, all four bases could react with the acrylic groups at alkaline pH, therefore attachment of oligonucleotides does not depend on their base contents. The tested surfaces offer the capability to permanently immobilize several unmodified biomolecular species in an array format at high density, using high-precision robotic spotters. Applying unmodified biomolecules (e.g., oligonucleotides) results in greater cost efficiency in microarray production. The generated microarrays can be applied in DNA-microarray, protein-microarray and chemical-microarray studies.

Acknowledgments

This work was supported in part by Biotechnológia 2000 grant no. BIO-0005/2000 and Biotechnológia 2002 grant no. BIO-0006/2002 from OMFB. The authors wish to thank S. Bottka and Z. Kupihár for the preparation and purification of the oligonucleotides.

References

1. Johnston, M., *Gene chips: array of hope for understanding gene regulation*, Curr. Biol., 8 (1998) 171–174.
2. Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D. and Brown, P.O., *Genome-wide analysis of DNA copy-number changes using cDNA microarrays*, Nature Genet., 23 (1999) 41–46.
3. Hacia, J.G., Fan, J.B., Ryder, O., Jin, L., Edgemon, K., Ghandour, G., Mayer, R.A., Sun, B., Hsie, L., Robbins, C.M., Brody, L.C., Wang, D., Lander, E.S., Lipshutz, R., Fodor, S.P. and Collins, F.S., *Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays*, Nature Genet., 22 (1999) 164–167.
4. Dill, K., Stanker, L.H. and Young, C.R., *Detection of salmonella in poultry using a silicon chip-based biosensor*, J. Biochem. Biophys. Methods, 41 (1999) 61–67.
5. Zvara, A., Hackler Jr., L., Nagy, B.Z., Micsik, T. and Puskás, L.G., *New molecular methods for classification, diagnosis and therapy prediction of hematological malignancies*, Pathol. Oncol. Res., 8 (2002) 231–240.
6. Ónody, A., Zvara, A., Hackler Jr., L., Vígh, L., Ferdinandy, P. and Puskás, L.G., *Effect of classic preconditioning on the gene expression pattern of rat hearts: a DNA microarray study*, FEBS Letters, 536 (2003) 35–40.
7. Kurian, K.M., Watson, C.J. and Wyllie A.H., *DNA chip technology*, J. Pathol., 187 (1999) 267–271.
8. Drmanac, S., Kita, D., Labat, I., Hauser, B., Schmidt, C., Burczak, J.D. and Drmanac R., *Accurate sequencing by hybridization for DNA diagnostics and individual genomics*, Nat. Biotech., 16 (1998) 54–58.
9. Milosavljevic, A., Savkovic, S., Crkvenjakov, R., Salbego, D., Serrato, H., Kreuzer, H., Gemmell, A., Batus, S., Grujic, D., Carnahan, S., Paunesku, T. and Tepavcevic J., *DNA sequence recognition by hybridization to short oligomers: experimental verification of the method on the E. coli genome*, Genomics, 37 (1996) 77–86.
10. Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P. and Fodor, S.P.A., *Light-Generated Oligonucleotide Arrays for Rapid DNA Sequence Analysis*, Proc. Natl. Acad. Sci. USA. 91 (1994) 5022–5026.
11. Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., Kobayashi, S., Davis, C., Dai, H., He, Y.D., Stephaniants, S.B., Cavet, G., Walker, W.L., West, A., Coffey, E., Shoemaker, D.D., Stoughton, R., Blanchard, A.P., Friend, S.H. and Linsley, P.S., *Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer*, Nat. Biotech., 19 (2001) 342–347.
12. Lal, S.P., Christopherson, R.I. and dos Remedios C.G., *Antibody arrays: an embryonic but rapidly growing technology*, Drug Discov Today, 7 (2002) (18 Suppl):S143–9.
13. MacBeath, G. *Protein microarrays and proteomics*, Nat. Genet., 32 (2002) Suppl:526–32.
14. Khandurina, J. and Guttman, A., *Microchip based HTS analysis of combinatorial libraries*, Curr. Opin. Chem. Biol. 6 (2002) 359–366.
15. Ferraton, N., Delair, T., Laayoun, A., Cros, P. and Mandrand B., *Covalent immobilization of nucleic acid probes onto reactive synthetic polymers*, J. Appl. Polym., Sci. 66 (1997) 233–242.
16. Rogers, Y.H., Jiang-Baucom, P., Huang, Z.J., Bogdanov, V., Anderson, S. and Boyce-Jacino, M.T., *Immobilization of oligonucleotides onto a glass support via disulfide bonds: A method for preparation of DNA microarrays*, Anal. Biochem., 266 (1999) 23–30.
17. Joos, B., Kuster, H. and Cone R., *Covalent attachment of hybridizable oligonucleotides to glass supports*, Anal. Biochem., 247 (1997) 96–101.
18. Rasmussen, S.R., Larsen, M.R., and Rasmussen S.E., *Covalent immobilization of DNA onto polystyrene microwells: the molecules are only bound at the 5' end*, Anal. Biochem., 198 (1991) 138–142.
19. Salo, H., Virta, P., Hakala, H., Prakash, T.P., Kawasaki, A.M., Manoharan, M. and Lonnberg H., *Aminooxy functionalized oligonucleotides: preparation, on-support derivatization, and postsynthetic attachment to polymer support*, (1999) Bioconjug. Chem., 10, 815–823.
20. McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguyen, T. and Hinsberg W., *Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists*, Proc. Natl. Acad. Sci. U.S.A., 93 (1996) 13555–13560.
21. Beier, M. and Hoheisel, J.D., *Versatile derivatisation of solid support media for covalent bonding on DNA-microchips*, Nucleic Acids Res., 27 (1999) 1970–1977.
22. Kele, Z., Janáky, T., Mészáros, T., Fehér, A., Dudits, D. and Szabó, P. T., *Capillary chromatography/microelectrospray mass spectrometry used for the identification of putative cyclin-dependent kinase inhibitory protein in Medicago*, Rapid Commun Mass Spectrom., 12 (1998) 1564–1568.
23. Darvas, F. and Kovács, L., *CMT: A solution phase combinatorial approach. Synthesis and yield prediction of phenazins.* In J. P. Devlin (ed.), *High-Throughput Screening*, Marcel Dekker Inc., New York, (1997) 223–242.
24. Darvas, F., Dorman, G., Ürge, L., Szabó, I., Rónai, Z. and Sasváry-Székely, M., *Combinatorial Chemistry. Facing the Challenge of Chemical Genomics*, Pure Appl. Chem., 73 (2001) 1487–1498.

Research Report

RNA Amplification Results in Reproducible Microarray Data with Slight Ratio Bias

BioTechniques 32:1330-1340 (June 2002)

László G. Puskás, Ágnes Zvara, László Hackler Jr., and Paul Van Hummelen[1]
Hungarian Academy of Sciences, Szeged, Hungary and [1]Flanders Interuniversity Institute for Biotechnology, Leuven, Belgium

ABSTRACT

Microarray expression analysis demands large amounts of RNA that are often not available. RNA amplification techniques have been developed to overcome this problem, but limited data are available regarding the reproducibility and maintenance of original transcript ratios. We optimized and validated two amplification techniques: a modified in vitro transcription for the linear amplification of 3 µg total RNA and a SMART™ PCR-based technique for the exponential amplification of 50 ng total RNA. To determine bias between transcript ratios, we compared the expression profiles in mouse testis versus spleen between the two amplification methods and a standard labeling protocol, using microarrays containing 4596 cDNAs spotted in duplicate. With each method, replicate hybridizations were highly reproducible. However, when comparing the amplification methods to standard labeling, correlation coefficients were lower. Twelve genes that exhibited inconsistent or contradictory expression ratios among the three methods were verified by quantitative RT-PCR. The amplification methods showed slightly more discrepancies in the expression ratios when compared to quantitative RT-PCR results but were more sensitive in terms of detecting expressed genes. In conclusion, although amplification methods introduce slight changes in the transcript ratios compared to standard labeling, they are highly reproducible. For small sample size, in vitro transcription is the preferred method, but one should never combine different labeling strategies within a single study.

INTRODUCTION

DNA microarray technology provides a rapid and comprehensive approach to monitor simultaneously the expression levels of thousands of genes among diverse biological samples in a comparative way (5,13–15). Standard protocols of gene expression analysis require large amounts of RNA, typically 20–200 µg total RNA or 0.5–2 µg poly(A) RNA (10,17,18). To obtain these amounts of RNA, milligrams of tissue or approximately 10^7 cells are needed. Alternatively, one can amplify the RNA. One way to amplify RNA linearly is to make use of in vitro transcription, where a T7 promoter is linked to the oligo(dT) primer and used for first-strand cDNA synthesis (17). The probe can be labeled during or after the in vitro amplification step using fluorescently labeled NTPs or dNTPs, respectively (RNA amplification protocol; http://www.microarrays.org/pdfs/ ModifiedEberwine.pdf). Using in vitro transcription, the starting material can be limited to 100 ng to 10 µg total RNA or as little as the amount of RNA present in one cell, if two consecutive in vitro transcription cycles are performed (12). Another amplification strategy (RT-PCR) uses a logarithmic amplification step in which the starting material can require as little as 1 ng total RNA. This is usually accomplished by (*i*) ligating an anchor to the ends of the cDNA and amplifying the complex mixture using primers complementary to the anchor sequence (1,2) or (*ii*) using primers with 3′-end random stretches of nucleotides that can hybridize to the transcripts during reverse transcription (4,6).

One should be prudent using the amplification steps because possible bias can be introduced by these techniques. Theoretically, the risk of distorting the initial transcript level is highest for the logarithmic amplification methods. PCR-based methods are thought to be very sensitive for small changes in experimental conditions, especially for rare and long transcripts. Selective sequence amplification could also occur in the case of linear amplifications. Otherwise, in vitro transcription may also introduce bias of the initial transcript level caused by sequence-specific efficiencies of RNA polymerase, partly deriving from the uncontrolled termination of the enzyme because of a very long poly(A) tail or strong secondary structure of the message.

Although several laboratories have already used amplification steps for microarray analysis (3,7,8,11), the effect on the ratios has not been rigorously studied or published. Therefore, we present a comparative study of these amplification methods using standard labeling without amplification to investigate their reproducibility, reliability, and sensitivity.

MATERIALS AND METHODS

Construction of Microarrays

A total of 4596 PCR-amplified cDNA fragments from sequenced, verified IMAGE clones (Mouse Gem I; Incyte, USA) were spotted twice, once on each side of the slide. The cDNA inserts were amplified with M13 primers, purified with MultiScreen® PCR plate (Millipore, Brussels, Belgium), and resuspended in 20 µL 50% DMSO in an average concentration of 100 ng/µL. PCR products were arrayed on Type V silane-coated slides using a Generation III printer (both from Amersham Biosciences, Buckinghamshire, UK). The slides were blocked just before hybridization in 3.5% SSC, 0.2% SDS, and 1% BSA for 10 min at 60°C.

RNA Samples

To avoid variability caused by RNA preparation, we used highly purified commercially available poly(A) and total RNA from BD Biosciences Clontech (Erembodegem, Belgium) and Ambion (Austin, TX, USA). All experiments were performed with RNA from mouse adult spleen and testis.

Direct Labeling of RNA during Reverse Transcription

First-strand cDNA probes were generated by the reverse transcription of 2 µg poly(A) RNA using an anchored oligo(dT) (d-T_{25}-dA/C/G) primer (0.4 µM; Genset, Paris, France), 0.1 mM d(G/T/A)TPs, 0.05 mM dCTP (Amersham Biosciences), 0.05 mM Cy3-dCTP or Cy5-dCTP (Amersham Biosciences), 1× first-strand buffer, 10 mM DTT, and 200 U SUPERSCRIPT II™ (Invitrogen, Merelbeke, Belgium) in 20 µL total volume. The RNA and primers were denatured at 75°C for 5 min and cooled on ice before we added the other reaction components. After 2 h incubation at 42°C, mRNA was hydrolyzed in 250 mM NaOH for 15 min at 37°C. The sample was neutralized with 10 µL 2 M MOPS (0.6 M final concentration) and purified using QIAquick™ columns (Qiagen, Hilden, Germany).

Linear Amplification by In Vitro Transcription

Antisense RNA amplification was performed using a modified protocol of in vitro transcription available on http://www.microarrays.org/pdfs/ModifiedEberwine.pdf and published by Phillips et al. (12) and Van Gelder et al. (16). For the first-strand cDNA synthesis, 3 µg total RNA were mixed with 2 µg HPLC-purified, anchored oligo(dT) + T7 promoter (5′-GGCCAGTGAATT-GTAATACGACTCACTATAGGGAG-GCGG-T_{24}(ACG)-3′) (Eurogentec, Belgium), 40 U RNAseOUT™ (Invitrogen), and 0.9 M D(+)trehalose (Sigma, Bornem, Belgium) in a total volume of 11 µL and heated at 75°C for 5 min. The following was added to this mixture to give a 20 µL final volume: 4 µL 5× first-strand buffer (Invitrogen), 2 µL 0.1 M DTT, 1 µL 10 mM dNTP mixture, 1 µL 1.7 M D(+)trehalose, and 1 µL 200 U SUPERSCRIPT II (20 µL final volume). The sample was then incubated in a Uno II thermal cycler (Biometra) at 37°C for 5 min, 45°C for 10 min, and 10 cycles of 60°C for 2 min and 55°C for 2 min. To the first-strand reaction mixture, 33.4 µL 5× second-strand synthesis buffer (Invitrogen), 3.4 µL 10 mM dNTP mixture, 1 µL 10 U/µL *E. coli* DNA ligase (Invitrogen), 4 µL 10 U/µL *E. coli* DNA polymerase I (Invitrogen), and 1 µL 2 U/µL *E. coli* RNase H (Invitrogen), 103.8 µL water were added and incubated at 16°C for 2 h. The synthesized dsDNA was purified with a QIAquick kit. Antisense RNA was synthesized in a total volume of 20 µL using an AmpliScribe T7 High-Yield transcription kit (Epicentre, Madison, WI, USA), according to the manufacturer's instructions. The RNA was then purified with an RNeasy® purification kit (Qiagen). From this mixture, 3 µg amplified RNA were labeled and purified, using random nonamers in the standard protocol previously described (Genset, Paris, France).

Exponential Amplification by SMART™ cDNA Synthesis and PCR

The SMART method is based on intrinsic terminal transferase activity of a point mutant RNase H reverse transcriptase. When the reverse transcription reaches the end of the mRNA template, it automatically adds a stretch of C-rich sequence to the ends of newly synthesized strands of cDNA. An oligonucleotide with a G stretch at its 3′-end could hybridize to the C stretch, creating an extended template for the reverse transcription. cDNA synthesis continues to the end of this oligomer. Second-strand cDNA synthesis is then easily accomplished using PCR with the G-extended primer and an oligo(dT) primer.

