

Gene expression profiling of single cells on large-scale oligonucleotide arrays

Claudia H. Hartmann and Christoph A. Klein*

Institut für Immunologie, Ludwig-Maximilians Universität, Goethestr. 31, 80336 München, Germany

Received July 10, 2006; Revised September 6, 2006; Accepted September 25, 2006

ABSTRACT

Over the last decade, important insights into the regulation of cellular responses to various stimuli were gained by global gene expression analyses of cell populations. More recently, specific cell functions and underlying regulatory networks of rare cells isolated from their natural environment moved to the center of attention. However, low cell numbers still hinder gene expression profiling of rare *ex vivo* material in biomedical research. Therefore, we developed a robust method for gene expression profiling of single cells on high-density oligonucleotide arrays with excellent coverage of low abundance transcripts. The protocol was extensively tested with freshly isolated single cells of very low mRNA content including single epithelial, mature and immature dendritic cells and hematopoietic stem cells. Quantitative PCR confirmed that the PCR-based global amplification method did not change the relative ratios of transcript abundance and unsupervised hierarchical cluster analysis revealed that the histogenetic origin of an individual cell is correctly reflected by the gene expression profile. Moreover, the gene expression data from dendritic cells demonstrate that cellular differentiation and pathway activation can be monitored in individual cells.

INTRODUCTION

At least 200 cell types can be discriminated in the human body (1) that may also pass through various states of cellular differentiation. In addition, each cell may be engaged in cellular functions such as proliferation, migration, senescence, or may be in an activated or in a resting state. Extrinsic factors (such as drugs) or intrinsic damage (such as mutations) may additionally affect cell function. Important insights into cellular responses were gained by widespread applications of gene expression analyses of cell populations (2,3). However, established methods for array-based gene

expression analysis require at least 50 000 cells (4), which lead to the development of different methods for amplifying mRNA based either on linear amplification by T7 RNA polymerase (5) or on exponential amplification via PCR (6). So far, linear amplification methods have not reached sensitivity for global gene expression profiling of single cells, in contrast to PCR-based protocols (7,8).

Relatively few protocols have been published that demonstrate successful analysis of single cells on cDNA or oligonucleotide arrays. In most cases protocols had been established by using dilutions of total RNA, but were seemingly less successful using real single cells. One likely reason for this discrepancy may be that pipetting errors exclude precise quantification of mRNA copy numbers down to 5–10 copies per cell for individual transcripts or 2–6 pg of mRNA as single cell equivalents (1). The perhaps most frequently applied protocol for single cell amplification by Brady and Iscove (6) which had also been used for array analysis (8) was the starting point of our own modified version (7). We previously found that the applied conditions of limited processivity for reverse-transcriptase during cDNA synthesis (e.g. low concentration of cDNA synthesis primers and nucleotides and short reaction time) and the poly-T primer for the PCR amplification severely reduced sensitivity and developed a protocol to overcome these shortcomings (7,9). However, even this protocol did not allow analysis of single cells with extremely low mRNA content (see below), although dilution experiments and direct analysis of human cancer cells had demonstrated a higher sensitivity than protocols based on the Brady procedure. We therefore aimed to develop a robust method for genome-wide gene expression profiling of single cells on large-scale oligonucleotide microarrays with exquisite sensitivity for low abundant transcripts, suited for all cell types alike.

Here we present a PCR-based method that avoids distortion of transcript abundance by solid-phase purification of mRNA allowing cDNA synthesis and amplification under optimal enzymatic conditions. Preservation of transcript ratios was carefully evaluated by quantitative PCR (qPCR) directly on the single cell level and not by using diluted RNA. Last but not least, a large set of microarray experiments using individually isolated cells of various histogenetic origins confirmed the high sensitivity and reproducibility of the method for genome-wide analyses.

*To whom correspondence should be addressed. Tel/Fax: +49 89 218075696; Email: christoph.klein@med.uni-muenchen.de

MATERIALS AND METHODS

mRNA isolation, cDNA synthesis and global amplification

Single cells isolated by micromanipulation were placed in 5 μ l lysis buffer (Active Motif, Rixensart, Belgium) supplemented with 1 μ g protease (Active Motif), 1 μ l biotinylated oligo-dT peptide nucleic acids (PNAs) (Midi-Kit, Active Motif, dissolved in 400 μ l water) and 10 ng tRNA (Roche, Mannheim). The proteolytic digest was performed at 45°C for 10 min, followed by 1 min at 70°C and 15 min at 22°C for PNA annealing. mRNA was isolated with 4 μ l streptavidin beads (Active Motif) during 45 min rotation at room temperature. Ten microliters of cDNA wash buffer 1 (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT and 0.25% Igepal) were added and the tubes placed into a magnetic rack. The supernatant containing the genomic DNA was collected and the beads were washed using 20 μ l of cDNA wash buffer 2 (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT, 0.5% Tween-20). The supernatant was removed and the step repeated with 20 μ l cDNA wash buffer 1. Reverse transcription was carried out under rotation for 45 min at 44°C in a mix containing 0.5 mM dNTPs, 200 U Superscript II (Invitrogen, Karlsruhe), 30 μ M CFL15CN8 primer (C₁₅GTC-TAGAN₈), 15 μ M CFL15CT24 primer (C₁₅GTCTA-GAT₂₄VN), 0.25% Igepal, 10 mM DTT (Invitrogen) and the buffer supplied by the manufacturer in a final volume of 20 μ l. Primers were annealed at room temperature for 10–15 min before addition of the enzyme. After completing the reverse transcription, beads were washed in 20 μ l tailing wash buffer (50 mM KH₂PO₄ pH 7, 1 mM DTT, 0.25% Igepal) and resuspended in 10 μ l tailing buffer (10 mM KH₂PO₄ pH 7, 4 mM MgCl₂, 0.1 mM DTT, 200 μ M dGTP). After denaturation of the cDNA-mRNA hybrids at 94°C for 4 min, 10 U Terminal deoxynucleotide transferase (Amersham, Freiburg) were added and the samples incubated at 37°C for 60 min for the G-tailing reaction. After inactivation of the tailing enzyme (70°C, 5 min), PCR mix I [4 μ l buffer 1 (Expand Long Template, Roche), 3% deionized formamide] was added to each sample. A hotstart PCR was performed adding 5.5 μ l PCR mix II [350 mM dNTPs, 1.2 μ M CP₂ primer (TCAGAATTCATGC₁₅) and 5 U Pol Mix (Expand Long Template)]. Forty cycles were run in a MJ research PCR machine: 20 cycles of 15 s at 94°C, 30 s at 65°C, 2 min at 68°C and 20 cycles with an elongation of the extension time of 10 s and a final elongation step of 7 min at 68°C.

