RNA expression profiling at the single molecule level

Jan Hesse,1 Jaroslaw Jacak,1 Maria Kasper,2 Gerhard Regl,2 Thomas Eichberger,2 Martina Winklmayr,2 Fritz Aberger,2 Max Sonnleitner,3 Robert Schlapak,3 Stefan Howorka,3 Leila Muresan,4 Anna-Maria Frischauf,2,5 and Gerhard J. Schütz1,5

1Biophysics Institute, Johannes Kepler University Linz, A-4040 Linz, Austria; 2Division of Genomics, Department of Molecular Biology, University of Salzburg, A-5020 Salzburg, Austria; 3Center for Biomedical Nanotechnology, Upper Austrian Research GmbH, A-4020 Linz, Austria; 4Department of Knowledge-based Mathematical Systems, Johannes Kepler University Linz, A-4040 Linz, Austria

We developed a microarray platform for PCR amplification-independent expression profiling of minute samples. A novel scanning system combined with specialized biochips enables detection down to individual fluorescent oligonucleotide molecules specifically hybridized to their complementary sequence over the entire biochip surface of cm² size. A detection limit of 1.3 fM target oligonucleotide concentration—corresponding to only 39,000 molecules in the sample solution—and a dynamic range of 4.7 orders of magnitude have been achieved. The applicability of the system to PCR amplification-independent gene-expression profiling of minute samples was demonstrated by complex hybridization of cDNA derived from the equivalent of only 10⁴ cells, which matches results obtained in ensemble studies on large samples. By counting each hybridized molecule on the microarray, the method is insensitive to gene-specific variations of the labeling, thereby representing a principle advance to conventional ensemble-based microarray analysis.

[Supplemental material is available online at www.genome.org.]

Microarray analysis of the global transcriptional state of a small cell number represents a major target for both medical diagnostics and basic research. In particular, precise characterization of minute patient’s samples such as malignant and multidrug-resistant cancer stem cells is decisive for a detailed molecular understanding of the oncogenic mechanisms driving tumor growth as well as for diagnosis and individualized therapy; furthermore, global molecular characterization of small subregions within heterogeneous tissues or of tiny biological units such as stem-cell niches is of fundamental importance for a more detailed understanding of key biological processes such as tissue homeostasis in normal and diseased states.

Recent technological advances allow for routine isolation of small, yet defined subpopulations of cells such as cancer stem cells from heterogeneous fluid or tissue samples using Fluorescence Activated Cell Sorting (FACS) and Laser Capture Microdissection (LCM). Subsequent purification and biochemical processing can be performed with minimum loss of material via microfluidic devices (Paegel et al. 2003; Hong et al. 2004). Still, there is an urgent need for a robust technology to determine the global expression profiles of such minute samples. Current protocols include an amplification step for RNA/DNA, with linear amplification being the preferred method over PCR-based approaches. While linear amplification is less prone to distortions of the relative abundance of mRNA species in the original sample, the method is time consuming and requires sophisticated protocols that are difficult to establish (Van Gelder et al. 1990; Kacharmina et al. 1999; Mahadevappa and Warrington 1999; Wang et al. 2000; Baugh et al. 2001; Hu et al. 2002; Taylor et al. 2004). These problems could be avoided by direct hybridization of unamplified labeled DNA and readout with a highly sensitive detection system.

In conventional DNA microarray readout, the sensitivity is limited by standard formats of biochip substrates. Their thickness of ~1 mm requires the implementation of imaging optics with a long working distance, at the expense of detection efficiency. Moreover, impurities within the substrate material typically generate a strong fluorescence background, which impedes ultrasensitive fluorescence detection on such biochips.

Setting up a robust scanning technology for cm² areas with single-molecule sensitivity is a challenging task. To reliably identify single dye molecules, imaging at the resolution limit is inevitable; the concomitant short depth of focus of ~400 nm demands precise adjustment of the focal plane. A high-precision scanning stage operating at the resolution limit, a robust focus-hold system, and a thin biochip with a surface of high purity over large areas must therefore be regarded as major requirements for DNA microarray analysis at the single molecule level.

We report here the development of a system for PCR amplification-independent DNA microarray analysis down to the single-molecule level, which includes an ultrasensitive detection device and accessory biochips. Hybridization of fluorescent oligonucleotides was used to characterize the platform in terms of sensitivity and dynamic range. Complex hybridization of biological material corresponding to only 10⁴ cells demonstrates the applicability of the system for biomedical diagnostics.