Fifty nanograms of total RNA were reverse-transcribed in 20 µL, according to the manufacturer's protocol for the SMART PCR cDNA synthesis kit (BD Biosciences Clontech), with an anchored poly(dT) primer (CDS) and a SMART II primer. Three microliters of the reverse transcription reaction were added to 100 µL PCR mixture (1× Advantage® 2 PCR buffer, 0.2 mM dNTPs, 30 nM PCR primer, 1× Advantage 2 polymerase mixture containing 40 pmol CDS and SMART PCR primers). Amplification was performed in a Uno II thermal cycler at 95°C for 1 min, 17–25 cycles of 95°C for 25 s, 65°C for 40 s, and 68°C for 6 min. To determine the optimal number of PCR cycles, 6-µL aliquots were removed from the reaction every second cycle after the first 17 cycles and analyzed on a 1.5% agarose gel. After optimization, we performed two 100-µL reactions using the same parameters and optimal number of cycles to get amplified cDNA. After purification with QIAquick, the products were labeled in two 60-µL reactions using asymmetric amplification containing 1× Advantage 2 PCR buffer, 0.2 mM d(A/T/G)TPs, 0.04 mM dCTP, 0.05 mM Cy3-dCTP or Cy5-dCTP, 60 nM dT_{25}(A/C/G) primer, and 1× Advantage 2 polymerase mixture in a Uno II thermal cycler at 95°C for 1 min, and then 10 cycles of 95°C for 25 s, 50°C for 40 s, 62°C for 3 min, and 68°C for 5 min. The amplified and labeled probes were purified using QIAquick columns.

Array Hybridization and Posthybridization Processes

The probes were resuspended in 30 µL hybridization solution (50% for-

mamide, 5× SSC, 0.1% SDS, 100 µg/mL salmon sperm DNA) and prehybridized at 42°C for 30 min to block hybridizations of polyA/T tails of the cDNA on the arrays by adding poly(dA) for the direct labeling of RNA during reverse transcription and SMART or poly(dT) for the in vitro transcription hybridizations. Mouse COT-1 DNA® (1 mg/mL; Invitrogen) was added to the mixture and placed on the array under a glass coverslip. Slides were then incubated for 18 h at 42°C in a humid hybridization cabinet (Amersham Biosciences). Posthybridization washings were performed for 10 min at 56°C in 1× SSC, 0.1% SDS, twice for 10 min at 56°C in 0.1× SSC, 0.1% SDS, and for 2 min at 37°C in 0.1× SSC.

Scanning and Data Analysis

Arrays were scanned at 532 and 635 nm using a Generation III scanner. Image analysis was performed using ArrayVision (Imaging Research, ON, Canada). Spot intensities were corrected for background and filtered based on two standard deviations above the background. Ratios were normalized by a linear regression between $\log_{10}$ ratio (Cy5/Cy3) and $\log_{10}$ total intensity of Cy5*Cy3. For duplicate spots, the average ratios of Cy5/Cy3 were used for further analysis. Scatter plots were visualized using http://www.Spotfire.net (Spotfire AB, Göteborg, Sweden) and Statview (SAS Institute). The correlations and *P* values were calculated using Statview.

Single-Gene RT-PCR and Southern Blot Analysis

Semiquantitative RT-PCR was performed similarly as described in Huntsman et al. (9). Briefly, 100 ng mouse spleen and testis mRNA were reverse-transcribed using specific primers complementary to the 3′-end of the coding region of the desired gene with the SUPERSCRIPT II system in total volume of 30 µL. Gene-specific primers used for the amplification of the desired clones were combined with 0.5 µL reverse transcriptase mixture as template in total volume of 50 µL. To avoid overamplification, the optimal number of PCR cycles was determined by agarose gel electrophoresis and ethidium bromide staining (18–26 cycles were performed according to the abundance of the appropriate mRNA). PCR products were separated with electrophoresis using 2% agarose gel and then transferred onto nylon membranes (Amersham Biosciences) by capillary action. ^{32}P-labeled probes were synthesized with PCR using plasmids with the corresponding cDNA insert that were spotted on the array with M13 forward and reverse primers in a 100 µL total volume containing 1× PCR buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.5, 0.1% Triton® X-100), 1.5 mM $MgCl_2$, 125 µM each dNTP, 5 µCi [α-^{32}P]dCTP (3000 Ci/mmol), 50 pmol each primer, and 3 U Platinum *Taq* DNA polymerase (Invitrogen). Labeled PCR fragments were purified with QIAquick, resuspended in hybridization buffer, denaturated, and applied onto the membrane. Hybridizations and washing steps were performed using standard protocols (18). Membranes were exposed to phosphorimager

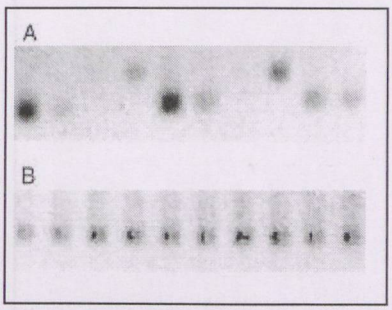


Figure 1. Autoradiograph showing inconsistent expression. Autoradiograph of 12 genes analyzed by semiquantitative RT-PCR between testis and spleen. Five of these genes showed inconsistent or contradictory expression ratios among the methods tested. (A) The following genes are shown: lane 1, AA049537; lane 2, AA050371; lane 3, AA058055; lane 4, AA175441; and lane 5, W13718. (B) As a control, the expression of the GAPDH gene is shown. T, testis. S, spleen.

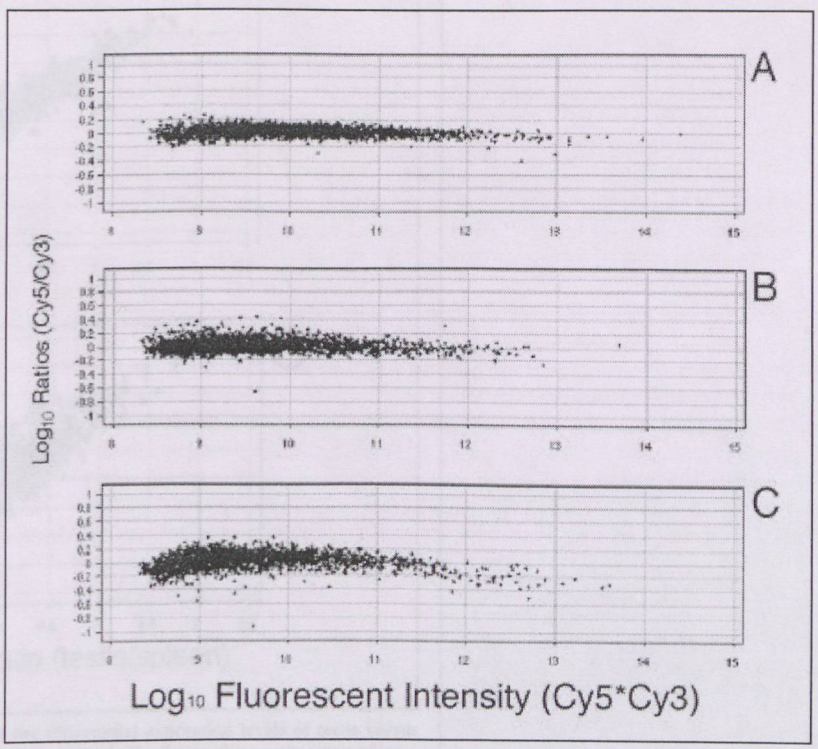


Figure 2. Reproducibility of spleen-versus-spleen microarray analysis performed by three different labeling methods. For each method, mouse spleen RNA was labeled with Cy5 and Cy3 separately and hybridized together. The figure shows scatter plots of $\log_{10}$ (Cy5/Cy3) ratios of all 9192 spots (two replicate spots/gene) against $\log_{10}$(intensity of Cy3*Cy5). (A) Standard labeling starting with 2 µg poly(A) RNA. (B) IVT starting with 3 µg total RNA. (C) SMART starting with 50 ng total RNA. Only ratios that had fluorescent signals significantly above background were considered for spots. No further selections were made.

screens (Amersham Biosciences), and intensity ratios were calculated with ImageQuant 4.2 software. ^{32}P-labeled products were normalized to amplified levels of the ubiquitously expressed glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Figure 1 illustrates the images after autoradiography.

RESULTS AND DISCUSSION

The accuracy and sensitivity of DNA microarray analysis is a very important consideration in monitoring the transcriptome. Indeed, small changes in expression levels could be of great value in studying biochemical or other bi-

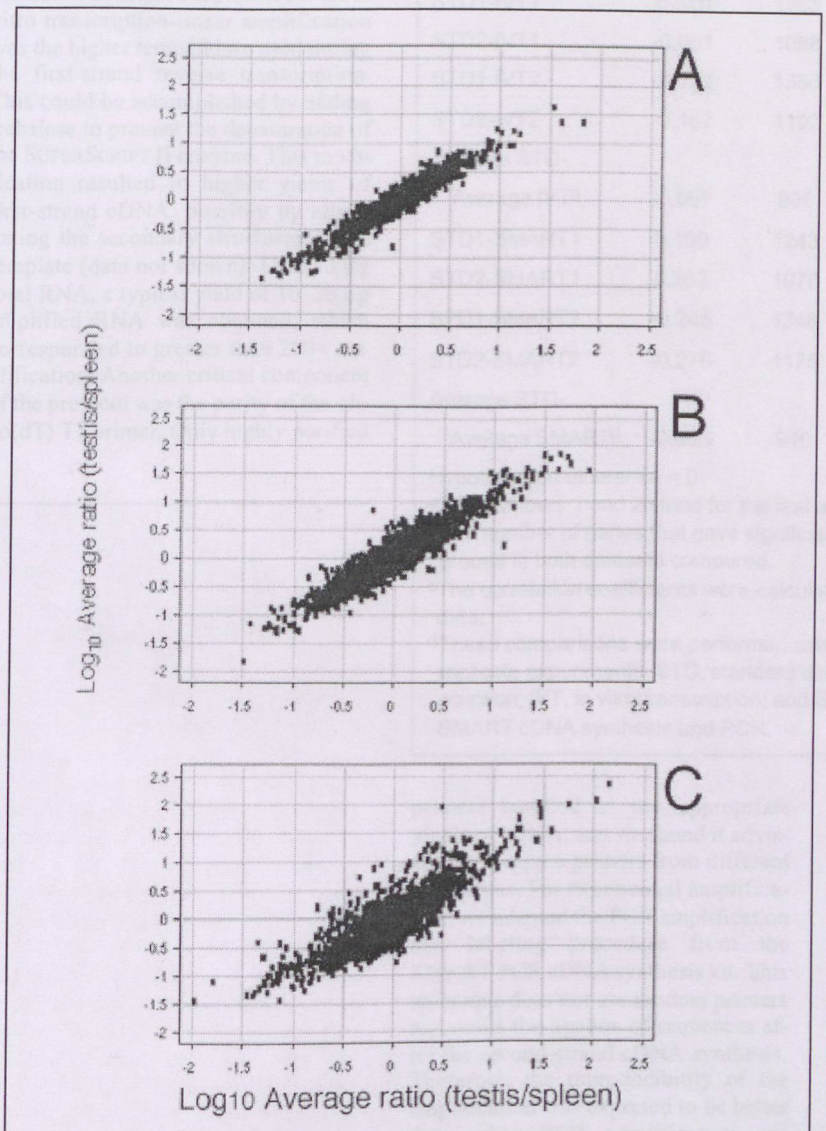


Figure 3. Reproducibility and preservation of the differential expression levels of testis versus spleen among replicate experiments for each labeling method. The figures show scatter plots of $\log_{10}$ average ratio (testis/spleen) against the same results of an independent replicate experiment within each labeling method. (A) Standard labeling (STD1 and STD2); (B) linear sample amplification (IVT1 and IVT2); and (C) exponential sample amplification by SMART-PCR (SMART1 and SMART2). The only average values considered were those for which both replicate spots had fluorescent signals significantly above background and also had a less than 2-fold difference from one another. See Table 1 for the number of genes used in this comparison.

ological functions. Another challenge of microarray gene expression analysis is often the limiting amount of RNA that can be obtained from small or limited biological samples. Thus, we optimized two sample amplification techniques: a linear amplification-based in vitro transcription method and an exponential amplification-based SMART method. The major adaptation for the in vitro transcription-linear amplification was the higher temperature used during the first-strand reverse transcription. This could be accomplished by adding trehalose to prevent the denaturation of the SUPERSCRIPT II enzyme. This modification resulted in higher yields of first-strand cDNA, possibly by eliminating the secondary structures of the template (data not shown). Using 3 µg total RNA, a typical yield of 10–20 µg amplified RNA was obtained, which corresponded to greater than 200× amplification. Another critical component of the protocol was the purity of the oligo(dT) T7 primer. Only highly purified primers resulted in the appropriate yields of cDNA, and we found it advisable to compare primers from different companies. For exponential amplification, we adapted the PCR amplification and labeling procedure from the SMART PCR cDNA synthesis kit. This technique does not use random primers and omits the ligation of sequences after the second-strand cDNA synthesis. Therefore, the reproducibility of the amplification was expected to be better than other PCR amplification approaches in which the contamination of genomic DNA or poly(A)- RNA can be amplified, thus leading to nonspecific hybridization. After the PCR amplification of the cDNA, the probe was generated by asymmetric PCR, resulting in only one labeled strand. This leads to a considerable enhancement of the signal (data not shown). The optimal number of PCR cycles had to be determined in every case. More than 10% of the spots exhibited nonreproducible intensity data (more than 2.5-fold difference between the intensities from duplicate experiments) caused by the negative effect of overcycling (data not shown). We used 50 ng total RNA as starting material, and the approximately 2000× amplification rate of SMART resulted in sufficient labeled probe for two hybridization experiments.