Sample labelling and array hybridization

Primary cDNA amplification products were labelled in the presence of 3% formamide, 2.4 μ M CP2-BGL primer (TCA-GAATTCATGCCGCCCCCGGCC), dNTPs (0.35 mM dATP and dGTP, 0.3 mM dTTP and dCTP) and 0.05 mM labelled nucleotides. Sample cDNA was labelled with digoxigenin-dUTP (Roche) and aminodigoxigenin-dCTP (PerkinElmer, Rodgau-Jügesheim), reference cDNA with biotin-dUTP (Roche) and biotin-dCTP (Invitrogen). Universal reference cDNA was pooled from various mouse tissues and murine cell lines. The primers were removed by a digest with 30 U *Bgl*II (Fermentas, St Leon-Rot) and the samples

purified on a purification column (Qiagen, Hilden). Sample and reference cDNA were coprecipitated with 50 μ g Cot1-DNA (Invitrogen) and 15 μ g *E.coli*-DNA. Murine Operon 70mer oligonucleotide arrays (Version 2) spotted on Ultra-GAPS slides (Operon, Köln resp. Corning, Schiphol-Rijk, Netherlands) were prehybridized with 5 \times SSC + 0.1% SDS + 0.1% BSA at 42°C and hybridized in an Arraybooster hybridization (Implen, Munich) at 42°C overnight. Washing steps following hybridization were at 50°C twice in 1 \times SSC + 0.1% SDS for 10 min, twice in 0.5 \times SSC + 0.1% SDS for 10 min, and twice in 0.1 \times SSC for 30 min. Unspecific binding of labelled proteins was blocked with 1% blocking reagent for nucleic acid hybridization (Roche) followed by a staining procedure with anti-Dig-Cy5 and Streptavidin-Cy3 (each 16 μ g/ml, Jackson Laboratories). Excess antibody/streptavidin was removed with 4 \times SSC + 0.2% Tween-20 and slides were scanned on a Genepix 4000A scanner (Axon Instruments, Union City). Experiments shown in Table 2 were performed using human cells hybridized to Human Genome OpArrays version 4. Protocol details are available upon request.

Data analysis

GenePix result files were loaded into GeneSpring and normalized using Lowess. Features that were called 'absent' in all samples were removed from further analysis. Samples were clustered using the GeneSpring condition tree function based on standard correlation. Of the 16 928 genes on the array, 15 988 were present in at least one sample. In Figure 6 only single cells were clustered based upon the 508 genes whose expression was correlated with the sample type at $P < 0.001$ by multivariate analysis using the GeneSpring ANOVA function 'Find significant parameter'.

For analysis of reproducibility among technical replicates, the GeneSpring sample correlation function in the 'Find similar samples'-category was used.

The lipopolysaccharide (LPS) pathway was analyzed by importing the KEGG murine Tlr4 signaling pathway into GeneSpring. The mean expression values of all samples from mature and immature dendritic cells (DCs) can then be displayed for each gene.

Evaluation of amplification bias

Single cells and ten cells of the epithelial TUBO cell line were isolated under the microscope. Single cell equivalents were generated by lysing a pool of ten cells and splitting the lysate into ten technical replicates. One half of each sample type was processed according to the global amplification protocol, the other half was only reverse transcribed. The resulting cDNA was used for gene-specific qPCR.

Quantitative PCR

Real-time PCR was performed using a LightCycler (Roche) and Fast Start Master SYBR Green Kits (Roche). Analysis was done using the RelQuant software (Roche) with PCR efficiency normalization and a reference sample included in every run. Measurements showing unspecific products in the melting curve analysis were discarded from further analysis. All expression levels are given relative to *Gapdh*, which served as a loading control. Primer sequences are provided upon request.

Culture and stimulation of murine DCs

Bone marrow cells were isolated from femur and tibia of two C57/black6 mice and cultivated in the presence of 25 ng/ml GM-CSF for six days to generate dendritic cells (DCs). Immature DCs were harvested and enriched using Magnetic affinity cell sorting (MACS) technology according to the manufacturer's instructions (Miltenyi, Bergisch Gladbach) selecting for CD11c and depleting of cells expressing CD86. Samples were stained with α CD11c-FITC (BD Pharmingen, Heidelberg) and single positive cells were isolated with a micromanipulation device (Eppendorf, Hamburg). A second culture dish was stimulated with 1 μ g/ml LPS (Sigma) overnight and cultured for two more days. Mature DCs were positively selected for CD86 and CD11c using MACS. Cells were stained and isolated as immature DCs.

Ex vivo isolation of murine hematopoietic progenitor cells

Bone marrow cells were isolated from femur and tibia of a BALB/c mouse. Differentiated, lineage marker positive cells were depleted using MACS technology according to the manufacturer's instructions (Miltenyi, Bergisch Gladbach). Lineage marker negative cells were stained with Hoechst 33342 (Sigma) for 30 min as previously described (10), followed by staining for CD34 (BD Pharmingen). Single cells were isolated using a micromanipulation device (Eppendorf).

RESULTS

Isolation of mRNA and amplification

We previously described a protocol for global amplification and analysis of single cell transcripts. It was based on isolation of the cellular mRNA by oligo-dT beads, cDNA synthesis using random octamer and oligo-dT₍₁₅₎ primers, poly-dG tailing, and PCR amplification using a single primer under very stringent conditions adequate for CG-rich sequences (7). The solid-phase capturing of the mRNA enabled depleting the cDNA from free cDNA synthesis primers and abundant dNTPs thus avoiding the subsequent amplification of tailed primers or inefficient tailing. Together with the introduction of random primers (resulting in a fragmentation of the cDNA to a size amplifiable by PCR) the protocol therefore allowed for optimal conditions of all enzymatic reactions such as high concentrations of dNTPs, primer and enzyme, all of which were individually shown to contribute to increased sensitivity (7). The highest increase of sensitivity (100-fold) over the original Brady procedure was achieved by the use of poly-G tailing instead of poly-A tailing and the subsequent use of a single poly-C-containing primer (7). The protocol was suitable for analysis of individual human cells (7,9) but unfortunately failed repeatedly for single mouse cells, which contain only 10–20% of the total RNA isolated from human cells (Figure 1B and C, data not shown). Since cDNA synthesis and PCR conditions had been shown to be extremely sensitive, we reasoned that