Results and Discussion

Biochip surfaces of arbitrary size were imaged at the sensitivity of single-dye molecules using a home-built scanning system (Hesse et al. 2004; Sonnleitner et al. 2005). All images were recorded with a pixel size of 200 nm, which yields best performance for single-molecule identification at the given optical resolution.
limit of 450 nm (full width at half maximum of the point spread function); 40 min scanning time was required to record an area of 1 cm² on the chip surface. To enable imaging at high-detection efficiency, DNA microarrays were established on the basis of 150-µm thick aldehyde-functionalized glass coverslips, which were selected for low autofluorescence (Schlapak et al. 2005).

We first tested the binding specificity by hybridizing a Cy5-labeled 60mer target oligonucleotide (c = 100 fM) to its complementary sequence (probe oligonucleotide) bound to the biochip surface (Fig. 1). Homogenously distributed diffraction-limited peaks were observed over the entire image; the corresponding surface density was 1000-fold higher when compared with a surface functionalized with control oligonucleotide. To ascertain whether each fluorescence peak corresponds to a single Cy5-oligonucleotide molecule, we determined the brightness of individual peaks as a measure of the number of colocalized dye molecules; fitting with a two-dimensional Gaussian profile (Schmidt et al. 1996) yielded a mean value of B = 2600 counts, which is in good agreement with the single-molecule brightness of Cy5 determined in a supported lipid bilayer doped with low amounts of Cy5-labeled lipids (2400 counts). The high single-molecule brightness and the low background noise within a diffraction limited region (σ = 34 counts) obtained on the biochip surface allow for reliable (signal-to-background noise ratio >10) detection of more than 99% of all fluorescent oligonucleotide molecules.

To determine the detection limit and the dynamic range of the platform, we varied the target concentration c from 100 aM to 1 nM. Hybridization experiments were performed on microarrays with a regular checkerboard pattern of probe and control oligonucleotides using a spot size of 100 × 100 µm². Figure 2A shows a microarray hybridized with 100 fM Cy5-labeled target oligonucleotide. Binding efficiency was homogenous over the chip surface, characterized by a spot-to-spot signal variation of only 13%. Single oligonucleotide molecules bound to the chip surface become visible when the data are plotted at higher magnification (Fig. 2B). At low surface densities of N ≤ 1000 molecules per spot, more than 95% of all molecules are well separated in the image and were counted via automated algorithms.

At higher surface densities, peaks begin to overlap; in such cases, the number of hybridized molecules per spot was inferred from the corresponding integrated fluorescence signal and the single molecule brightness B. Residual signals due to unspecific adsorption (Ñ) on control spots were subtracted for calculating the specific signal Ñ. A constant concentration of Cy3-labeled target oligonucleotide (c = 10 fM) was included in all experiments; the fluorescence signal S_Cy3(c) is a measure of the binding capacity of each spot and was used to correct for chip-to-chip variations with (S_Cy3(c)/S_Cy3(all conc)) as linear normalization. Figure 2C shows the binding curve for specific hybridization (red) and unspecific adsorption (black). A linear relation between specifically hybridized oligonucleotides and sample concentration was found over six orders of magnitude. At all applied concentrations, the average distance between two neighboring molecules is >30 nm; therefore, fluorescence quenching has no significant effect on the molecular brightness.

A constant unspecific signal of Ñ = 13.2 molecules per spot was observed for c < 1 pM, due to residual surface and buffer impurities. At sample concentrations c ≥ 1 pM unspecific oligonucleotide binding resulted in a linear increase of the unspecific signal, approximately four orders of magnitude below the specific signal.

For the definition of the detection limit, two major noise contributions have to be considered; first, the signal noise, which includes variations in the brightness of the fluorophore and of the background; and second, variations due to the stochastic nature of the binding process. While for conventional microarray analysis the first term is dominant and limits the accuracy of signal quantification, its influence on single-molecule counting is negligible due to the high-detection reliability. Here, the ultimate statistical limit of the binding process defines the sensitivity of the system. Assuming Poissonian-binding statistics, the signal-to-noise ratio (SNR) is described by SNR = Ñ/√N. We find that Ñ = 112 molecules specifically hybridized to one spot are sufficient for data analysis with a SNR of 10. For the given experimental conditions, this number corresponds to a sample concentration of 1.3 fM (~39,000 target molecules in the sample volume of 50 µL), defining the limit of detection of the system.

The high instrumental dynamic range of the developed system enables signal quantification in a range of from 1.3 fM to 1 nM. The operational dynamic range, however, is currently limited by unspecific adsorption of fluorescent oligonucleotides to the target control spots. At c = 67 pM, the unspecific signal equals the specific signal recorded at the detection limit of c = 1.3 fM, yielding an operational dynamic range of 4.7 orders of magnitude.