In this work, we primarily focused on two questions: (*i*) "Are amplification methods reproducible?" and (*ii*) "Do amplification methods introduce bias in relative transcript levels?" For the first question, we compared SpleenCy5/

Table 1. Two-Tailed Paired *t* Test of Testis/Spleen Ratios within and between Labeling Methods

Comparison[a]	Mean Difference	No. of Genes[b]	*t* value	*P* value	Correlation Coefficient[c]
STD1-STD2	-0.065	1194	-1.360	0.174	0.95
IVT1-IVT2	0.620	1934	0.772	0.440	0.93
SMART1-SMART2	-0.361	1740	-3.771	0.001	0.85
STD1-IVT1	0.010	1253	0.111	0.911	0.84
STD2-IVT1	-0.001	1088	-0.005	0.996	0.85
STD1-IVT2	-0.172	1366	-1.296	0.195	0.78
STD2-IVT2	-0.167	1192	-1.414	0.157	0.77
Average STD-Average IVT[d]	-0.097	965	-1.097	0.273	0.83
STD1-SMART1	0.199	1243	4.436	0.0001	0.84
STD2-SMART1	0.262	1078	4.292	0.0001	0.81
STD1-SMART2	-0.245	1348	-1.761	0.078	0.70
STD2-SMART2	-0.270	1175	-1.636	0.102	0.70
Average STD-Average SMART[d]	-0.035	940	-0.376	0.707	0.79

Hypothesized difference = 0

[a]The numbers 1 and 2 stand for the first and replicate experiment, respectively.

[b]The number of genes that gave significant fluorescent signals above background in both datasets compared.

[c]The correlation coefficients were calculated after the $\log_{10}$ transformation of the data.

[d]These comparisons were performed using the average expression ratios from replicate experiments. STD, standard direct labeling of RNA during reverse transcription; IVT, in vitro transcription; and SMART, exponential amplification by SMART cDNA synthesis and PCR.

Table 2. Comparison of 12 Differentially Expressed Genes

GenBank® Accession Nos.	Qualitative RT-PCR[a]	STD1	STD2	IVT1	IVT2	SMART1	SMART2
AA049537	6.55	**7.14**	**6.25**	**3.57**	2.00	-1.25	—
AA058055	3.23	**5.26**	**2.94**	-3.85	-7.69	—	-5.26
W82946[a]	—	—	—	-16.67	-14.29	-7.69	-20.00
AA050371	-2.94	**-2.70**	**-2.44**	**-12.50**	**-11.11**	—	**-6.67**
AA118976	-2.38	**-3.70**	**-8.33**	-1.61	1.56	3.45	2.04
W36382	5.15	1.06	1.15	**4.35**	**9.09**	1.61	**3.45**
W47799[a]	—	100.00	—	—	50.00	90.90	30.30
AA175329	-20.00	**-11.11**	—	**-50.00**	**-50.00**	**-33.33**	**-20.00**
AA175441	-12.50	—	—	**-10.00**	**-12.50**	**-5.56**	**-7.69**
W13718	1.59	-6.25	-4.35	**1.32**	**1.41**	**1.43**	2.22
AA124396	1.85	**1.03**	**1.10**	4.55	4.35	3.03	5.03
AA011839	-4.17	—	—	**-11.11**	**-10.00**	**-11.11**	**-9.09**
Correctly classified (bold)	—	6	5	7	6	4	5
Incorrectly classified	—	2	2	3	4	4	4
No Data	2	3	5	1	0	2	1

Comparison of differentially expressed genes from mouse testis and spleen showed inconsistent or contradictory expression rations among labeling methods between microarray analysis and semiquantitative RT-PCR. Negative values are underexpressed ratios.

[a]In the cases of W82946 and W47799, no amplification products were obtained from either spleen or testis RNA by quantitative RT-PCR.

SpleenCy3 ratio distributions between each method (Figure 2) and compared TestisCy5/SpleenCy3 replicate experiments within each method (Figure 3). For the second question, we compared directly the TestisCy5/SpleenCy3 of the amplification methods with the standard labeling procedures (Figures 4 and 5). In addition, 12 genes with ratios that were inconsistent or contradictory among the three methods were verified via quantitative RT-PCR (see Figure 1).

The reproducibility was very high for the direct labeling of RNA during reverse transcription and in vitro transcription but slightly lower for SMART. For SpleenCy5/SpleenCy3 hybridizations, over 99% of the spots showed a Cy5/Cy3 ratio that was within a 1.5-fold variation from the expected mean; for in vitro transcription, this was 96%, and for SMART, 84%. Using standard reverse transcription and in vitro transcription protocols, approximately 0.3% of the clones have a ratio greater than 2 or less than 0.5, while in the case of SMART, up to 1.5% of these outliers were observed. This could also be observed in the spleen/testis hybridizations. These hybridizations were repeated for each method, starting with labeling a different batch of RNA extracts from BD Biosciences Clontech. The correlation coefficients for Cy5/Cy3 ratios among the repeats were high for all three methods (Table 1). However, using a Paired *t* test, SMART replicates were significantly different statistically. Overall, SMART scored lower in these comparisons than the standard reverse transcription and in vitro transcription and was highly dependent on a careful optimization of the number of PCR cycles in each experiment.

Although the reproducibility was high within each method, the direct comparison between labeling strategies did reveal different testis/spleen ratios and different absolute fluorescent intensities for some genes. Comparing testis/spleen ratios of the amplification methods to the standard reverse transcription again showed lower correlation coefficients than among repeats within each method (Table 1). The correlation coefficients among the methods ranged from 0.77 to 0.85 for the comparison between in vitro transcription and reverse transcription and from 0.70 to 0.84 for the comparison between SMART and reverse transcription. Correlation coefficients between in vitro transcription and reverse transcription were not statistically significant, but they were between the SMART and reverse transcription protocols depending on the repeat. Focusing on the greater than 2-fold differen-

tially expressed genes, in vitro transcription and SMART, respectively, did not detect 20%–23% of the total number of differentially expressed genes measured by reverse transcription (Figure 5). Otherwise, there were 16–33 more differentially expressed genes determined by the amplification methods. Another important observation was that absolute fluorescent intensities were different among the three methods (Figure 6). This difference in absolute fluorescence was consistent among the repeat experiments and therefore likely to be dependent on the method used. However, it did not immediately affect the expression ratios as long as the same amplification method was used for all samples tested. Interestingly, the fraction of spots with fluorescent signals significantly above background was different among the various methods, being 12%–16% higher for in vitro transcription and SMART (data not shown).

Because the standard labeling method was not necessarily the best method to conserve the initial transcript ratios, a set of genes that did not show consistent testis/spleen transcript ratios among the three methods was selected for confirmation by quantitative RT-PCR (Table 2). The results could be grouped by (*i*) correctly classified differential expression (ratio of microarray data was in agreement with quantitative RT-PCR), (*ii*) incorrectly classified differential expression, or (*iii*) missing data (no fluorescent signals above background). Two clones did not result in RT-PCR products, probably due to suboptimal primer binding. The results showed that only 2 out of 12 genes were incorrectly classified by reverse transcription, 3–4 in in vitro transcription repeats, and 4–5 in SMART repeats. Interestingly, after the in vitro transcription amplification, less missing data were observed than in the two other methods, which may mean that amplifying the starting material may increase the sensitivity of the assay.

Overall, our results showed that amplifying RNA starting material can produce reproducible microarray data but induces slight distortions in the initial transcript levels. However, for in vitro transcription, the percentage of potentially misinterpreted gene expression levels is negligible, and more data points above the background level could be observed. Therefore, we suggest that direct labeling is preferred when micrograms of poly(A) RNA are available and that in vitro transcription can be considered a good alternative if only a few micrograms of total RNA can be obtained or when large-scale extractions may nega-

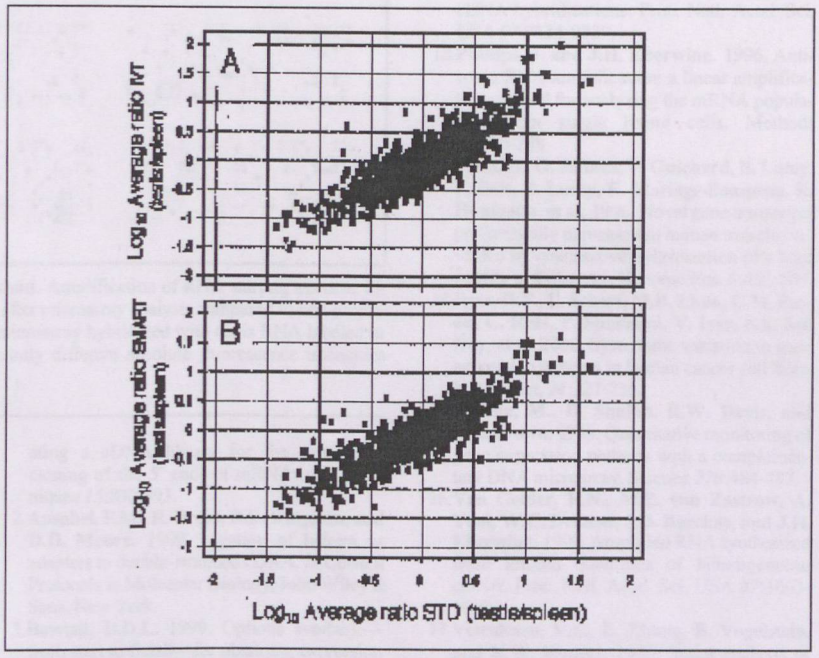


Figure 4. Reproducibility and preservation of the differential expression levels of testis versus spleen among labeling methods. The figure shows scatter plots of $\log_{10}$ average ratio (testis/spleen) from (A) STD1 versus IVT1 and (B) STD1 versus SMART1. The only average values considered were those for which both replicate spots had fluorescent signals significantly above background and also had a less than 2-fold difference from one another.

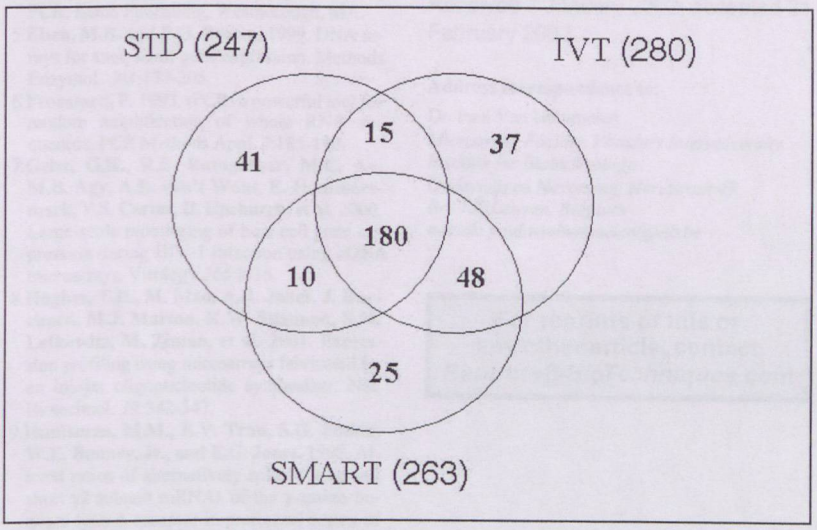


Figure 5. Venn diagram of differentially expressed genes. Comparison of standard labeling (STD), linear amplification (IVT), and logarithmic amplification (SMART) was made. The total number of differentially expressed genes per method is given between brackets. A 2-fold threshold allowing a 10% window determined the differentially expressed genes.

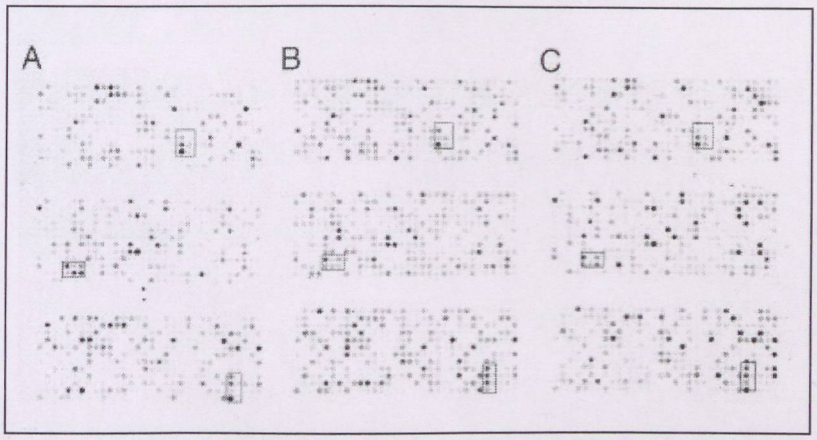


Figure 6. Microarray results for each labeling method. Amplification of RNA starting material induces changes of the absolute fluorescence intensities after microarray analysis compared to non-amplification. The images represent the same portion of a microarray hybridized with testis RNA labeled in Cy5. As an example, some spots exhibiting significantly different absolute fluorescence intensities among the techniques are boxed.

tively affect the RNA quality. SMART amplification can be used when the RNA is really limiting, such as in the case of cells obtained after cell sorting or laser capture microscopy. However, using different labeling or amplification strategies within one experiment should be avoided to retain the high reproducibility between the samples. Nevertheless, it is clear that the sooner robust methods become available that can label small amounts of RNA without amplification, the better.

ACKNOWLEDGMENTS

The authors thank Dr. Matt Coleman for fruitful discussions and K. Coddens, R. Maes, and K. Seeuws from the Flanders Interuniversity Institute for Biotechnology Microarray Facility for their excellent technical help. This work was supported in part by Biotechnológia 2000 grant nos. BIO-0005/2000 from OMFB and F034637 from OTKA and the Flanders Interuniversity Institute for Biotechnology. L.G.P. was supported by a European Molecular Biology Organization (EMBO) short fellowship (ref. no. ASTF 9633).