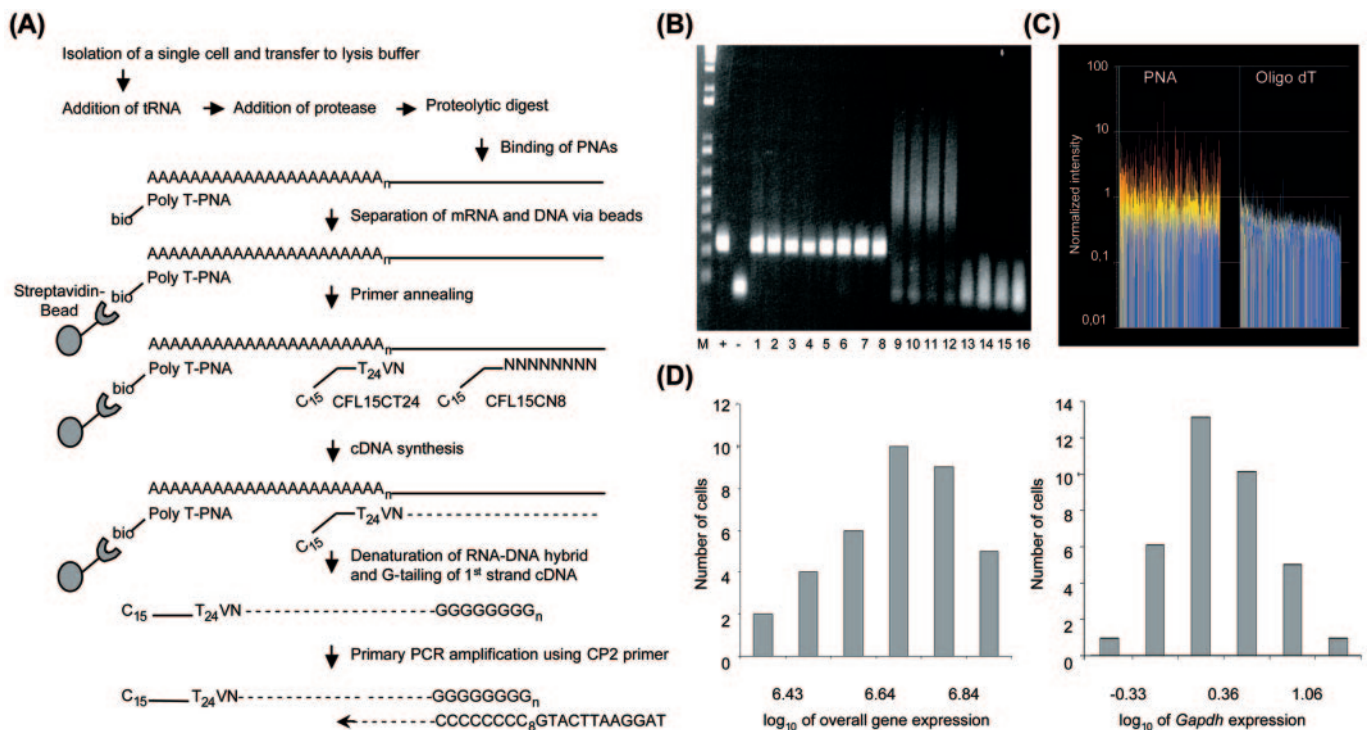


Figure 1. Flowchart of the protocol and increased sensitivity by PNAs. (A) Schematic overview of the protocol. The synthesis of the cDNA and subsequent steps are only shown for the primer binding to the poly-A tail of the mRNA. (B) Single cell mRNA from mouse cells was isolated with both methods and tested for β -actin (samples 1–4, PNAs; samples 5–8, oligo-dT) and for the total amount of the amplification products (samples 9–12, PNAs; samples 13–16, oligo-dT). Note that longer sequences are difficult to detect using oligo-dT beads. (C) Signal intensities for all probes on large-scale oligonucleotide arrays. The left histogram depicts intensities obtained after mRNA isolation with poly-T PNAs and the right one with oligo-dT beads. High signal intensities are in red and yellow, while blue colors represent low signal intensities. (D) The overall gene expression (left) and expression of *Gapdh* (right) are distributed log-normally.

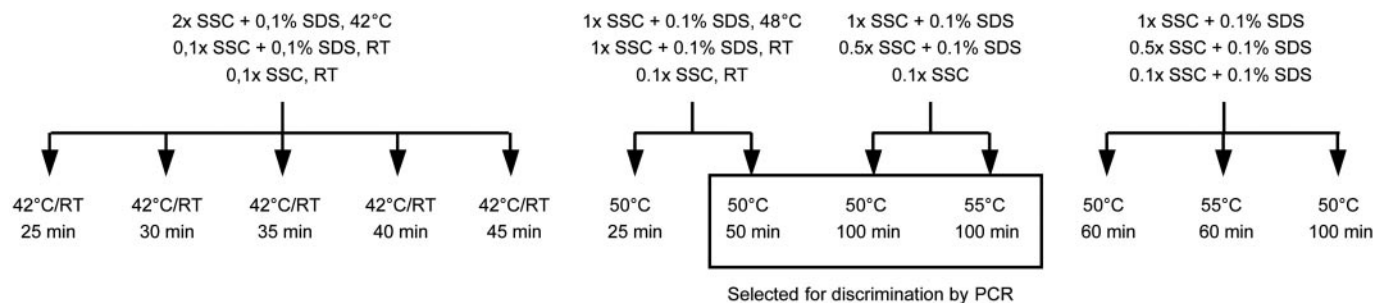


Figure 2. Hybridization of single cell cDNA to oligonucleotide arrays. Washing conditions of varying stringency (increasing from left to right) were tested for sensitive and specific array hybridization. The most suitable conditions were further discriminated by PCR.

one crucial step could be the solid-phase isolation of the mRNA itself.

The process of isolating and amplifying cellular mRNA was improved to a great extent by introducing two modifications. First, application of biotinylated poly-T PNAs bound to streptavidin beads instead of oligo-dT beads for the isolation of the mRNA (Figure 1A) significantly increased yields from single murine cells (Figure 1B). Using PNAs signal intensities were 4–5 times higher (median 2.3-fold) as compared to oligo-dT beads. Likewise the dynamic range increased for about 10-fold (Figure 1C). As a first quality check of the modified protocol, we tested whether the transcriptional activity of single cells (by summing up the intensity signals for all genes on the array for each cell) and specifically of *Gapdh* ($n = 36$) was log-normally distributed as has been observed previously (11). This was indeed the case (Figure 1D), suggesting that the method preserves this biological characteristic of single cell gene expression. Taken together, it appears that rare transcripts are isolated with much higher probability by PNAs because of their much higher affinity to the poly-A tail of the mRNA as compared with oligo-dT beads. Second, adding a longer oligo-dT-containing cDNA synthesis primer to the random octamer further improved results (Table 1).

Conditions for microarray analysis

We then sought to establish hybridization conditions for amplified single cell cDNA on large-scale microarrays containing 17 000 oligonucleotides of 70 nt length.

Sensitivity and specificity of hybridization techniques are controlled to a large extent by the stringency of the post-hybridization washing procedure. Controlling for specificity is especially important when single cell amplicates are hybridized due to the presence of contaminating bacterial sequences in the sample. This bacterial DNA is inevitably introduced with the various recombinant enzyme preparations added during the process of amplification and is co-amplified and labelled, which may result in false-positive hybridization signals. To establish hybridization and washing conditions for the oligonucleotide arrays, a reference positive control sample was co-hybridized with a negative control consisting of a product of the whole protocol with no cell being added. Critical factors for the hybridization of amplified single cell cDNA also comprised the specific sequences of the oligonucleotides on the array, the surface characteristics of the glass slide (epoxide or silan surfaces), and the spotting

buffers and spotting conditions (data not shown). We therefore recommend performing a short series of hybridization experiments with any type of microarray in a manner similar to Figure 2.