To demonstrate the applicability of the system to gene-expression profiling, we used arrays of 125 distinct 67-70mer oligonucleotides spotted in eight replicates (1000 probe features per array). Complex cDNA was synthesized from 200 ng of total RNA from the equivalent of only 10⁵ cells of a human keratinoocyte cell line (HaCaT) and used for hybridization (Fig. 3A). Samples were scanned at a speed of 2.4 sec/spot; in total, the analysis took ~6 sec/spot. Spots indicated by arrows (1, 2, and 3) are shown below at high magnification. Well-separated peaks were detected, e.g., for BCL2 (1) and S100A8 (3), and counted for determination of N, yielding 450 ± 100 and 890 ± 31-bound molecules per spot, respectively. Interestingly, the two cDNA species show differences in the single-molecule brightness. This result points out a potential bias in conventional microarray analysis; variations in the brightness of the target molecules affect the spot signal in ensemble measurements, hampering a precise de-

---

**Figure 1.** Single-fluorescent oligonucleotide molecules hybridized to functionalized biochips. (A) Surface functionalized with probe oligonucleotide upon hybridization with 100 fM of Cy5-labeled target oligonucleotide. Individual diffraction-limited peaks, corresponding to single oligonucleotide molecules, can be resolved at a surface density of 0.8 µm⁻². (B) Surface functionalized with control sequence yielded a 1000-fold lower signal compared with specific hybridization. We ensured that the control can bind to a complementary strand by hybridizing with 60-pM complementary fluorescent (Cy5-labeled) Xenopus XAG2-cDNA (Supplemental Fig. S1). (C) The high single-molecule brightness of 2600 counts and the low background noise of 34 counts allow reliable detection of more than 99% of all surface-bound dye molecules.
expression profiling at the single molecule level

Figure 2. Validation of the system. (A) 60mer probe (red squares) and target (yellow squares) oligonucleotides were spotted. Specific signals after hybridization with 100 fM Cy5-labeled target oligonucleotide show a spot-to-spot variation of only 13%. For illustration, data were software-binned (10 × 10). (B) Fluorescence images of spots hybridized with different concentrations of Cy5-labeled target-oligonucleotide. Individual Cy5-labeled oligonucleotide molecules can be detected in the range of from 100 aM to 10 fM; at higher concentrations, single molecule signals begin to overlap. (Insets) Details within the spots at 3.5-fold higher magnification. In the second row, the corresponding images of control spots are shown. A constant concentration of Cy3-labeled target-oligonucleotide was included in all experiments (third row). (see Supplemental Fig. S2 for images of higher target concentrations). (C) Number of specifically hybridized (red) and unspecifically adsorbed (black) oligonucleotide molecules per spot vs. sample concentration. Three independent experiments, each containing 16 replicate spots, were performed for each concentration. The surface density increases linearly with c over six orders of magnitude (surface-density ∝ $c^{0.91}$). Assuming a minimum signal-to-noise ratio of 10 as criterion for reliable analysis (dashed line), a detection-limit of $c = 1.3$ fM can be specified. Error bars indicate the standard deviations of the signals. Note that due to the logarithmic scaling, the error bars are only visible at low concentrations (see also Supplemental Table S1).

termination of relative mRNA abundances. In this example, ensemble analysis would even yield a reversed expression ratio of the two genes.

For validation, we compared the expression profiles obtained with the platform described above to those of a conventional microarray platform. A 100-fold higher amount of cDNA from the same RNA preparation was hybridized to commercially available aldehyde substrates (Schott Nexcelon) under identical conditions and scanned with a conventional confocal microarray scanner (ScanArray 4000, Perkin Elmer). The signals derived from both measurements are in good agreement, yielding a Pearson correlation coefficient of 0.85. The high consistency of the data obtained here using different microarray platforms and different sample amounts appears remarkable, as typically a considerable divergence of gene-expression results has been observed when comparing commercial products (Tan et al. 2003; Irizarry et al. 2005). Tan et al. (2003) find hardly any correlation when comparing hybridization on different platforms (coefficients ranging from 0.5 to 0.6). However, differences in the capture sequences of the commercial microarrays may dominate the measured signal differences. Irizarry et al. (2005) have therefore used a refined comparison method, which analyzes whether different laboratories can identify differential gene expression on the same sample. As identical platforms were used for comparison, a higher correlation coefficient has been obtained (∼0.91–0.98). The intermediate result revealed in our study reflects the fact that different platforms containing the same oligoset have been investigated.