REFERENCES

1. **Apte, A.N. and P.D. Siebert.** 1993. Anchor-ligated cDNA libraries: a technique for generating a cDNA library for the immediate cloning of the 5′ ends of mRNAs. BioTechniques *15*:890-893.
2. **Ausubel, F.M., R. Brent, R.E. Kingston, and D.D. Moore.** 1999. Ligation of linkers or adapters to double-stranded cDNA. *In* Current Protocols in Molecular Biology, John Wiley & Sons, New York.
3. **Bowtell, D.D.L.** 1999. Options available—from start to finish—for obtaining expression data by microarray. Nat. Genet. *21*:25-32.
4. **Chenchik, A., Y.Y. Zhu, L. Diatchenko, R. Li, J. Hill, and P.D. Siebert.** 1998. Generation and use of high-quality cDNA from small amounts of total RNA by SMART PCR, p. 305-319. *In* P.D. Siebert and J.W. Larrick (Eds.), Gene Cloning and Analysis by RT-PCR. Eaton Publishing, Westborough, MA.
5. **Eisen, M.B. and P.O. Brown.** 1999. DNA arrays for analysis of gene expression. Methods Enzymol. *303*:179-205.
6. **Froussard, P.** 1993. rPCR: a powerful tool for random amplification of whole RNA sequences. PCR Methods Appl. *2*:185-190.
7. **Geiss, G.K., R.E. Bumgarner, M.C. An, M.B. Agy, A.B. van't Wout, E. Hammersmark, V.S. Carter, D. Upchurch, et al.** 2000. Large-scale monitoring of host cell gene expression during HIV-1 infection using cDNA microarrays. Virology *266*:8-16.
8. **Hughes, T.R., M. Mao, A.R. Jones, J. Burchard, M.J. Marton, K.W. Shannon, S.M. Lefkowitz, M. Ziman, et al.** 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat. Biotechnol. *19*:342-347.
9. **Huntsman, M.M., B.V. Tran, S.G. Potkin, W.E. Bunney, Jr., and E.G. Jones.** 1998. Altered ratios of alternatively spliced long and short γ2 subunit mRNAs of the γ-amino butyrate type A receptor in prefrontal cortex of schizophrenics. Proc. Natl. Acad. Sci. USA *95*:15066-15071.
10. **Khan, J., R. Simon, M. Bittner, Y. Chen, S.B. Leighton, T. Pohida, P.D. Smith, Y. Jiang, et al.** 1998. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res. *58*:5009-5013.
11. **Lee, M.-L.T., F.C. Kuo, G.A. Whitmore, and J. Sklar.** 2000. Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. Proc. Natl. Acad. Sci. USA *97*:9834-9839.
12. **Phillips, J. and J.H. Eberwine.** 1996. Antisense RNA amplification: a linear amplification method for analyzing the mRNA population from single living cells. Methods *10*:283-288.
13. **Pietu, G., O. Alibert, V. Guichard, B. Lamy, F. Bois, E. Leroy, R. Mariage-Sampson, R. Houlgatte, et al.** 1996. Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. Genome Res. *6*:492-503.
14. **Ross, D.T., U. Scherf, M.B. Eisen, C.M. Perou, C. Rees, P. Spellman, V. Iyer, S.S. Jeffrey, et al.** 2000. Systematic variation in gene expression patterns in human cancer cell lines. Nat. Genet. *24*:227-235.
15. **Schena, M., D. Shalon, R.W. Davis, and P.O. Brown.** 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*:484-487.
16. **Van Gelder, R.N., M.E. von Zastrow, A. Yool, W.C. Dement, J.D. Barchas, and J.H. Eberwine.** 1990. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc. Natl. Acad. Sci. USA *87*:1663-1667.
17. **Velculescu, V.E., L. Zhang, B. Vogelstein, and K.W. Kinzler.** 1995. Serial analysis of gene expression. Science *270*:484-487.
18. **Watson, A., A. Mazumder, M. Stewart, and S. Balasubramanian.** 1998. Technology for microarray analysis of gene expression. Curr. Opin. Biotechnol. *9*:609-614.

Received 7 January 2002; accepted 21 February 2002.

Address correspondence to:
Dr. Paul Van Hummelen
Microarray Facility, Flanders Interuniversity Institute for Biotechnology
Onderwijs en Navorsing, Herestraat 49
B-3000 Leuven, Belgium
e-mail: paul.vanhummelen@vib.be

Short Technical Report

Production of Bulk Amounts of Universal RNA for DNA Microarrays

BioTechniques 33:898-904 (October 2002)

L.G. Puskás, Á. Zvara, L. Hackler Jr., T. Micsik, and P. van Hummelen[1]
Hungarian Academy of Sciences, Szeged, Hungary, and [1]Flanders Interuniversity Institute for Biotechnology (VIB), Leuven, Belgium

ABSTRACT

In DNA microarray technology, repeatability and reliability are very important to compare multiple RNA samples from different experiments. The application of common or universal RNA as a standard control equalizes the differences in hybridization parameters and array variations. For this purpose, high-quality reference RNA is necessary in bulk amounts. A novel approach was developed to get milligrams of sense or antisense RNA, starting from micrograms of pooled total RNA from different cell lines, tissues, or organisms. This method is inexpensive and allows further labeling procedures using poly(dT) or random oligomers as primers. In addition, amplified, sense reference RNA is suitable for standard labeling protocols, while the antisense reference RNA can be used with antisense RNA from the linear sample amplification method. Here we produced universal RNA for human, rat, and alfalfa and demonstrated the quality using specific cDNA microarrays.

INTRODUCTION

In DNA microarray technology for each hybridization, a mixture of two fluorescent labeled probes is usually applied, where one labeled probe is obtained from a control (untreated, unaffected) and the other is from a treated or affected sample (5). This direct comparative hybridization method allows a quantitative comparison of the relative abundance of individual sequences, although experimental variation introduced by the uneven incorporation of labels, differences in hybridization, washing, and reading often occurs (3,5). It can cause discrepancies in relative and accurate comparisons of separate experimental results, especially when they are performed by different research laboratories or at different times. Many efforts have been made to minimize these variations by purchasing accurate hybridization chambers sold by different companies, although they cannot solve the problems caused by the incorporation and uneven distribution of samples on microarrays. While kits are available to optimize the labeling step, there is no consensus, even in protocols applied by different groups. The most reliable approach to overcome these problems is the indirect comparison of signal intensities. In these kinds of experiments, each mRNA sample is compared to a reference mRNA pool composed of an equal mixture of all experimental RNA targets. The importance of this approach was first suggested by Eisen and Brown (2) and used in the comparison of different cancer cell lines (7) and different non-Hodgkin's lymphomas (1). The application of a common reference RNA pool in each cDNA microarray experiment allows the relative gene expression in multiple samples to be analyzed. It is very difficult to obtain reliable reference RNA not only because of its need for complexity but also because of the large quantity. Today, universal reference RNA is commercially available but is extremely expensive for routine use. To obtain reference RNA from pooled, high-quality RNA, RNA isolation should be performed from cell lines representing different tissues or from different tissue samples. However, this is costly and time consuming if one wants to obtain milligrams of highly purified RNA.

We developed a novel amplification technique based on PCR amplification and a modified version of an in vitro amplification (http://www.microarrays.org/pdfs/ModifiedEberwine.pdf) to obtain high-quality reference RNA in bulk amounts, starting from micrograms of mixed total RNA. Amplification of mRNAs without distortion of their initial ratios was previously presented (6). Both antisense and sense amplified RNA can be synthesized with this method and can be used as standards in diverse microarray studies.

MATERIALS AND METHODS

Construction of cDNA Microarrays

PCR-amplified cDNA fragments were spotted in duplicate on amino-alkyl-silane-treated microscope slides (Silane-Prep™ Slides; Sigma, St. Louis, MO, USA) using a MicroGrid Total Array System printer (BioRobotics, Cambridge, UK). The cDNA clones were obtained from mixed libraries, cloned in pBluescript® SK II- (New England Biolabs, Hertfordshire, UK) or pGEM® (Promega, Madison, WI, USA) plasmids with standard cloning techniques. The cDNA inserts were amplified with vector-specific primers, purified with the Montage™ PCR_{96} Cleanup Kit (Millipore, Bedford, MA, USA), resuspended in 50% DMSO/water, and loaded into a 384-well format microplate. Post-processing of DNA arrays was performed as described previously (6).

RNA Extraction and RNA Pools

For the production of rat reference RNA, commercially available total RNA were pooled. RNA (5 µg) from each of the following tissues were mixed: liver, brain, thymus, heart, lung, spleen, testicle, ovary, and kidney. Rat embryo total RNA (35 µg) was added to this mixture. All of the RNA was purchased from Ambion (Austin, TX, USA). For human reference RNA, 60 µg Universal Human Reference RNA (pooled from 10 different cell lines representing different tissues; Stratagene, La Jolla, CA, USA) was mixed with 5 µg human heart total RNA (Ambion), 5 µg human fetal brain total RNA (BD Biosciences Clontech, Palo Alto, CA, USA), 5 µg human thyroid total RNA, and 5 µg total RNA from human carotid tissue. Total RNA from thyroid and carotid tissue were purified with NucleoSpin® RNA II extraction kit (Macherey-Nagel, Düren, Germany). For alfalfa reference RNA, 14 µg total RNA from the following tissues were pooled: flower, leaf, root, seed, and germ. This RNA was mixed with 10 µg total RNA prepared from protoplasts. For RNA purification the FastRNA® kit (BIO 101, Vista, CA, USA) was used. The quality of the extracted RNA was assessed by gel electrophoresis, as well as by $A_{260/280}$ ratios.

Reference cDNA Synthesis

Thirty-three microliters pooled total RNA (80 µg) from different organisms were reverse transcribed in a total volume of 60 µL. For the synthesis of antisense reference RNA, total RNA was mixed with 3 µL (100 pmol/µL) $T7T_{25}V$ and 3 µL (100 pmol/µL) FOR primers, and 3 µL (20 U/µL) RNasin® (MBI Fermentas, Vilnius, Lithuania), heated to 75°C for 5 min and cooled on ice. To this mixture, 12 µL 5× first-strand buffer (MBI Fermentas), 3 µL 10 mM dNTP mixture, and 3 µL, 200 U/µL RNase H^- point mutant Moloney murine leukemia virus (MMLV) reverse transcriptase (MBI Fermentas) were added and incubated at 42°C for 2 h. For the production of sense reference RNA, the same protocol is used as for cDNA synthesis, but $REVT_{25}V$ and T7FOR primers were used instead of $T7T_{25}V$ and FOR primers. After cDNA synthesis, 240 µL TE (10 mM Tris-HCl, pH 7.5, 0.2 mM EDTA) were added.

Exponential Amplification of Reference cDNA

From the first-strand cDNA synthesis reaction mixture, 0.5 µL were amplified in a total volume of 100 µL PCR with $REVT_{25}V$ and T7FOR, or REVT and T7FOR primers in case of sense reference RNA production, or $T7T_{25}V$ and FOR, or T7T and FOR primer pairs in case of antisense amplified RNA production. The reaction mixture contained 50 nM final concentration of each primers, 1× PCR buffer (Amersham Biosciences, Piscataway, NJ,

USA), 4 U *Taq* DNA polymerase (Amersham Biosciences), and 200 µM dNTP. Amplification was done under the following conditions: initial denaturation at 95°C for 1 min, 22 cycles of denaturation at 95°C for 35 s, annealing at 58°C for 40 s, and extension at 68°C for 4 min. PCR products were purified with a PCR purification kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions, but eluted twice with 40 µL 1/5 Elution buffer. The concentration of the eluted dsDNA was determined by UV absorbance measurement at 260 nm. The usual yield was approximately 2 µg DNA from 100 µL PCR. Forty reactions were performed from each cDNA synthesis reaction mixture from the different pooled samples. The purified PCR products were pooled, vacuum-concentrated using an Automatic SpeedVac AS160 (Savant Instrument, Farmingdale, NY, USA) and dissolved in DEPC-treated water to a final concentration of 500 ng/µL.

Production of Reference RNA Using In Vitro Transcription

PCR-amplified cDNA template (20 µg) was transcribed in a total volume of 200 µL using Ribomax Large Scale RNA Production System (Promega), according to the manufacturer's instructions. The RNA was diluted with 0.3 mL DEPC-treated water, purified using 15 columns from the RNeasy® purification kit (Qiagen), and measured spectroscopically. The usual yield was 1.5–2 mg amplified RNA, which corresponds to 150–200× amplification of the PCR-amplified DNA. For long-term storage, RNA was aliquoted (aliquots contained 30 µg amplified RNA) in the presence of RNasin (final concentration of 0.1 U/µL), dried using speed vacuum, and stored at -80°C.

Probe Preparation and Labeling

Amplified reference RNA (2.5 µg) was labeled with Cy3 during reverse transcription using 0.4 µM random nonamer as primer, 0.1 mM d(G/T/A)TPs, 0.05 mM dCTP (Amersham Biosciences), 20 U RNasin (MBI Fermentas), 1× first-strand buffer, 200 U RNAse H^- point mutant MMLV reverse transcriptase, and 0.05 mM Cy3-dCTP (Amersham Biosciences) in 20 µL total volume. The RNA, primer, and RNasin were denatured at 75°C for 5 min and cooled on ice before adding the remaining reaction components. After 2 h incubation at 37°C, the heteroduplexes were denatured and the mRNA was hydrolyzed with NaOH (250 mM final concentration) for 15 min at 37°C and neutralized with 10 µL 2 M MOPS (pH 6.0). The labeled cDNA was purified with a PCR purification kit, according to the manufacturer's instructions, but

Table 1. Sequences of Oligodeoxynucleotides Used

Name	Sequence (5′→3′)
$T7T_{25}V$	GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG$(T)_{25}V$
$REVT_{25}V$	TGTCTGCAGTGGTAACAACGCAGAGTACG$(T)_{25}V$
T7T	GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGGTTT
REVT	TGTCTGCAGTGGTAACAACGCAGAGTACGTTT
T7FOR	GGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGGG
FOR	TGTCTGCAGTGGTAACAACGCAGAGTACGCGGG

eluted twice with 40 µL 1/2 Elution buffer. The probes were concentrated by speed vacuum and dissolved in 20 µL hybridization buffer (50% formamide, 5× SSC, 0.1% SDS, 100 µg/mL salmon sperm DNA).

Array Hybridization and Post-hybridization Processes

In case of human heart expression monitoring, prehybridization was done by adding 1 µL 4 mg/mL poly(T), 2 µL 2 mg/mL human Cot DNA (Invitrogen), and 2 µL 1 mg/mL λ DNA (MBI Fermentas) to the hybridization mixture. For hybridization of labeled reference cDNA, 2 µL 1 mg/mL λ DNA and 2 µL 1 mg/mL salmon sperm DNA were added. The probe mixture was incubated at 42°C for 30 min after denaturation by heating for 5 min at 80°C. Twenty microliters of the mixture were placed on the blocked array under a 24 × 32 mm coverslip (Menzel-Glaser, Germany). To the edges of the coverslip DPX Mountant (Fluka) was poured to prevent evaporation. Slides were incubated at 42°C for 18 h in a humid hybridization chamber. After hybridization, the mountant was removed and the arrays were washed as described (6). Scanning and data analysis were performed as described previously (4). Briefly, following image analysis, genes were labeled ON or OFF according to a predetermined intensity threshold. The threshold was set at 1.5 times of the local background intensity. This cut-off level was determined from the "CH1GTB2" values (obtained from the ScanAlyze2 software, version 2.32; www.microarrays.org/software.html), which corresponds to a fraction of pixels in the spot greater than 1.5 times the background. Only if the mean spot intensity was greater than this threshold was a spot considered significantly above background. Each clone was spotted twice; thus, an average intensity could be calculated for the replicate spots.