Hybridization quality was assessed by the present call of positive and negative control samples, signal intensities of control spots, and fluorescence background on the glass slide. We first established washing conditions that prevented detection of unspecific sequences amplified in the negative control samples (Figure 2). Subsequently, in all further experiments a transcript was called present when the intensity value passed the threshold provided by the GeneSpring error model (<http://www.chem.agilent.com>). For single cells, optimal washing conditions resulted in present calls for 25–30% of all oligonucleotides on the array, confirming high sensitivity for low abundant transcripts on the microarray as well. To validate the array data we randomly selected 44 genes, with and without hybridization signal, from the 17 000 oligonucleotides on the array and compared expression data of PCR and hybridization in 13 single cells. Seventy percent of randomly selected oligonucleotides on the array gave concordant results with gene-specific PCR. While the 8.8% of transcripts solely detected by PCR may indicate the higher sensitivity of PCR, we determined probe design as cause for the discrepancy observed for 13 of the 26 sequences resulting in false positives on the array (data not shown). Thus, of the divergent results one-third each is explained by probe design and one third by characteristics of the particular detection method.

Finally, we evaluated our global amplification and hybridization protocol to answer the question, which of the different technical modifications had the highest impact on the correlation coefficient of technical replicates in microarray experiments. Interestingly, mRNA isolation by PNAs resulted in the single largest improvement, while the choice of cDNA synthesis primers, the number of cycles for reamplification/labelling, the usage of a single fluorescence-labelled nucleotide (dUTP) versus two fluorescence-labelled nucleotides (dUTP + dCTP), and the application of an automated hybridization chamber contributed to a lesser extent to reliable measurements (Table 1). The resulting correlation coefficient for single cell equivalents (i.e. three individually isolated cells, pooled and lysed and then divided into three equal parts for individual amplification and hybridization) that combined all the superior technical means was ~0.9 and was similar to values obtained with 100 and 200 pooled cells (Table 2).

Quantitative evaluation of global amplification and hybridization by qPCR

Generally, quantification of transcript abundance requires that their relative ratios remain unchanged by the amplification procedure. Exponential PCR-based methods for global mRNA amplification are often assumed to introduce large quantitative changes in transcript scores, because differences of amplification efficiency between individual transcripts may be transmitted exponentially during amplification. But also for the so-called linear amplification techniques based on Eberwine's T7-method (12), it is still unclear whether the original proportions of mRNAs are preserved (13–15). To test for the introduction of systematic bias during the amplification procedure of the novel PCR-based protocol, we isolated individual single cells and pools of ten cells from the murine mammary carcinoma cell line TUBO. Instead of diluting micrograms of total RNA down to the single cell level, we generated single cell equivalents by pooling ten cells (isolated by micromanipulation) in lysis buffer and dividing the mRNA into ten technical replicates. The samples were subjected to either cDNA synthesis followed by global amplification or to reverse transcription only. The resulting cDNA, either amplified or not, was used as template for transcript-specific qPCR (Figure 3A). Thereby, we could directly compare the effect of the global amplification on single cell-derived transcripts for a limited number of sequences. Because most transcripts are known to be present in copy numbers of only 10–15 in a single cell (1), at most 10 assays could be performed using the

unamplified cDNA. We found that global amplification did not change the transcript ratios because the mean relative expression ratios were similar for all samples (Figure 3B). Also, variability coefficients as a measure of noise were 18.5% and 16.5% for amplified and non-amplified single cell equivalents, respectively. As expected, variability coefficients were highest for individual single cells (amplified 49.8% and non-amplified 110.9%) reflecting cellular heterogeneity. The global amplification method was sensitive enough to measure very rare transcripts in single cells that could hardly be detected in the unamplified pools of ten cells (Figure 3B and C).

It has been proposed that the amplification reaction should be stopped before reaching the plateau phase or that certain sequences are amplified preferentially (16–18). We investigated single cells and stopped amplification during the exponential phase with half of the sample and continued amplification with the remaining half to the plateau phase. Interestingly, qPCR results demonstrated that relative transcript ratios were not altered when the template cDNA was taken from the plateau-phase of the PCR. In addition, the correlation coefficients of the hybridization results relative to non-amplified total RNA dropped in all cases, when cycling was stopped at the exponential phase of amplification as compared with the samples taken from the plateau phase (Figure 3 and data not shown). We therefore do not recommend stopping the primary amplification too early.

After activation of murine DCs by LPS, isolation of activated and non-activated cells, global single cell mRNA amplification, and array hybridization we selected ten genes that were either upregulated on the array, downregulated or unchanged during the maturation process. When we applied real-time PCR, all genes were identically classified by the two methods and transcriptional regulation was found in a similar order of magnitude (Figure 4A and B). The observed differences between array and PCR results for single cells were similar to those published for conventional gene expression profiling studies (19). Thus, all qPCR experiments confirmed that the global amplification method does not change relative transcript abundances and that the established hybridization conditions enable quantification of transcript numbers from single cells with similar accuracy as methods using large amounts of starting RNA.

Table 1. Factors influencing experimental outcome

Protocol modification	Increase of correlation coefficient to total RNA
Extraction of mRNA with PNAs versus oligo-dT beads	0.41
cDNA synthesis primer containing 24 thymidine nucleotides [oligo-dT ₍₂₄₎] versus 15 thymidine nucleotides [oligo-dT ₍₁₅₎]	0.13
Labelling reaction applying 16 cycles versus 10 cycles PCR	0.09
Labelling with dig-dUTP and dig-dCTP versus dig-dUTP alone	0.04
Use of hybridization station versus manual hybridization	0.10

Table 2. Correlation coefficients between technical replicates

	Single cell equiv. 1 ^a	Single cell equiv. 2	Single cell equiv. 3	100 cell pool 1	100 cell pool 2	200 cells donor A, sample 1 ^b	200 cells donor A, sample 2	200 cells donor B, sample 1	200 cells donor B, sample 2
Single cell equiv. 1 ^a	1	0.897	0.868	0.794	0.772	n.d. ^c	n.d.	n.d.	n.d.
Single cell equiv. 2	0.902	1	0.909	0.842	0.830	n.d.	n.d.	n.d.	n.d.
Single cell equiv. 3	0.869	0.917	1	0.875	0.860	n.d.	n.d.	n.d.	n.d.
100 cell pool 1	0.779	0.813	0.846	1	0.852	n.d.	n.d.	n.d.	n.d.
100 cell pool 2	0.809	0.849	0.880	0.920	1	n.d.	n.d.	n.d.	n.d.
200 cells donor A, sample 1 ^b	n.d.	n.d.	n.d.	n.d.	n.d.	1	0.921	0.865	0.828
200 cells donor A, sample 2	n.d.	n.d.	n.d.	n.d.	n.d.	0.919	1	0.868	0.870
200 cells donor B, sample 1	n.d.	n.d.	n.d.	n.d.	n.d.	0.796	0.823	1	0.867
200 cells donor B, sample 2	n.d.	n.d.	n.d.	n.d.	n.d.	0.822	0.865	0.909	1

^aCell line HT-29.