The system presented enables expression profiling of minute amounts of sample material. Material from the equivalent of only $10^8$ cells can be reliably analyzed without prior PCR amplification steps. Connection to sample isolation and micro-fluidic purification techniques offers the exciting perspective to directly analyze minute cell populations in biomedical applications. In addition to sensitivity, 4.7 orders of magnitude dynamic range significantly exceeds specifications of conventional microarray platforms and approaches the complete biological range of mRNA expression levels of up to six orders of magnitude (Holland 2002).

Methods

Single molecule biochip reader

The detection system was set up on an epifluorescence microscope (Axiovert 200, Zeiss) (Hesse et al. 2004). Ar+ and Kr+ ion lasers (Inova, Coherent) were used for selective fluorescence excitation of Cy3 at 514 nm and Cy5/Alexa647 at 647 nm, respectively. Samples were illuminated in objective-type total internal reflection (TIR) configuration using a 100× oil immersion objective (NA = 1.45, oil immersion, Zeiss). After appropriate filtering using standard Cy3 and Cy5 filter sets (Chroma Technology Corp.), fluorescence was imaged onto a back-illuminated CCD camera (SPELIC10:100B, Princeton Instruments; quantum efficiency = 90%, gain = 0.77counts/e−). Biochip readout was performed in time-delay and integration (TDI) mode. For this, samples were shifted using a motorized xy-stage (Scan IM 120 × 100, Märzhäuser) synchronized to the line-shift of the camera. During the TDI-scan, data were streamed into the RAM of the PC and stored on the hard disk in the course of repositioning for the next scan. The reader was equipped with an automated focus hold system operating during the scanning process; the back-reflected laser beam is imaged on a two-segment photodiode, and the differential signal is used to control a piezo (PIFOC, Physik Instrumente) for fast refocusing.

Genome Research 3 www.genome.org
Microarray preparation

Aldehyde-derivatized glass cover slips were prepared as described (Schlapak et al. 2005). For evaluation of sensitivity and dynamic range, probe oligonucleotide ([NH2-C12]ATAGAAATGCA GCGTGAGATCTTAGAGATGGAACCGAAGCTGTACG- GCCTGTGCCACCC) (VBC-Genomics) and control oligonucleotides with a sequence identical to the target ([NH2-C12]GGTG GCACAAGCCTACGTCCTGTTTCAATCTGTAAGAGAT CCACCGCTGCAATTCTAT) were spotted with a MicroGridII arrayer (Genomic Solutions) using Stealth SMP3 Microarray Spotting Pins (TeleChem International, Inc.) onto 150-µm aldehyde coated coverslips (Schlapak et al. 2005).

Target oligonucleotides labeled with Cy3 or Cy5 at the 5' end (Cy3/Cy5-GGTGGCACAAAGCCTACAGCTCGTTC ATCTCTGCAAGATCCAGGCATTCTAT) (VBC-Genomics) and control oligonucleotides were used for evaluation of sensitivity and dynamic range of the developed platform.

For RNA profiling, 96 oligonucleotides of the Human Genome Oligo Set V3 (Operon) and 29 oligonucleotides designed using OligoWiz 1.0 (http://www.cbs.dtu.dk/services/OligoWiz/) were printed. Oligonucleotides were dissolved in 3XSSC/1.5 M betaine to a final concentration of 20 µM.

After printing, the slides were stored under dry conditions at 4°C for 4 h at room temperature. Stable covalent coupling to the chip surface and removal of unbound probe oligonucleotides was achieved by extensive washing (1 h) with blocking buffer (0.1 M Na-carbonate, 0.1 M 4-aminobutanoic acid, 10 µM NaBH3CN at pH 8.0), followed by a washing step with 1XSSC buffer. Directly after washing, slides were used for hybridization experiments.

RNA preparation

Total RNA from HaCaT cells was extracted with TRI reagent (Molecular Research Center) according to the manufacturer’s instructions (Molecular Probes, Invitrogen). In brief, 20 µg of total RNA was reverse transcribed with Superscript II (Invitrogen Life Technologies) in the presence of Alexa Fluor 647-aha-dUTP (Molecular Probes, Invitrogen). Labeled cDNA was purified with CyScribeTM GFXTM purification columns (Amersham Biosciences) and vacuum dried using a Centrivap Concentrator (Labconco).