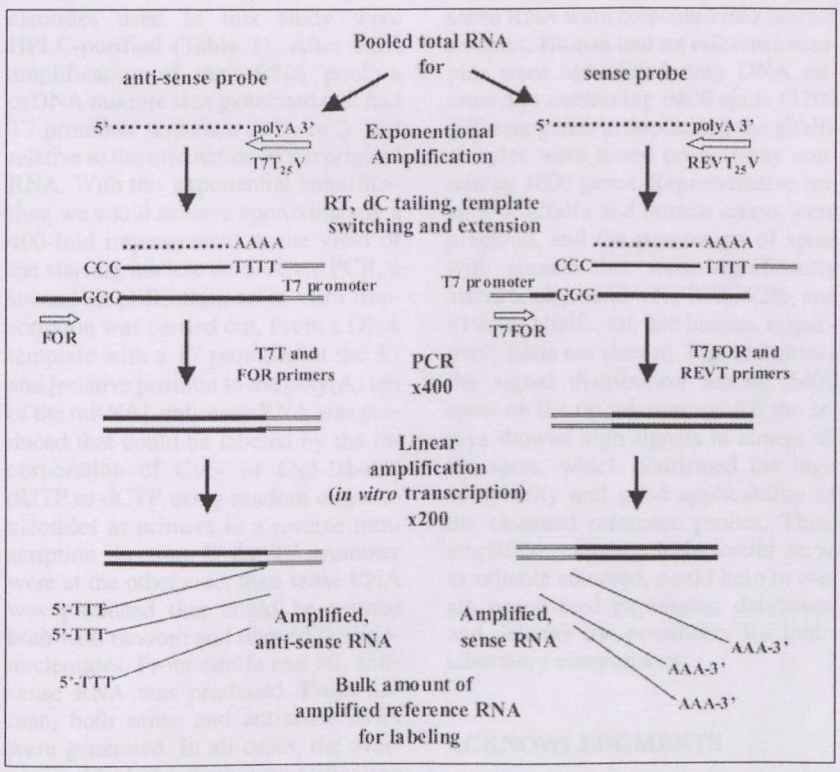


Figure 1. Summary of the method used in this study. A bulk amount of sense and antisense reference RNA can be obtained with combinations of exponential (PCR) and linear (in vitro transcription) amplification methods.

RESULTS AND DISCUSSION

The application of common reference RNA provides an internal control and thus normalizes differences in hybridization parameters and array variations. The reference RNA should represent a mixture of different gene products with the highest possible complexity. The reference RNA, produced by our method corresponds to poly(A)+ sequences and can be prepared in bulk amounts sufficient for thousands of experiments at low cost.

Mixtures of total RNA isolated from different tissues deriving from different organisms, including human, rat, and alfalfa were prepared. In case of a human total RNA pool, commercially available Universal Human Reference RNA was mixed with additional total RNA isolated from different tissues to increase the complexity of the mixture. In case of the other samples, purified or commercially available total RNA were mixed. We used these mixtures as starting material for amplification to obtain labeled sense or antisense reference cDNA.

To obtain a bulk amount of reference RNA starting from 80 µg pooled total RNA, a double amplification technique was applied. The summary of the method can be seen in Figure 1. At first, total RNA pools from different organisms were reverse-transcribed using a modified SMART™ cDNA synthesis technique (BD Biosciences Clontech). For the production of antisense probe, $T7T_{25}V$ and FOR primers were added

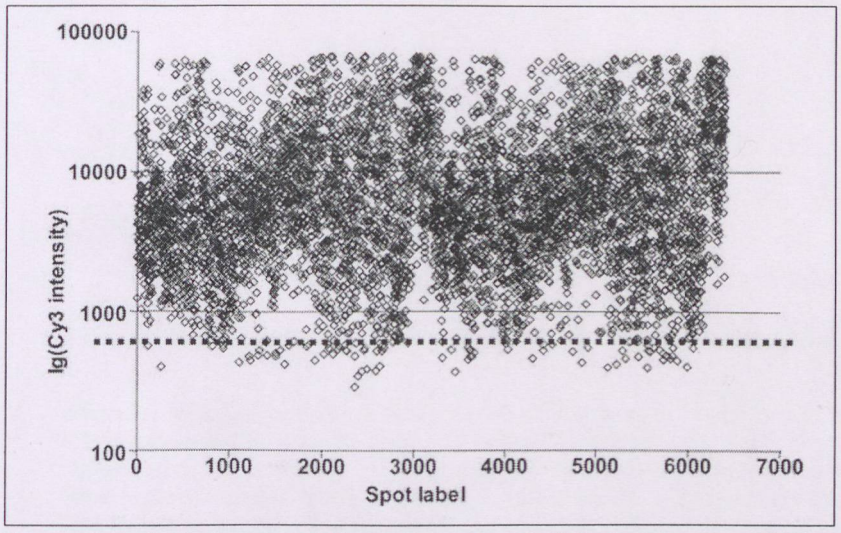


Figure 2. Signal distribution among 6400 spots (for number of individual samples, see spot label on x-axis) on the rat microarray demonstrates that 92% of the spots have significantly higher intensity than background (the ratio of the mean spot and local background intensities was higher than 1.5).

to the reverse transcription mixture, while in case of sense probe preparation, $REVT_{25}V$ and T7FOR primers were used. T7 denotes for the T7 RNA polymerase promoter sequence at the 5′-end of the primers. All the oligonucleotides used in this study were HPLC-purified (Table 1). After PCR amplification of the cDNA pool, a dsDNA mixture was generated that had T7 promoter sequence at 5′- or 3′-end relative to the orientation of the original RNA. With this exponential amplification, we could achieve approximately a 400-fold improvement in the yield of the starting nucleic acid. After PCR, a second amplification, an in vitro transcription was carried out. From a DNA template with a T7 promoter at the 3′-end [relative position to the poly(A) tail of the mRNA], antisense RNA was produced that could be labeled by the incorporation of Cy5- or Cy3-labeled dUTP or dCTP using random oligonucleotides as primers in a reverse transcription reaction. If the T7 promoter were at the other side, then sense RNA was produced that could be primed both with random and oligo(dT) oligonucleotides. From alfalfa and rat, antisense RNA was produced. From human, both sense and antisense RNA were generated. In all cases, the average yield of the double amplification was 1.5–2 mg RNA, starting from 1/12 of the PCR-amplified templates.

To test the quality of the reference RNA, we synthesized Cy3 fluorescently labeled cDNA samples from all the organisms and hybridized them to the species-specific microarrays. Alfalfa, rat, human antisense RNA, and human sense RNA were converted into labeled samples. Human and rat reference samples were hybridized onto DNA microarrays containing 6400 spots (3200 different genes in duplicate); the alfalfa samples were tested on an array containing 1600 genes. Representative images of alfalfa and human arrays were prepared, and the percentage of spots with signals that were significantly above background were 90%, 92%, and 91% for alfalfa, rat, and human, respectively (data not shown). Figure 2 shows the signal distribution among 6400 spots on the rat microarray. All the arrays showed high signals in almost all the spots, which confirmed the high complexity and good applicability of the obtained reference probes. Thus, amplified reference RNA could serve as reliable standard, could help to create normalized expression databases, and delivers the possibility for interlaboratory comparisons.

ACKNOWLEDGMENTS

This work was supported by Hungarian National Science and Development Foundation grant no. NKFP 1/040, Biotechnológia 2000 grant no. OMFB, BIO-0005/2000, and Hungarian National Scientific Research Foundation (OTKA) grant no. F034637. The authors thank S. Bottka and Z. Kupihár for the preparation and purification of the oligonucleotides, and A. Ábrahám and R.T. Dudásné for their technical support.

REFERENCES

1. **Alizadeh, A.A., M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J.C. Boldrick, H. Sabet, et al.** 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*:503-511.
2. **Eisen, M.B. and P.O. Brown.** 1999. DNA arrays for analysis of gene expression. Methods Enzymol. *303*:179.
3. **Hughes, T.R., M. Mao, A.R. Jones, J. Burchard, M.J. Marton, K.W. Shannon, S.M. Lefkowitz, et al.** 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat. Biotechnol. *19*:342-347.
4. **Kitajka, K., L.G. Puskás, A. Zvara, L. Hackler, Jr., G. Barceló-Coblijn, Y.K. Yeo, and T. Farkas.** 2002. The role of n-3 polyunsaturated fatty acids in brain. Modulation of rat brain gene expression by dietary n-3 fatty acids. Proc. Natl. Acad. Sci. USA *99*:2619-2624.
5. **Lee, M-L.T., F.C. Kuo, G.A. Whitmore, and J. Sklar.** 2000. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc. Natl. Acad. Sci. USA *97*:9834-9839.
6. **Puskás, L.G., A. Zvara, L. Hackler, Jr., and P. Van Hummelen.** 2002. RNA amplification results in reproducible microarray data with slight ratio biases. BioTechniques. *32*:1330-1340.
7. **Scherf, U., D.T. Ross, M. Waltham, L.H. Smith, J.K. Lee, L. Tanabe, K.W. Kohn, W.C. Reinhold, et al.** 2000. A gene expression database for the molecular pharmacology of cancer. Nat. Genet. *24*:236-244.

Received 28 February 2002; accepted 2 May 2002.

Address correspondence to:

Dr. László G. Puskás
DNA-chip Laboratory
Biological Research Center
Hungarian Academy of Sciences
Szeged, P.O. Box 521, H-6701, Hungary
e-mail: pusi@nucleus.szbk.u-szeged.hu

Article is available online at http://www.webio.hu/por/2002/8/4/0231

New Molecular Methods for Classification, Diagnosis and Therapy Prediction of Hematological Malignancies

Ágnes ZVARA, László HACKLER jr., P. Zsolt NAGY, Tamás MICSIK, László G. PUSKÁS

Laboratory of Functional Genomics, Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary

Normal functions of the cell are based on the precise regulation of various genes. If this strict regulation and the hierarchy of genes becomes upset due to flaws in this system, the result will be cellular dysfunction which eventually may lead to carcinogenic transformation. Two basic challenges of the classification of cancers are the discovery of new molecular markers characteristic to defined disease groups and the classification of already diagnosed or new cases into existing groups. This precise classification may open the door to tailored treatment or predict the expected outcome of the disease. Today, there is a universal access available to the databases containing sequences and localization of the genes within the confines of Human Genome project. It provides significant help for the discovery of characteristic alterations and systematic analysis of gene expression patterns. This

Keywords: molecular methods, DNA chip, cancer classification

Classification of tumors, current universally used according to tissue of origin or histological characteristics. However, genetic defects that literally determine the abnormal behavior of tumor cells can be analyzed only with precisely using the knowledge of almost the total sequence of the human genome. New, high-throughput molecular methods have been developed (microarray analysis of the whole genome, comparative genomic hybridization, gene expression monitoring with cDNA microarrays, array-based protein expression and interaction analysis) to obtain more information about different malignant transformations at different cellular levels (genomic, transcriptional, translational, protein). Results obtained with these methods can help to answer the questions arising

Received: July 18, 2002; accepted: Dec 11, 2002

Correspondence: László G. PUSKÁS, Laboratory of Functional Genomics, Biological Research Center, Hungarian Academy of Sciences, Temesvári krt. 62, H-6726 Szeged, Hungary

is important not only to understand normal functions of the cells, but it also contributes to the identification of new genes that are characteristic to given disease groups as markers and that are potential drug targets. Until the second half of the twentieth century the study of the function and regulation of genes was based on step-by-step investigation of individual genes. Regarding the fact, that the genomes of an increasing number of organisms have become known in whole or in part, numerous new techniques have been developed that facilitated the systematic analysis of gene functions. The aim of this study is to summarize the new molecular based possibilities for classification, diagnosis and prognosis of hematological malignancies, as well as to summarize the main results of these areas. (Pathology Oncology Research Vol 8, No 4, 231–239)

in connection with prevention, diagnosis and classification, therapy and outcome of the disease.[1,2] The most studied disease today, the cancer, is characterized by alterations from the point of view of functional genomics. It is known

Acute leukemia is a complex disease that can be classified into individual sub-groups based on morphological, immunological and cytogenetic characteristics and their response to chemotherapy. Classification of the disease and therapeutic grouping of the patients are currently quite difficult and costly tasks, which require several laboratory tests (cytomorphology, cytochemistry and immunophenotyping, cytogenetic and molecular diagnostic tests as well as a well-arranged medical specialist background (hematologist, oncologist, pathologist, cytogeneticist). Individual and total testing cannot provide sufficient basis for diagnosis. Despite the thorough and laborious diagnostic procedure, the classification of leukemias is usually not perfect and, even more, from a clinical standpoint, the identification of

PATHOLOGY ONCOLOGY RESEARCH Vol 8, No 4, 2002

Article is available online at http://www.webio.hu/por/2002/8/4/0231

ARTICLE

New Molecular Methods for Classification, Diagnosis and Therapy Prediction of Hematological Malignancies

Ágnes ZVARA, László HACKLER Jr., B Zsolt NAGY, Tamás MICSIK, László G PUSKÁS

Laboratory of Functional Genomics, Biological Research Center, Hungarian Academy of Sciences, Szeged, Hungary

Normal functions of the cell are based on the precise regulation of various genes. If this strict regulation and the hierarchy of genes becomes upset due to flaws in this system, the result will be cellular dysfunction which eventually may lead to carcinogenic transformation. Two basic challenges of the classification of cancers are the discovery of new molecular markers characteristic to defined disease groups and the classification of already diagnosed or new cases into existing groups. This precise classification may open the door to tailored treatment or project the expected outcome of the disease. Today there is unlimited access available to the databases containing sequences and localization of the genes within the confines of Human Genome project. It provides significant help for the discovery of chromosome abnormalities and systematic analysis of gene expression patterns. This is important not only to understand normal functions of the cells, but it also contributes to the identification of new genes that are characteristic to given disease groups as markers and that are potential drug targets. Until the second half of the twentieth century the study of the function and regulation of genes was based on step-by-step investigation of individual genes. Regarding the fact, that the genomes of an increasing number of organisms have become known in whole or in part, numerous new techniques have been developed that facilitated the systematic analysis of gene functions. The aim of this study is to summarize the new, molecular based possibilities for classification, diagnosis and prognosis of hematological malignancies, as well as to summarize the main results of these areas. (Pathology Oncology Research Vol 8, No 4, 231–240)

Keywords: molecular methods, DNA-chip, oncohematology

Classification of human cancers traditionally based on tissue of origin or histological characteristics. However, genetic defects that basically determine the abnormal behavior of tumor cells, can be analyzed more effectively using the knowledge of almost the total sequence of the human genome. New, high throughput molecular methods have been developed (mutation analysis of the whole genome, comparative genomic hybridization, gene expression monitoring with cDNA microarrays, array-based protein expression and interaction analysis) to obtain more information about different malignant transformations at different system levels (genome, transcriptome, proteome) (*Table 1*). Results obtained with these methods can help to answer the questions arising in connection with prevention, diagnosis and classification, therapy and outcome of the disease.[8,31,61] The most studied disease among the carcinogenic transformations, from the point of view of functional genomics, is leukemia.