^bSamples were generated by fluorescence-activated cell sorting (FACS) sorting of hematopoietic stem cells (HSCs) from two different human donors.

^cNot determined.

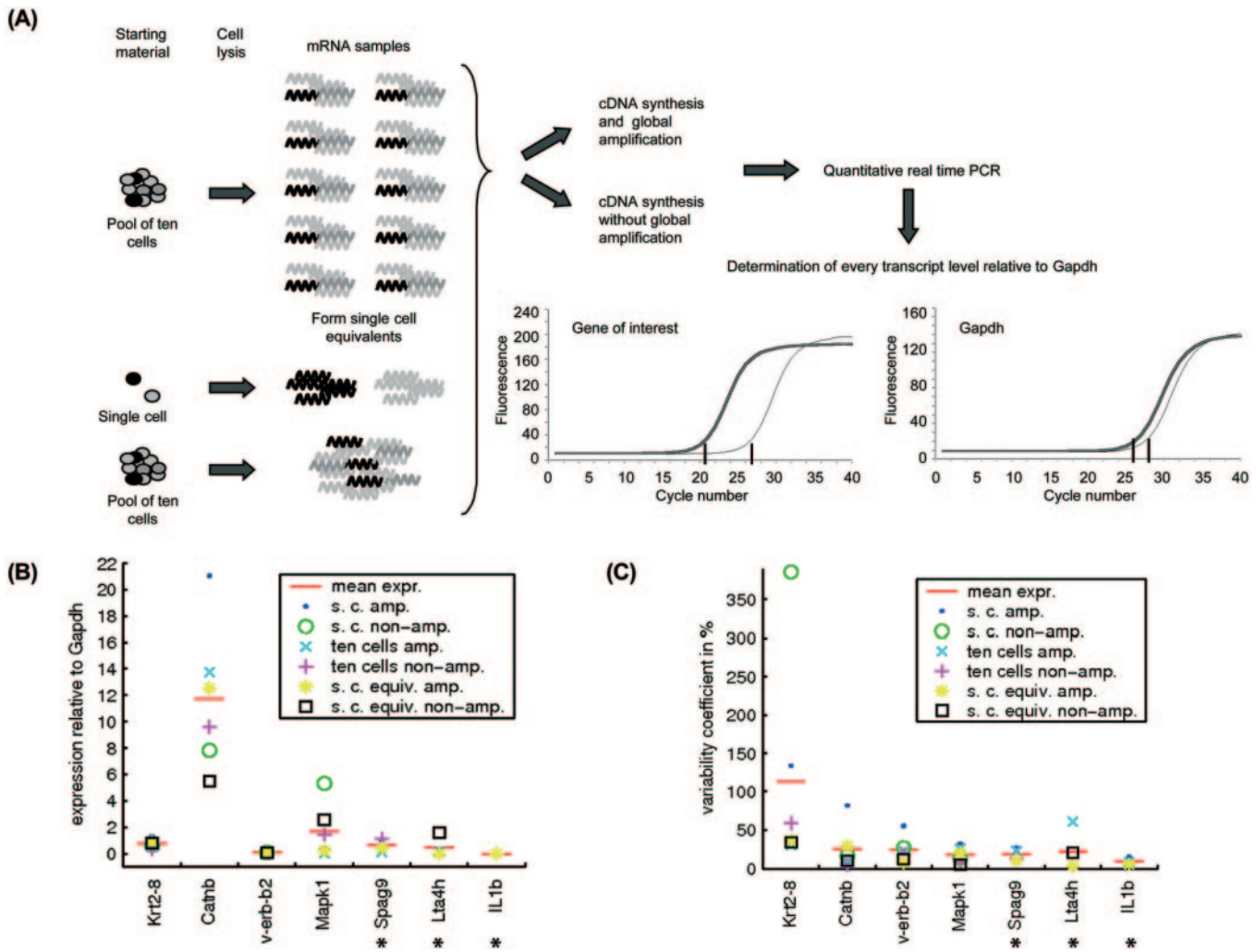


Figure 3. Quantitative assessment of amplification bias. (A) Experimental design to compare gene expression levels from samples with and without global amplification. Grey shades indicate various functional states of cells or mRNAs expressed during these states. (B and C) Expression levels and coefficients of variability of randomly selected genes for the different groups. Asterisks indicate transcripts that could not be detected without prior global cDNA amplification. Confidence intervals for the coefficients of variability are given in Supplementary Table 1.

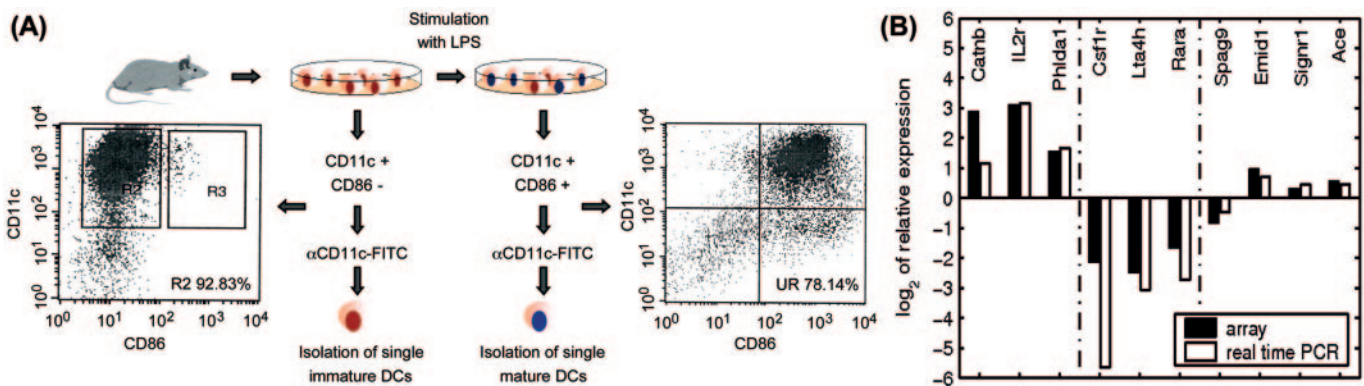


Figure 4. Analysis of mature and immature DCs. (A) Bone marrow was cultivated to generate DCs. CD11c positive cells (DCs) were purified using MACS and purification was controlled by FACS analysis. Single CD11c positive cells were then isolated by micromanipulation under microscopic control. Mature DCs were enriched using CD86 as marker after overnight incubation with LPS, followed by two additional days of culture, and were isolated after CD11c staining. (B) Comparison of array and qPCR results for differentially regulated genes. Genes are grouped as up-, down- or unregulated as indicated by the dashed lines.