Hybridization

Hybridization of 60mer Cy3 and Cy5 oligonucleotides was carried out for 60 min at 24°C in 4XSSC/0.1% SDS. Identical subarrays on the same biochip were separated by an adhesive siliccone mask (Secure Seal, Schleicher & Schuell), creating six hybridization chambers with 50 µL of sample volume and incubated with target oligonucleotides at different concentrations. cDNA was dissolved in 3XSSC/0.1% SDS, heated to 96°C for 3 min, and hybridized under LiftetSlip (Erie Scientific Company) at 50°C for 16 h in 3XSSC/0.1% SDS using a semiautomated hybridization station (SlideBooster, Advalytix AG; mixing to pause ratio: 3:7, mixing power: 27). In all experiments, hy-
bridization was followed by sequential washes in 1× SSC/0.1% SDS; 0.1× SSC/0.1% SDS, and 0.1× SSC. Directly after washing, the microarray was covered with 0.1× SSC buffer and LifterSlip, and scanned.

Conventional microarrays were hybridized with labeled cDNA synthesized from 20 μg of total RNA. Following precipitation with EtOH, cDNA was dissolved in 3× SSC/0.1% SDS, heated to 96°C for 3 min, and hybridized overnight at 50°C in a semi-automated hybridization station (SlideBooster, Advalytix AG) at a mixing:pause ratio of 3:7, mixing power: 27. Slides were washed at 30°C in a semiautomated wash station (Advalytix AG) two times for 10 min in 2× SSC/0.1% SDS, two times for 10 min in 0.2× SSC/0.1% SDS, and two times for 5 min in 0.2× SSC. Slides were dried by spinning for 3 min at 250g and subsequently scanned on a ScanArray 4000 confocal microarray scanner (PerkinElmer).

Data analysis

To increase analysis speed, the grid-fitting step was performed on a software-binned (10×10) image. After application of a global threshold by the Otsu method and removal of features smaller than the expected spot size, spot diameter and separation were computed from the remaining features. Based on these values, a grid was constructed and aligned to the brightest spots. For subsequent analysis, subimages (1000×1000 pixel) were loaded at full resolution.

Diffraction-limited peaks are small, approximately symmetric features over the background noise. For counting the total number of peaks per spot, \( N \), we used the à trous wavelet filter bank method as described previously (Starc et al. 1998; Olivo-Marin 2002). We apply the Anscombe transform

\[
A(x, y) = 2 \sqrt{\frac{I(x, y) + \frac{3}{8}}{2}}
\]

to convert the Poisson noise present in the images \( I \) to Gaussian noise as required by the à trous wavelet filter bank method. The resulting image \( A \) is decomposed via a sequence of low-pass and band-pass filters. For low-pass filtering, we convolve the image \( A \) with B-spline kernels of order \( J \), \( b_j \), where \( b_j = \left[ \frac{1}{1 + 3 \cdot |x|} \right] \). Low-pass filtering is iterated with kernels \( b_j \) where \( b_{j+1} \) is obtained by inserting lines and columns of zeros between the lines and columns of \( b_j \), yielding

\[
A_{j+1} = A_j \cdot b_j.
\]

The high-pass filter is obtained as the difference of two consecutive low-pass filtered images

\[
W_j = A_{j-1} - A_j.
\]

Each wavelet plane \( W_j \) contains information on feature of scale \( j \) and Gaussian noise \( \sigma_j \). We estimate the variance of the noise via robust statistics methods \( \sigma_j^2 = \text{median}[W(x, y) - m]^2 \), where \( m \) is the median of \( W(x, y) \). \( \sigma_j \) allows separation of the significant wavelet coefficients from those due to noise:

\[
W_j(x, y) = \begin{cases} 
W(x, y) & \text{if} \ W(x, y) \geq 3 \cdot \sigma_j \\
0 & \text{otherwise}
\end{cases}
\]

Due to the redundancy of the wavelet planes \( W_j \), salient features tend to produce significant coefficients across scales. A good way to detect salient features is by computing the multiscale product of the denoised wavelet planes, where non-zero values correspond to the position of the peaks:

\[
P(x, y) = \prod_{j=1}^J W_j(x, y).
\]

Optimal detection was achieved with \( J = 4 \). Finally, the number of peaks, \( N \), is determined as the number of points \( (x, y) \) for which \( P(x, y) > 0 \) holds and \( A(x, y) \) is a local maximum (in the original image). The intensity of diffraction-limited peaks was determined by fitting with a two-dimensional Gaussian profile as described (Schmidt et al. 1996).

Acknowledgments

This study was supported by the GEN-AU program of the Austrian Federal Ministry of Education, Science, and Culture, by the Austrian Research Fund, by the state of Upper Austria, and by the Priority Program “Life Sciences and Health” of the University of Salzburg.

References


Received December 2, 2005; accepted in revised form May 4, 2006.