Acute leukemia is a complex disease that can be classified into individual sub-groups based on morphological, immunological and cytogenetic characteristics and their response to chemotherapy. Classification of the disease and therapeutic grouping of the patients are currently quite difficult and costly tasks, which require serious laboratory (e.g. immuno-phenotyping), cytogenetic and molecular diagnostic tests as well as a wide-ranging medical specialist background (hematologist, oncologist, pathologist, cytogeneticist). Individual and fast testing cannot provide sufficient basis for diagnosis. Despite the thorough and laborious diagnostic processes, the classification of leukemias is usually not perfect and error-prone.[52] From a clinical standpoint, the classification of

Received: Oct 18, 2002; *accepted:* Dec 15, 2002

Correspondence: László PUSKÁS, Laboratory of Functional Genomics, Biological Research Center, Hungarian Academy of Sciences, Temesvári krt. 62, H-6726 Szeged, Hungary

Table 1. **Comparative table of different methods for DNA, mRNA and protein analysis at different system levels.**

System level	*Technique*	*Application*	*Throughput*		*Biological requirement*
			Biological sample Sample No.	*Gene, transcript or protein No.*	
Genome (genes)	Sequencing	Detection deletions/ amplifications, mutations and polymorphic sites (SNP) analysis	medium	low	Genomic DNA from fresh or archive biological sample
	Chromosome band analysis (FISH)	Detection of chromosome abnormalities, deletions/ amplifications	low	low	Genomic DNA from fresh or archive biological sample
	Microsatellite analysis	Detection of chromosome abnormalities, deletions/ amplifications	medium	low	Genomic DNA from fresh or archive biological sample
	Array-based comparative genomic hybridization (CGH)	Detection of chromosome abnormalities, deletions/ amplifications	medium	high	Genomic DNA from fresh or archive biological sample
	Microarray-based methylation pattern analysis	Analysis of methylation pattern of the genome	medium	high	Genomic DNA from fresh biological sample
	DNA or oligonucleotide microarray	Mutation and polymorphic sites (SNP) analysis, detection of chromosome abnormalities, deletions/ amplifications	medium	high	Genomic DNA from fresh or archive biological sample
Transcriptome (mRNA)	Northern blot analysis	Gene expression analysis	low	low	RNA from fresh biological sample
	Differential display	Differential gene expression analysis	low	high	RNA from fresh biological sample
	Subtractive hybridization	Differential gene expression analysis	low	high	RNA from fresh bio logical sample
	cDNA Representational Difference analysis (RDA	Differential gene expression analysis	low	high	RNA from fresh bio logical sample
	Serial analysis of gene expression (SAGE)	Differential gene expression analysis	low	high	RNA from fresh biological sample
	Real –time PCR	Comparative gene expression analysis	high	low	RNA from fresh biological sample
	cDNA microarray	Expression profiling, differential gene expression analysis	medium	high	RNA from fresh biological sample
Proteome (protein)	Two-dimensional gel-electrophoresis	Analysis of protein profile	low	high	Denatured protein lysates
	Two-hybrid system	Protein-protein interactions	low	high	Cloned cDNA of protein of interest and cDNA library from given organism
	Protein arrays	Analysis of protein profile, protein-protein, protein-DNA interactions	medium	high	Native protein mixtures or denatured protein lysates

leukemias is important for two reasons: firstly to understand the initial causes of the disease and to project a view on the final processes; and secondly for diagnosis supported by molecular results toprovide optimal therapy and recovery.

Acute myeloid leukemia (AML) is classified into eight morphological subgroups (M0-M7), which all show a correlation with clinical appearance, cytogenetic markers and severity of the disease.[56] Classification of acute myeloid leukemias is currently based on cytogenetic differences instead of the earlier FAB system based on morphology (existence or lack of certain chromosome translocations). Leukemias can also be classified based on special chromosome abnormalities, which can be identified in more than 90 % of the cases.[27] A few chromosome abnormalities have therapeutic and prognostic significance.[5,26,2]

Acute lymphoid leukemias (ALL) can be classified into T and B cell types, and further subgroups identified based on cell surface markers, with the help of immunhistochemistry. The classification has prognostic significance in cases of both leukemias, but only in the statistical sense, because patients (especially AML) with the same cytogenetic classification may show individual variations. It is a frequent observation that leukemias classified into the same group based on various criteria follow a completely different clinical course, and react completely different ways to the same therapeutic intervention.

The first classification of ALL reflected various clinical outcomes of the disease and fine cytomorphological differences. At the end of the 1960's researchers tried to classify the disease with the help of a new method – enzyme based histochemistry. They managed to show that some of the leukemias proved to be acid-Schiff positive, while others proved to be myeloperoxidase positive with the periodate- acid- dying method. This gave the first basis for the molecular level classification of leukemias. According to this, we can differentiate between ALL with either thymus-T-cell-precursor origin (T-ALL) or bone-marrow-B-cell precursor origin (B-ALL) and AML. This classification gained further confirmation in the 1970's using antibodies recognizing cell surface antigens.[57] This review gives a short overview of the new, molecular-based, high throughput methods applied successfully for classification and therapy prediction of leukemias.

High throughput analysis of chromosome abnormalities based on comparative genomic hybridization (CGH)

Changes within the chromosome: deletion or amplification is quite frequent in neoplasias.[48] The specific rearrangements, in many cases, are characteristic to the individual tumor types and states. Genes mapped to the locations of these rearrangements can play roles in the formation of tumor. Their investigation can contribute to the characterization of the different tumors and tumor stages. Clinically distinct subgroups of ALL patients are characterized by chromosome translocations and chromosome imbalance. These abnormalities are the defining molecular features of ALL. In case of B-ALL the presence of a fusion oncoprotein may frequently be observed.[22] Patients with t(12;21)/TEl-AML1 and t(1;19)/E2A-PBX1 translocations is usually expected to give relatively good treatment outcome, especially by raising the dosage of chemotherapy, however the prognosis for patients with translocations t(9;22)/BCR-ABL and t(4;11)/MLL-AF4 are much worse.[6] Patients with B-ALL characterized by high chromosome number (hyperdiploid chromosome content, 50) have also good treatment outcome. T-ALL is characterized by repetitive but relatively rare chromosome translocation, which usually affects genes encoding for transcription factors (LYL1, HOX11, HOX11L2, TAL1), their abnormal expression impairs early thymocyte differentiation and finally leads to leukemia.[22]

Techniques based on *chromosome band analysis, fluorescence in situ hybridization (FISH)* have proven to be useful in detecting rearrangements and imbalances, but less informative in recognizing potentially amplified or deleted regions. Using this method, the detection of nucleotide sequences on DNA molecule is performed indirectly, by first hybridizing the seeker nucleotide sequences (probes) with the DNA (also called target). If the probes are synthesized with incorporated fluorescent molecules or antigenic sites that can be recognized with fluorescent antibodies, the direct visualization of the relative position of the probes is possible.[73] In addition, these techniques require analysis of metaphases after cell culture, which are time-consuming and misleading in the aspect of selecting subclones having growth advantage. The technique based on *microsatellite analysis* is a higher resolution possibility for chromosome region instability than FISH. Microsatellites are short, highly polymorphic tandem repeat sequences suitable for detection of chromosome imbalances. PCR amplification of these certain region give direct information about the copy number alteration of a given chromosome segment. The technique was applied successfully to detect submicroscopic chromosomal deletions in hematological malignancies, as well.[49, 65]

Comparative genomic hybridization (CGH) is a rapid method that does not require cell cultures and provide more accurate information about the possible occurrence of chromosome amplifications than do cytogenetic methods.[34] It provides a lot of information about the genomic balance of tumor cells, mono- or trisomies, amplifications and deletions in a simple experimental procedure. Using this technique, DNA purified from the tumor-tissue is labeled with green-fluorescent dye (Cy3),

while the DNA purified from the healthy control is labeled with another, red fluorescent dye (Cy5). They are hybridized together onto a DNA chip containing high number of genomic DNA fragments or cDNAs of already sequenced and identified genes immobilized on glass surface. After reading with a confocal laser scanner, differences in color can be measured and analyzed with computer software; the determined Cy3/Cy5 ratio indicates deletion, or amplification of the given gene (*Figure 1*).

Pinkel et al. performed an initial study in this area. They immobilized bacterium artificial chromosomes (BAC) and genomic fragments from human chromosome 20 on glass surface and demonstrated the feasibility of detecting both gains and losses with single copy sensitivity using array CHG.[50] Despite the fact, that the CGH method based on DNA microarrays is not suitable for the identification of small mutations (SNPs or deletions/amplifications of few nucleotide within genes), it offers a new solution for amplification/deletion analysis applied so far, because it is extremely well applicable and high throughput way for the overall analysis of the whole genome. It has the significant advantage of being less sensitive to cell contamination: a single gene-copy change may be detected from a sample containing up to 60% of normal, healthy cell contamination.[33] A great advantage is that good quality labeled probe can be obtained even from a small amount of paraffin-archived material. The results can be easily analyzed with the help of currently available databases that contain significant amount of information about the chromosomal location of the genes identified in the experiments.

In cases of adult T-cell leukemia, both deletion and amplification were successfully determined by this method. Testing of 64 patients showed amplification in 14q, 7q, and 3q chromosome regions, while in the regions 6q and 13q, deletions were observed. These chromosome changes were much more frequent in patients with an aggressive form of leukemia than in the indolent form. An increased number of chromosome imbalances were detected in patients, where the chance of survival was significantly lower.[68] In cases of acute myeloid leukemia, a high level of DNA amplification was detected at the chromosome region 8q24, the locus of the *c-myc* proto-oncogene using CGH.[10] The authors could subdivide the reported AML cases into two groups: cases with *c-myc* amplification together with further complex chromosome aberrations were associated with rapid disease progression and short survival time; while cases with *c-myc* amplification accompanied by a single other chromosome aberration (e.g. loss of one X chromosome or trisomy of chromosome 4) showed good response to chemotherapy and had prolonged survival time.

In these cases, data obtained from CGH experiments provided more information about the outcome of the disease, and distinguished between AML cases for diagnosis and therapy. The clonal identity of unusual metastases and the suspected primery tumor can be also successfully comfirmed by the CGH technique using cDNA microarray.[24]

In case of childhood ALL, deletion of the 12p chromosome region was also successfully shown with the CGH technique, which may offer favorable prognosis.[39] DNA chip-based CGH method may provide significant help in the identification of new genes that may play a role in tumorigenesis (novel oncogenes and tumor suppressors) including leukemia, in the follow up examination of metastases during the course of tumor development, and in classification of different cases. CGH has been used to determinate DNA copy number alterations in many other cancer types, like renal cancer,[74] breast cancer[7] or thyroid cancers.[76]

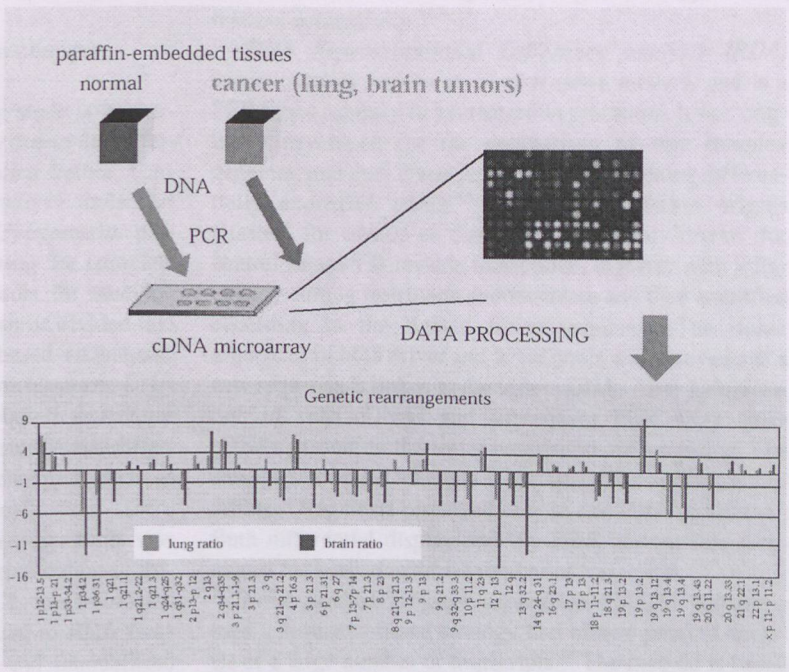


Figure 1. Comparative Genomic Hybridization for the detection of genetic rearrangements. Genetic rearrangements of two tumor tissues (brain and lung) have been compared in order to confirm that the brain metastasis derived from the lung cancer

Classification based on DNA-methylation-pattern analysis by DNA microarray

Aberrant DNA methylation pattern of CpG islands is one of the earliest and most common alterations in human malignancies. Several studies confirm the hypothesis that aberrant methylation occurs in a tumor-specific manner. Gain or loss of function of wide range of genes was shown in these cases.[38] Abnormal methylation patterns can be detected in CpG-rich regulatory elements either in introns or in exons.[12] Tumor cells are generally characterized by the hypermethylation of tumor suppressor genes and, in contrast, hypomethylation of the whole DNA molecule. This general hypomethylation can be detected relatively early, before the development of the actual tumor. Correlation between hypomethylation and increased gene expression can be detected in cases of large number of oncogenes.[21,18] With a genome-wide approach suitable for whole genome screening, it was shown that the methylation pattern may vary among different tumors.[15,35] The methylation pattern of the as whole genome, similarly to the gene expression pattern, is characteristic to the given cell, tissue, or disease as a molecular fingerprint. Analysis of genome wide methylation profiles enables the caracterization of new tumor classes, or the classification newly diagnosed cases into already existing groups based on their methylation pattern. In recent years a new method had emerged for the analysis of methylation pattern extending to the whole genome that is suitable for analyzing large number of genes simultaneously.[1,25,67]

Classification based on gene expression changes

A primary goal of expression profiling study is to characterize genes that expressed abnormally due to the different chromosomal abnormalities mentioned before. Currently there are two main ways to analyze molecular expression patterns: (1) generating mRNA-expression profile, the "transcriptome" and (2) analyzing the complete protein population, the "proteome". Tools for studying gene-expression at the transcript level can be divided into three major groups: (1) hybridization-based techniques, such as Northern blotting, subtractive hybridization, DNA microarrays or macroarrays, (2) PCR-based techniques such as differential display, RDA (representational difference analysis) (3) Sequence-based techniques such as SAGE (serial analysis of gene expression)[42]

The breakthrough in studying gene-expression was *Northern blot analysis* that made the identification of different mRNA in a given sample possible. In this technique, labeled RNAs or cDNA probes hybridized to RNA fragments separated by gel electrophoresis and immobilized onto nylon filter.[4] This technique is still a useful method for detecting a transcript in a given mRNA population and confirming expression data obtained with other experiments.