Molecular portraits of single cells of different histogenetic origin

It has never been determined whether the phenotype of a single cell can be correctly retrieved from large gene expression studies and whether small changes of cellular differentiation of single cells are mirrored by their molecular profiles. We therefore hybridized cells of various histogenetic origin and differentiation state to the 17k array, including epithelial cells, dendritic cells and hematopoietic stem cells. From the epithelial TUBO cell line ten individual cells were isolated, four pools of ten and 100 cells, (unamplified) total RNA, and a series of globally amplified equivalents for 100 cells, 10 cells, 1 cell and 0.1 cell diluted from total RNA to assess the detection limit of the method. A single cell equivalent was adjusted to 6 pg mRNA (1). Eight single hematopoietic stem cells (HSCs) defined by either expression of the stem cell marker CD34 and/or exclusion of the Hoechst 33342 dye, which characterizes the so-called side population (10), were isolated from murine bone marrow. A minimal difference in cellular differentiation was assumed for the DCs, which were either stimulated by LPS (i.e. so-called mature DCs) or unstimulated (immature DCs). Hierarchical cluster analysis correctly grouped all but one cell according to their differentiation state and histogenetic origin (Figure 5). The high reliability of the method was further demonstrated by the finding that both technical replicates of two single TUBO cells as well as all samples from the dilution series of cell equivalents were grouped close to each other. Even the two 0.1 cell dilutions were not only grouped with the TUBO cells but also clustered together. Immature DCs formed a subgroup, from which mature DCs seemed to diverge, possibly reflecting different stages of maturation. Finally, all CD34+ HSCs clustered together—regardless of the Hoechst staining—whereas the single Hoechst-negative, CD34-negative stem cell was separated.

We then proceeded to identify genes that differentiate the different cell types on a single cell level. To this end, we excluded all samples from the dilution series and the technical replicates and concentrated only on the gene expression profiles of individually isolated and amplified single cells. The remaining 30 samples were subjected to multivariate analysis to define genes whose expression was significantly correlated with the sample type. A hierarchical cluster analysis based upon a list of genes that significantly differed between the groups ($P < 0.001$; $n = 508$) robustly separated the samples. Among the genes that were expressed in a tissue-specific manner, we found the epithelial marker keratin 18 and the tight junction gene claudin 3 to be significantly up-regulated in epithelium-derived cells, the maturation of DCs was associated with the elevated expression of ribosomal transcripts, and HSCs could be differentiated from the other cell types by expression of *Egln1*, *Sh3d2a* or *Dnmt3b* (Figure 6).

Pathway analysis in single cells

Cluster analysis already indicated the upregulation of several genes during the maturation of DCs (Figure 6). To get a closer look into the specificity of this process, we analyzed the activation of the LPS pathway for the two DC types. The current model of LPS-induced DC maturation holds that upon binding of the ligand to Tlr4, the MyD88-dependent

signaling pathway is activated, leading to the expression of inflammatory cytokines and costimulatory molecules (20,21). We implemented the gene expression data of all sixteen immature and mature DCs for the molecules represented in the KEGG pathway scheme. As expected, changes in gene expression associated with DC maturation induced by the applied conditions were confined to the MyD88-dependent signaling cascade. We observed upregulation of intracellular pathway components as well as of all costimulatory molecules (*CD80*, *CD86* and *CD40*) and inflammatory cytokines such as *IL-6* and in particular the chemokine *Rantes* (Figure 7). Neither the IRF-3 pathway nor the PI3-kinase cascade were activated under our experimental conditions. These results demonstrate that gene regulatory pathways can be explored on a single cell level.

DISCUSSION

Here, we present a highly sensitive protocol for global gene expression and gene regulatory pathway analysis of single cells on large-scale arrays. The protocol extends the application of our previously published method (7) to single cells with very low mRNA content, and to genome-wide oligonucleotide arrays. Its characteristics such as solid-phase capturing of nucleic acids and amplification by a single poly-dC-containing primer under stringent PCR conditions apparently lead to highly similar amplification efficiencies for all transcripts.

Here, we assessed the individual contribution of various protocol variations to successful microarray analyses. To this end, we measured for each modification the net gain of 'correlation' of a single cell sample to a standard consisting of total RNA. We found that the major increase in correlation was due to the usage of PNAs for mRNA isolation. Of minor importance were the choice of cDNA synthesis primers, the number of cycles for labelling, the method of labelling and the method of hybridization.

Since several protocols have been published claiming single cell sensitivity, but included mostly experiments with diluted total RNA, we concentrated on real single cells. When pooled RNA was employed, we isolated 2–10 cells and divided the mRNA to obtain single cell equivalents. We were able to define a hybridization protocol on Operon oligonucleotide arrays that resulted in excellent correlations for up- and downregulated genes with results obtained by qPCR. The method was vigorously tested on a large number of samples and the results clearly demonstrate that various cell types, including freshly *ex vivo* isolated HSCs, and thus not only cells with abundant amounts of mRNA, can be analyzed reproducibly on large-scale microarrays. In particular, the minute differentiation step from immature to mature DCs was correctly preserved.

Very recently, a PCR-based protocol for the amplification and analysis of single cell mRNA on Affymetrix GeneChip arrays has been published (22) which is based on a modification of the original Brady method. Interestingly, also this group found that by-products such as tailed cDNA synthesis primers had to be removed. Instead of using solid-phase capturing, this group introduced an exonuclease step to digest unincorporated primers. Otherwise, all previously identified

limitations (7) of the original Brady protocol apparently still apply. Since the authors used the Affymetrix GeneChip platform, a direct comparison with the method presented here is not possible. However, both protocols strongly support the conclusion that PCR-based single cell amplification methods are suited for large-scale gene expression profiling adding to the previously discussed advantages of PCR-based protocols over linear amplification protocols (22), which are increased sensitivity and easy handling of PCR-based methods. The protocol described here does not require

unusual equipment, is very easy to perform and robust, including the steps of cell isolation and processing. Isolation of rare cells from human tissues or animals often greatly extends the duration of the total experiment, so time considerations are very important. We found that—prior to cDNA synthesis—the protocol can be interrupted at any stage with the notable exception that the proteolytic digest has to be performed immediately after addition of the protease (Figure 1A and data not shown). In particular, isolated cells can be immediately frozen in lysis buffer containing tRNA