The first technique that enabled identification of differentially expressed genes was based on *subtractive hybridization*. This technique based on development of subtractive cDNA libraries, which are generated by hybridizing an mRNA pool of one origin to an mRNA pool of a different origin.[32] Transcripts that do not have a complementary strand in the hybridization step are then used for the construction of a cDNA library. A variety of refinements of this method have been developed to identify specific mRNAs.[64] Despite the fact that numerous genes were successfully identified with this method, it had serious disadvantages: 1. only a small fraction of the gene expression differences were successfully discovered, 2. it required a large amount of RNA sample, and 3. it proved to be quite laborious and time-consuming.

In 1992 it was replaced by a new technique, called *differential display PCR (DD-PCR)*. This technique was the first one-tube method to compare differentially expressed genes systematically.[44] mRNA obtained from two different cell populations or tissues was amplified by PCR after a reverse transcription (RT) step, and then the generated fragments, that reflected the expression pattern of the given cell populations or tissues, were separated by denaturing gel electrophoresis. Differentially expressed genes could be isolated from the gel, sequenced and identified. Numerous studies have been published, which - despite the serious disadvantages (maintenance the quantitative correlation after RT and PCR reactions, repeatability, and the elimination of false positive signals) - applied this method successfully.[51,71]

cDNA Representational Difference analysis (RDA) became widely known as an alternative method, and is a PCR based subtractive hybridization procedure. It was originally developed for the comparison of two complex genomes, and then it was later adapted for cloning differentially expressed genes.[46] mRNA of different origins ("tester" for treated or diseased sample and "driver" for control sample) is reverse transcribed, digested with a frequently cutting restriction endonuclease and then amplified according to the ligated linker sequence. The linker sequences of both driver and tester pools are removed and a new sequence is linked to the tester cDNA. After hybridization of both cDNAs and subsequent PCR, only those cDNAs present in the tester population are amplified. The advantage of this technique is the specific amplification of different fragments presented only in one DNA population. Both differential display and the RDA are suitable techniques for analyzing small number of transcripts.

Serial analysis of gene expression (SAGE) technique uses a sequence-based strategy that allows parallel analysis of a large number of transcripts.[70] The method is based on two principles: 1, a short, 8-9 base pair long nucleotide sequence tag contains enough information for the identification of the transcript, 2. concatenation of these short tags

allows the efficient analysis of transcripts in a serial manner by the sequencing of multiple tags within a single clone. Results obtained with this technique allow the determination of significant quantitative relationship between mRNA populations derived from various experimental procedures. In contrast to other methods mentioned before, it is much more sensitive in detecting low copy number transcripts. The aforementioned techniques are material-intensive and time consuming. For these reason, efforts had been undertaken to develop methods for high-throughput screening.

In recent years, a new technique, the *DNA microarray technology* (DNA-chip) has emerged offering the possibility of high-throughput systematic analysis of the "transcriptome". It is the most informative and most effective functional method of techniques mentioned so far, regarding the time and work necessity.[16,8,20,62] The arrays are constructed of thousands of DNA fragments either spotted or synthesized (Affymetrix) onto chemically activated glass slides. DNA fragments can be collections of short or long oligonucleotides or cDNAs of variable length. Oligonucleotide microarrays are suitable for the simultaneous detection of several thousands of single nucleotide mutations polymor, phismus (SNPs) in an amplified genomic DNA sample. This has a great significance in searching for polymorphic loci, or in detection of single or multiple mutations in medical samples. The most important and most informative application of DNA-chips is the parallel study of gene expression from different biological samples that focuses on the functionally active parts of the genome. DNA microarrays with sets of cDNA fragments on their surface can be used to obtain a molecular fingerprint of gene expression of cells.[9,41,45] The method has enabled large numbers of genes, from specific cell populations, to be studied in a single experiment. An important difference between the oligonucleotide and cDNA chips is the hybridization step. While oligonucleotide chips are hybridized by labeled RNA (the different sample RNAs hybridized onto different chip), in the case of cDNA chips, labeled cDNA samples are hybridized together onto one chip. This method is a reverse blotting technique, where mRNA populations gained from diverse biological sample (tissues or cell cultures) converted to cDNA in the presence of fluorescence dye (Cy3 or Cy5) labeled nucleotides. Using a co-hybridization strategy, with Cy3-labeled cDNA from the test sample and Cy5-labeled cDNA from the control sample, the relative intensity ratio on the microarrays can be determined and the expression pattern can be analyzed (*Figure 2.*). The quantity of the starting mRNA depends on the type of tissue (e.g. liver or brain tissues). The amount of RNA is strongly limited in those cases, where the amount of the sample tissue is small, for example in cases of biological sample obtained with laser microdissection or other operative methods, or

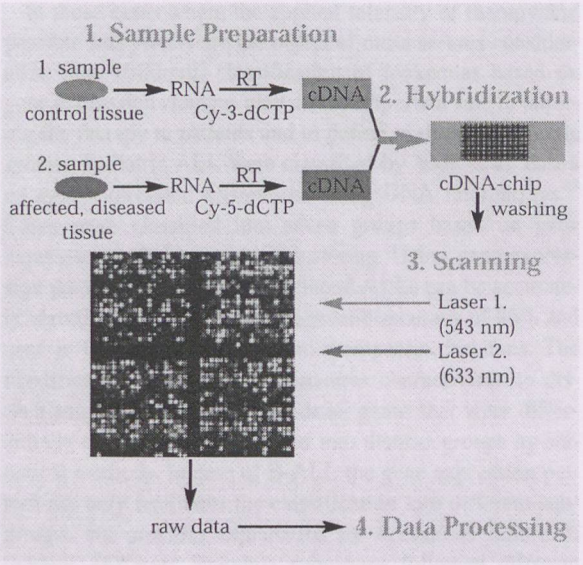


Figure 2. Application of cDNA arrays for the follow up detection of gene expression changes

in experimental systems where 1000–5000 cells are the object of the investigation. In these cases, amplification of the RNA sample or signal is necessary. During the amplification steps, however, it is very important to keep the quantitative ratios presented in the original mRNA population. The proper use of exponential (PCR) and linear (*in vitro* transcription, IVT) amplification can solve this problem.[53,54,40] By this technique, changes in cells generated by various effects (e.g. pharmaceutical treatment, pathological processes) can be traced, new biochemical markers and genes responsible for pathological phenotype can be discovered, drug effects can be followed and the treatment can be optimized. The differences in gene expression of the treated and untreated cells or tissues provide information about the regulation of the enzymatic pathways influenced by drugs, about the enzymes, transporters playing a role in drug resistance. Identification of gene expression patterns may provide vital information for the understanding of the pathological processes and may contribute to diagnostic decisions and therapies tailored to the individual patient.

The discovery of gene expression changes associated with the chromosome imbalances described in the previous section can have crucial diagnostic and therapeutic values. The molecular basis of T-cell acute lymphoid leukemias was largely discovered by the analysis of chromosome translocations and intrachromosomal rearrangements. These abnormalities typically reflect such arrangements where strictly regulated genes that important for cell-maturation (e.g. HOX11, LMO1, LMO2) become regulated by strong promoters or enhancers (e.g. T-cell receptor promoter). These rearrangements greatly increase the expression of these genes

and cause the abnormal maturation of precursor cells. Little is known about the downstream mechanisms that maintain the T-ALL phenotype. By using oligonucleotide chips, Ferrando et al. have shown that in case of T-ALL, five different T-cell oncogenes (HOX11, TAL1, LYL1, LMO1, LMO2) are frequently expressed abnormally in the absence of chromosome abnormalities. During the maturation of thymocytes, expression of these oncogenes is characteristic to specific developmental stages of the cells (e.g. LYL1+ [pro-T], HOX 11+ [early cortical thymocyte], TAL1+ [late cortical thymocyte]). They identified several gene expression signatures that were indicative of leukemic arrest at specific stages of thymocyte development. Samples could be grouped according to their shared oncogenic pathways by hierarchical clustering analysis.[23] These findings can help to form subgroups that are also important from the point of view of therapy.

Golub and colleagues applied successfully the DNA microarray technology to test whether the differencies in gene expression profile of ALL and AML can be used to distinguish between patients suffering from the two cancer types. They identified numerous genes that were associated specifically with either AML or ALL. Using statistical methods sets of genes were defined as "class predictors" helping the accurate classification. They also showed that this approach could be useful for characterization of new classes of tumors ("class discovery").[28]

In those cases where the applied intensity of therapy and possible side effects are the object of more serious consideration (e.g. children), classification of leukemias based on gene expression changes play a very important role in tailoring the therapy to patients and to define optimal therapeutic groups. Pediatric ALL were classified by Yeoh et al. based on gene expression differences using DNA microarrays.[69] Cases were classified into seven groups based on gene expression with hierarchical clustering. Using gene expression patterns subgroups of childhood ALLs can be accurately identified with an overall diagnostic accuracy of 96% and used in the class prediction and therapeutic decisions. The identification of molecular signatures characteristic to distinct subgroups was based on those genes that were differentially expressed and clustered into distinct groups by statistical methods. In case of B-ALL the gene expression pattern not only facilitates the classification into different subgroups, but provides explanation for occasional failure of therapy. Different leukemia subgroups followed different oncogene activation mechanisms, have different signal-transduction pathways, and therefore respond to therapies in different ways. Chemotherapy-induced secondary AML can also be identified by using distinct expression profiles (molecular signatures) despite the fact that secondary AML is thought to arise from a hematopoetic stem cells different from that giving rise the primary leukemia.[69]

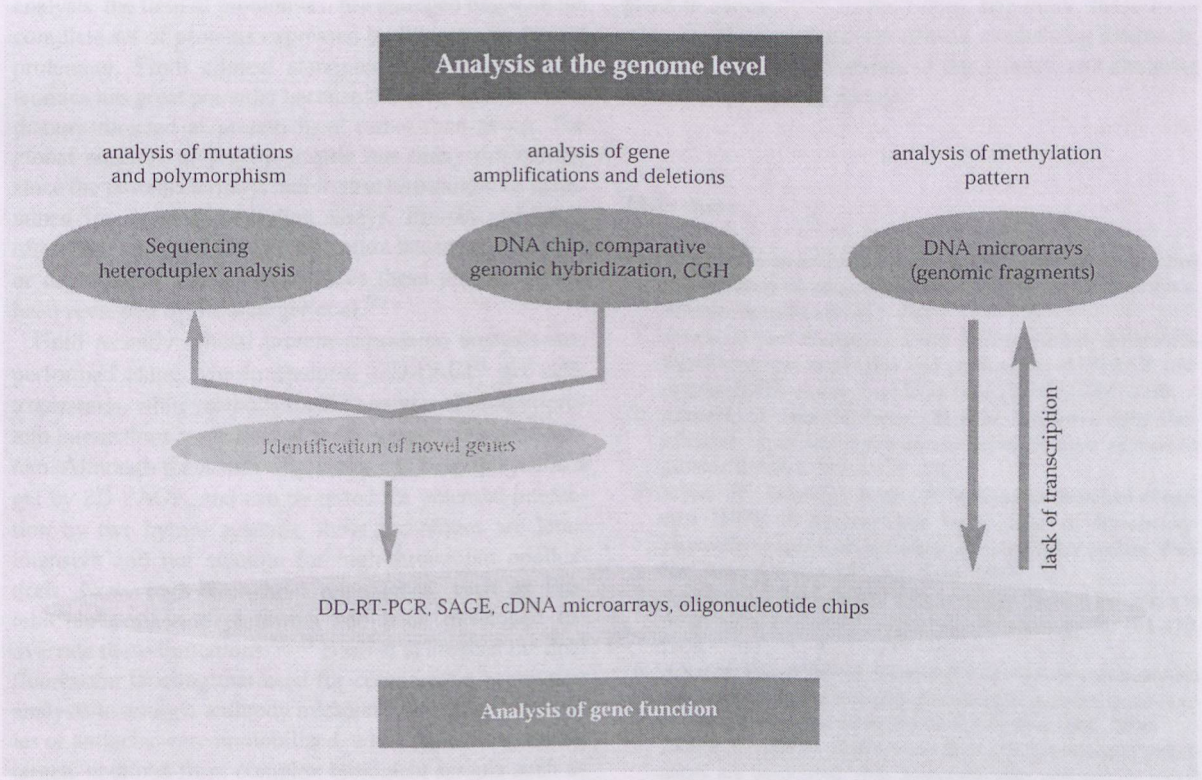


Figure 3. Analysis of genome and gene function with various techniques

Expression profiling with DNA-chip offers a viable and applicable alternative method in comparison to traditional methods (chromosome band analysis or *in situ* fluorescent hybridization) for diagnosis, treatment, and prediction of outcome of cancers especially in cases where the chromosome abnormalities cannot be detected with cytogenetic methods.[28,55,59] A general review by Kozian et al. is an excellent overview of different techniques used in differential gene expression analysis studies.[42]

Proteome level: Array-based analysis of protein expression pattern and protein function

Although transcript profiling offers good opportunity to identify genes that play a role in cancer formation, even complete mRNA fingerprints have their limitations. It is obvious that cellular functions are carried out by proteins, not by DNA or RNA. Numerous protein modifications, such as RNA splicing and posttranslational modification (e.g. phosphorylations, glycosylations), are known to influence protein function. The genomic or the transcript sequences do not give information about the different protein-protein interactions, or how and where these interaction occur inside the cells under various conditions. To obtain detailed information about a complex biological sample, information about many proteins, and protein-protein interaction is required. A new field of cell function analysis, the field of proteomics, has emerged based on the complete set of proteins expressed by the genome (called proteome). From clinical standpoint, the field of proteomics has great potential because the drug treatment and therapy directed at protein level rather than genes. The global analysis of protein profile has many difficulties, since the proteins tertiary, native structure should be maintained for successful binding assays. Protein regulation often based on reversible modification instead of synthesis or degradation. Strategies to solve these problems have been reviewed by Weinbarger et al.[72]