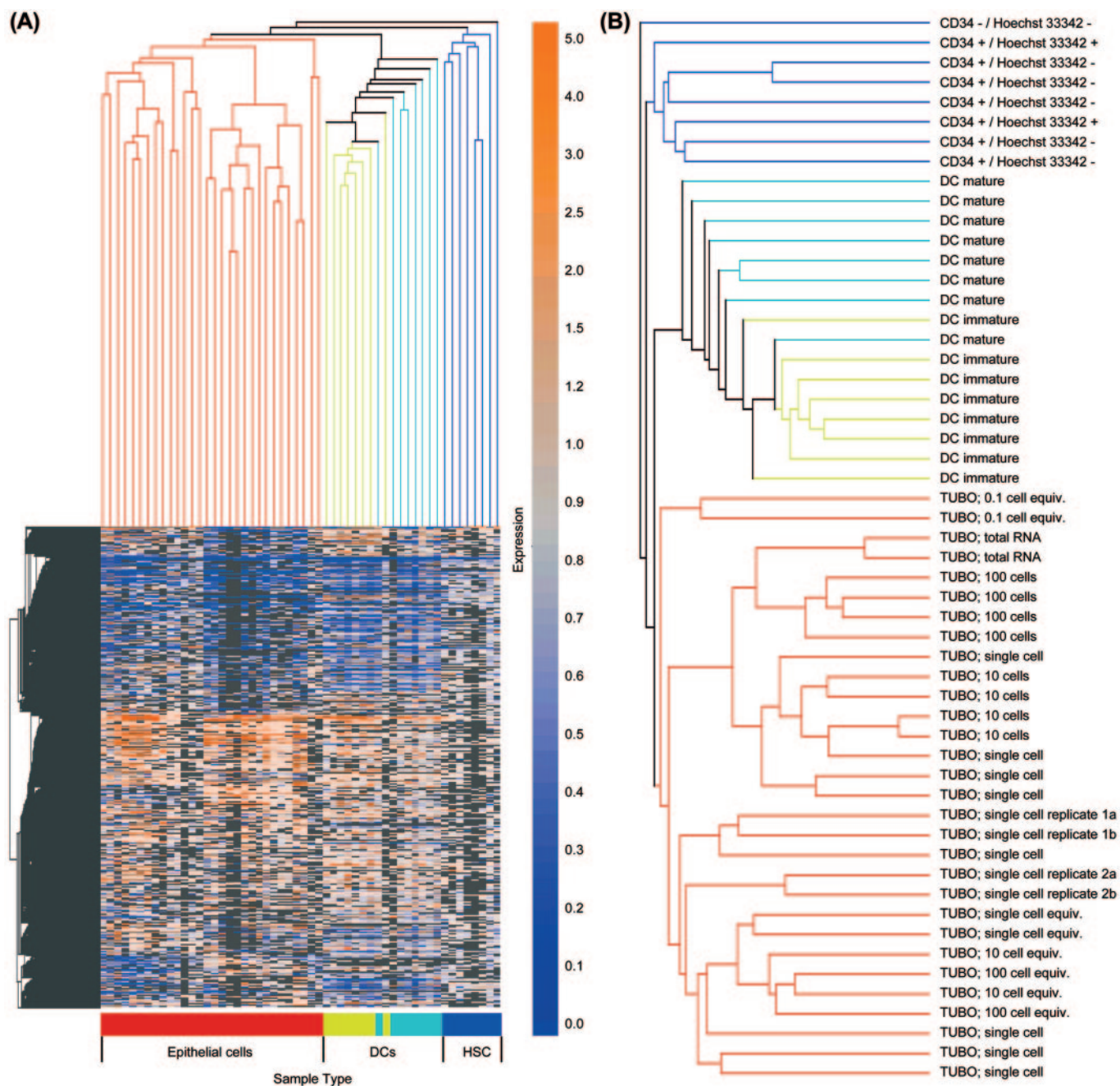


Figure 5. Unsupervised clustering of samples of various histogenetic origins. (A) A hierarchical clustering algorithm separates epithelial cells from DCs and HSCs. (B) Dendrogram from panel (A) shown in higher magnification with sample identifier. DC, dendritic cell; TUBO, epithelial mammary carcinoma cell line; equiv., equivalent. HSC, hematopoietic stem cell. Clustering of samples is based on values from all oligonucleotides that were found expressed in at least one sample (15 988 of 16 928 oligonucleotides).

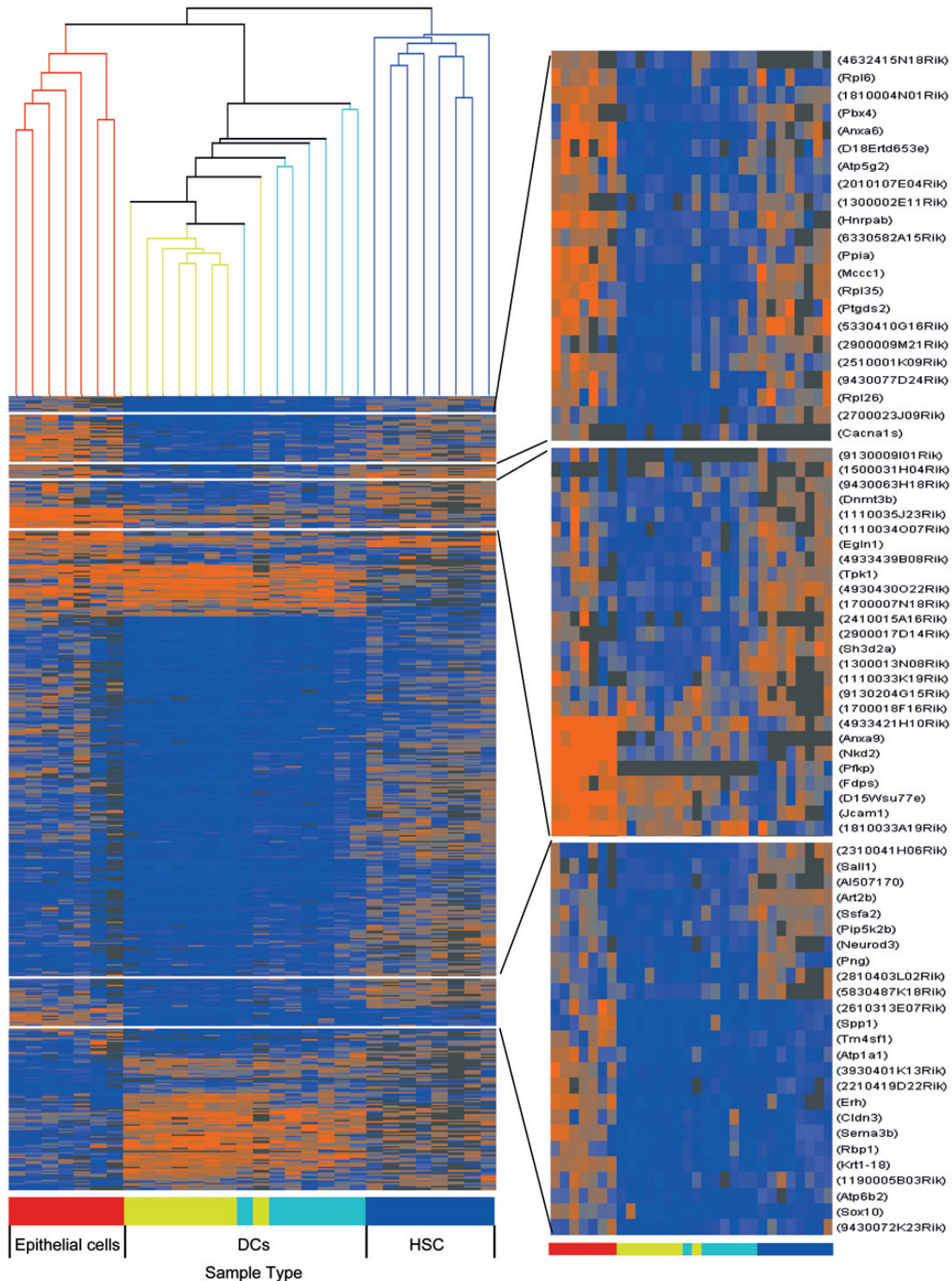


Figure 6. Molecular portraits of single cells. Unsupervised clustering of single cell samples was performed using those genes that were differentially expressed among cell types. Epithelial cells (red), DCs (immature: yellow, mature: cyan) and HSCs (blue) were grouped based upon the expression of the 508 genes for which P -values were <0.001 . Certain genes are upregulated during DC maturation (upper right panel), others are more highly expressed in the progenitor cells (middle right panel) and a third group is most prominent in the epithelial cells (lower right panel). Expression values are color-coded as in Figure 5.