Until recently, global protein expression analysis was performed using two-dimensional (2D-PAGE) gel electrophoresis, while methods for the analysis of protein-protein interactions were limited to the yeast two-hybrid system. Although thousands of proteins can be displayed in a gel by 2D-PAGE, and can be tested for potential interaction by two-hybrid systems, these techniques are labor intensive and not suitable for high-throughput applications. New, high-throughput approaches, such as ProteinChip® proteomic platforms had been developed to override these limitations.[2,11,19] Haab et al adapted the dual fluorescent labeling that used for comparative expression analysis to antigen-antibody microarrays.[29] Either antibodies or antigens were immobilized, while the corresponding targets obtained from complex biological sample such as serum were fluorescently labeled. The two samples were mixed and incubated simultaneously on the same microarray. Dual color detection system revealed the quantitative differences between the two samples. Since monoclonal antibody production is labor intensive, development of alternative methods, such as high-throughput generation of recombinant proteins have become crucial. The recently developed microarray-based protein-protein interaction (protein-protein, small synthetic molecules-protein, enzyme-substrate, receptor-ligand) assays give the possibilities for high-throughput protein function assays. Some of the current technologies for proteome profiling and the application of proteomics to the analysis of leukemias have been reviewed by Hannash et al. [30]

The technical developments of the last 15 years have revolutionized the molecular understanding, the discovery of functional genomic backgrounds, and the identification of diagnostic markers of leukemias and other cancers. A part of these techniques strive to detect genome wide abnormalities (mutations, gene amplifications-deletions, methylation pattern), while another part aims to analyze the products of active genes, the mRNA populations (transcriptome) (gene expression analyses, SAGE, DD-RT-PCR, DNA chip technology) and protein composition and function (proteome) of cells (*Figure 3*). The number of studies analyzing the functional genomics of cancers are growing exponentially, for example melanoma,[8,1,75] colon cancer,[45,17,63,62] thyroid tumor,[66,13,36] breast cancer,[37,14,43] prostate cancer.[60,58,47] In the future, hopefully, these techniques will be available for clinical use helping diagnosis, prognostication the outcome of the disease, and choosing the most appropriate therapy.

References

1. *Adorján P, Distler J, Lipscher E, et al:* Tumour class prediction and discovery by microarray-based DNA methylation analysis. Nucleic Acids Res 30:e21, 2002.
2. *Albala JS and Humphrey-Smith I:* Array-based proteomics: high-throughput expression and purification of IMAGE consortium cDNA clones. Curr Opin Mol Ther 680–684, 1999.
3. *Alizadeh AA, Ross DT, Perou CM, et al:* Towards a novel classification of human malignancies based on gene expression patterns. J Pathol 195: 41-52, 2001.
4. *Alwine JC, Kemp DJ, Stark GR:* Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc Natl Acad Sci USA 74: 5350-5354, 1977.
5. *Appelbaum FR:* Molecular Diagnosis and Clinical Decisions in Adult Acute Leukemia, Seminars in Hematology 36: 401-410, 1999.
6. *Arico M, Valsecchi MG, Camitta B et al:* Outcome of treatment in children with Philadelphia chromosome-positive acute lymphoblastic leukemia. N Engl J Med 342: 998-1006, 2000.
7. *Aubele M, Auer G, Braselmann H et al:* Chromosomal imbalances are associated with metastasis-free survival in breast cancer patients. Anal Cell Pathol 24: 77-87, 2002.

8. *Bittner M, Meltzer P, Chen Y et al:* Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature 406: 536-540, 2000.
9. *Blohm DH, Guiseppi-Elie A:* New developments in microarray technology. Curr Opin Biotechnol 12: 41-47, 2001.
10. *Bruckert P, Kappler R, Scherthan H et al:* Double minutes and c-MYC amplification in acute myelogenous leukemia: Are they prognostic factors? Cancer Genet Cytogenet 120: 73-79, 2000.
11. *Cahill DJ:* Protein arrays: a high-throughput solution for proteomics research? Proteomics: A Trends Guide, 47–51, 2000.
12. *Chan MF, Liang G and Jones PA:* Relationship between transcription and DNA methylation. Curr Top Microbiol Immunol 249: 75–86, 2000.
13. *Chen KT, Lin JD, Chao TC et al:* Identifying differentially expressed genes associated with metastasis of follicular thyroid cancer by cDNA expression array. Thyroid 11: 41-46, 2001.
14. *Clark J, Edwards S, John M et al:* Identification of amplified and expressed genes in breast cancer by comparative hybridization onto microarrays of randomly selected cDNA clones. Genes Chromosomes Cancer 34: 104-114, 2002.
15. *Costello JF, Fruhwald MC, Smiraglia DJ et al:* Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nature Genet 24: 132–138, 2000.
16. *DeRisi J, Penland L, Brown P et al:* Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nature Genet 14: 457-460, 1996.
17. *Dunican DS, McWilliam P, Tighe O et al:* Gene expression differences between the microsatellite instability (MIN) and chromosomal instability (CIN) phenotypes in colorectal cancer revealed by high-density cDNA array hybridization. Oncogene 21: 3253-3257, 2002.
18. *Eads CA, Danenberg KD, Kawakami K et al:* MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res 28-c32, 2000.
19. *Emili AQ and Cagney G:* Large-scale functional analysis using peptide or protein arrays. Nat. Biotechnol 18: 393–397, 2000.
20. *Ernst T, Hergenhahn M, Kenzelmann M et al:* Decrease and gain of gene expression are equally discriminatory markers for prostate carcinoma: a gene expression analysis on total and microdissected prostate tissue. Am J Pathol 160: 2169-2180, 2002.
21. *Esteller M, Corn PG, Baylin SB et al:* A gene hypermethylation profile of human cancer. Cancer Res 61: 3225–3229, 2001.
22. *Ferrando AA, Look AT:* Clinical implications of recurring chromosomal and associated molecular abnormalities in acute lymphoblastic leukemia. Semin Hematol 37: 381-95, 2000.
23. *Ferrando AA, Neuberg DS, Staunton J, et al:* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell, 1: 75-87, 2002.
24. *Furák JI, Troján T, Szöke L et al:* Development of Brain Metastasis 5 Years Before the Appearance of the Primary Lung Cancer: Messenger Metachronous Metastasis. Annals Thoracic Surg (in press).
25. *Gitan RS, Shi H, Chen CM et al:* Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. Genome Res 12: 158-164, 2001.
26. *Gilliland DG:* Molecular Genetics of Human Leukemia. Leukemia 12: S7-S12, 1998.
27. *Golub TR:* The Genetics of AML: An Update in Proceedings of the American Society of Hematology, pp. 102-111, 1999.
28. *Golub TR, Slonim DK, Tamayo P et al.:* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286: 531-537, 1999.
29. *Haab BB Dunham MJ, Brown PO:* Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. Genome Biol 2: Research 0004.1–0004.13, 2001.
30. *Hanash SM, Madoz-Gurpide J, Misek DE:* Identification of novel targets for cancer therapy using expression proteomics. Leukemia 16: 478-485, 2002.
31. *Hedenfalk I, Duggan D, Chen Y et al:* Gene-expression profiles in hereditary breast cancer. N Engl J Med 344: 539-548, 2001.
32. *Hedrick SM, Cohen DI, Nielsen EA et al:* Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature 308: 149-153, 1984.
33. *Hodgson G, Hager JH, Volik S et al:* Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. Nat Genet 29: 459-464, 2001.
34. *Houldsworth J, Chaganti RS:* Comparative genomic hybridization: an overview. Am J Pathol 145: 1253-1260, 1994.
35. *Huang TH-M, Perry MR and Laux DE:* Methylation profiling of CpG islands in human breast cancer cells. Hum Mol Genet 8: 459–470, 1999.
36. *Huang Y, Prasad M, Lemon WJ et al:* Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. Proc Natl Acad Sci USA 98: 15044-15049, 2001.
37. *Jiang Y, Harlocker SL, Molesh DA et al:* Discovery of differentially expressed genes in human breast cancer using subtracted cDNA libraries and cDNA microarrays. Oncogene 21: 2270-2282, 2002.
38. *Jones PA:* DNA methylation errors and cancer. Cancer Res 65: 2463–2467, 1996.
39. *Kanerva J, Niini T, Vettenranta K et al:* Loss at 12p detected by comparative genomic hybridization (CGH): association with TEL-AML1 fusion and favorable prognostic features in childhood acute lymphoblastic leukemia (ALL). A multi-institutional study. Med Pediatr Oncol 37: 419-425, 2001.
40. *Kitajka K, Puskás LG, Zvara Á et al:* The role of n-3 polyunsaturated fatty acids in brain: Modulation of rat brain gene expression by dietary n-3 fatty acids. Proc Natl Acad Sci USA 99, 2619-2624, 2002.
41. *Kopper L, Tímár J:* Gene expression profiles in the diagnosis and prognosis of cancer. Magy Onkol 46:3-9, 2002.
42. *Kozian DH, Kirschbaum BJ:* Comparative gene-expression analysis. Trends Biotechnol 17: 73-78, 1999.
43. *Kroll T, Odyvanova L, Clement JH et al:* Molecular characterization of breast cancer cell lines by expression profiling. J Cancer Res Clin Oncol 128: 125-134, 2002.
44. *Liang P, Pardee A et al:* Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257: 967-971, 1992.
45. *Lin YM, Furukawa Y, Tsunoda T et al:* Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 21: 4120-4128, 2002.
46. *Lisitsyn N, Wigler M et al:* Cloning the differences between two complex genomes. Science 259: 946-951, 1993.
47. *Luo JH, Yu YP, Cieply K et al:* Gene expression analysis of prostate cancers. Mol Carcinog 33: 25-35, 2002.
48. *Mitelman F, Mertens F, Johansson B:* A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. Nat Genet Spec No: 417-474, 1997.
49. *Novak U, Oppliger Leibundgut E, Hager J et a:* A high-resolution allelotype of B-cell chronic lymphocytic leukemia (B-CLL). Blood 100: 1787-1794, 2002.
50. *Pinkel D, Segraves R, Sudar D et al:* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 20: 207-11, 1998.

51. *Prashar Y, Weissman S et al:* Analysis of differential gene expression by display of 3' end restriction fragments of cDNAs. Proc Nat Acad Sci USA 93: 659-663, 1996.
52. *Pui CH, Campana D, Evans WE:* Childhood acute lymphoblastic leukaemia current status and future perspectives. Lancet Oncol 10: 597-607, 2001.
53. *Puskás LG, Zvara Á, Hackler JrL et al:* RNA amplification results in reproducible microarray data with slight ratio biases. Biotechniques 32: 1330-1342, 2002.
54. *Puskás LG, Zvara Á, Hackler JrL et al:* Production of bulk amounts of universal RNA for DNA-microarrays. Biotechniques 33: 898-900, 902, 904, 2002.
55. *Ramaswamy S, Tamayo P, Rifkin R et al:* Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci USA 98: 15149-15154, 2001.
56. *Rubnitz JE and Pui C-H:* Leukemias, in *Principles of Molecular Medicine*, pp. 233-239. (J.L. Jameson ed.) Humana Press, NJ, 1998.
57. *Schlossman SF, Chess L, Humphreys RE et al:* Distribution of Ia-like molecules on the surface of normal and leukemic human cells. Proc Natl Acad Sci USA 73: 1288-1292, 1976.
58. *Schwarze SR, DePrimo SE, Grabert LM et al:* Novel pathways associated with bypassing cellular senescence in human prostate epithelial cells. J Biol Chem 277: 14877-14883, 2002.
59. *Shipp MA, Ross KN, Tamayo P et al:* Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med. 8: 68-74, 2002.
60. *Shou J, Soriano R, Hayward SW et al:* Expression profiling of a human cell line model of prostatic cancer reveals a direct involvement of interferon signaling in prostate tumor progression. Proc Natl Acad Sci USA 99: 2830-2835, 2002.
61. *Sidransky D:* Emerging molecular markers of cancer. Nature Rev Cancer 2: 210-219, 2002.
62. *Stremmel C, Wein A, Hohenberger W et al:* DNA microarrays: a new diagnostic tool and its implications in colorectal cancer. Int J Colorectal Dis 17: 131-136, 2002.
63. *Suzuki H, Gabrielson E, Chen W et al:* A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. Nat Genet 31: 141-149, 2002.
64. *Swaroop A, Xu JZ, Agarwal N et al.:* A simple and efficient cDNA library subtraction procedure: isolation of human retina-specific cDNA clones. Nucleic Acids Res 19: 1954, 1991.
65. *Sweetser DA, Chen CS, Blomberg AA et al:* Loss of heterozygosity in childhood de novo acute myelogenous leukemia. Blood 98: 1188-94, 2001.
66. *Takano T, Hasegawa Y, Matsuzuka F et al:* Gene expression profiles in thyroid carcinomas Br J Cancer 83: 1495-502, 2000.
67. *Toyota M, Ho C, Ahuja N et al:* Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. Cancer Res 59: 2307–2312, 1999.
68. *Tsukasaki K, Krebs J, Nagai K et al:* Comparative genomic hybridization analysis in adult T-cell leukemia/lymphoma: correlation with clinical course. Blood 97: 3875-3881, 2001.
69. *Yeoh EJ, Ross ME, Shurtleff SAA et al:* Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell 1: 133-143, 2002.
70. *Velculescu VE, Zhang L, Vogelstein B et al:* Serial analysis of gene expression. Science 270: 464-467, 1995.
71. *Vogeli-Lange R, Burckert N, Boiler T et al:* Rapid selectionand classification of positive clones generated by mRNA differential display. Nucleic Acids Res 24: 1385-1386, 1996.
72. *Weinberger SR, Morris TS, Pawlak M:* Recent trends in protein biochip technology. Pharmacogenomics 1: 395-416, 2000.
73. *Wieser R:* Rearrangements of chromosome band 3q21 in myeloid leukemia. Leuk Lymphoma 43: 59-65, 2002.
74. *Wilhelm M, Veltman JA, Olshen AB et al:* Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer. Cancer Res 62: 957-960, 2002.
75. *de Wit NJ, Burtscher HJ, Weidle UH et al:* Differentially expressed genes identified in human melanoma cell lines with different metastatic behaviour using high density oligonucleotide arrays. Melanoma Res 12: 57-69, 2002.
76. *Wreesmann VB, Ghossein RA, Patel SG et al:* Genome-wide appraisal of thyroid cancer progression. Am J Pathol 161: 1549-1556, 2002.