without notable loss of transcripts. Successful application in an independent laboratory and therefore the robustness of the method was already demonstrated by gene expression analysis of single cardiomyocytes from the hearts of old

versus young mice. The experiments revealed a significant stochastic deregulation of gene expression with age (23), offering some basic insight as to how organ function may become impaired at old age. Thus, gene expression analysis

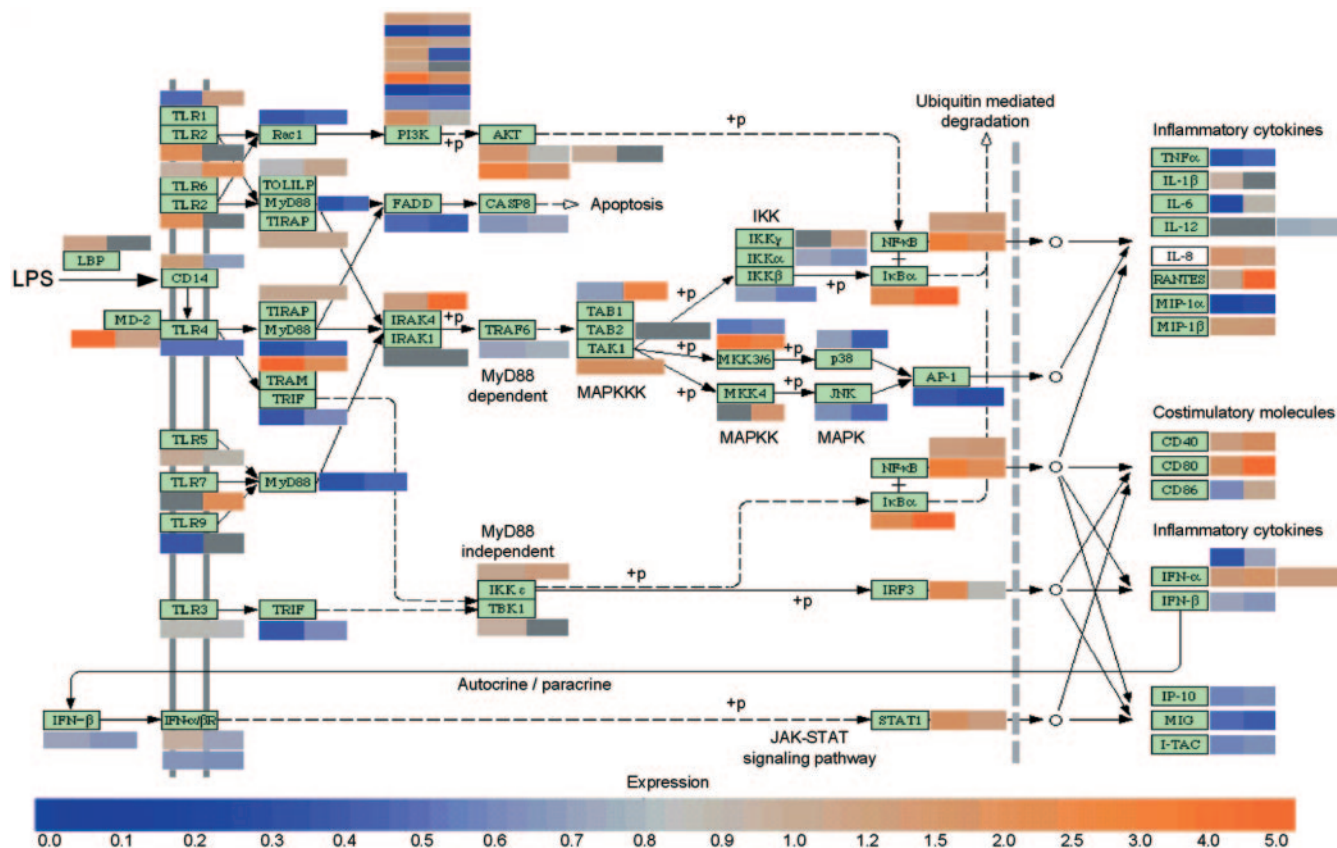


Figure 7. Tlr4 pathway analysis of immature and mature DCs. LPS binds to CD14 and activates Tlr4 in DCs. The mean expression values of the cells were implemented into the KEGG pathway scheme for the eight immature samples (depicted at the left of each colored box) and for the eight mature DCs (at the right). Levels of expression are indicated by the color code. Dark grey tone indicates missing expression on the arrays. Solid arrows indicate direct and dashed arrows indirect activation. (Detailed legend information can be found on the KEGG website: <http://www.genome.jp/kegg/>).

of single cells may be widely used for various biological questions and extend our understanding of complex processes in multicellular organisms.

Attempts to establish transcriptional networks from single cells are confronted with the fact that gene expression in single cells is inherently stochastic (24,25) and that the general expectation of highly ordered and consistent gene expression in a single cell is inadequate (26,27). Among the many factors that contribute to the stochasticity of gene expression and thereby to intercellular heterogeneity, low transcript abundance, small cell size and a large size of gene networks have been firmly established (26). To challenge the method, we used murine cells that are smaller and contain significantly less mRNA per cell than human cells, but have comparable gene networks. Despite these intrinsic difficulties it was possible to perform an analysis of the LPS/TLR4 pathway activation in stimulated DCs by implementing the gene expression values of our samples into the pathway scheme from the KEGG database. We found that maturation under the applied conditions correlates best with the activation of the IRAK4-NF κ B pathway on a transcriptional level. It is evident that refined bioinformatics methods must be developed to explore the transcriptional state of an individual cell in more detail. That this might be feasible is suggested by the finding that the global gene expression patterns correctly defined the histogenetic origin of tested

single cells. Given the probabilistic nature of gene expression in individual cells on one side and the inevitable experimental noise on the other side, it is remarkable that the single murine cells of various histogenetic origins were correctly classified. This finding suggests that the method will allow the definition of so far unknown differentiation stages of cells isolated from normal and diseased tissues.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Jan Vijg and Yousin Su for providing us with the Operon microarrays, and Melanie Werner-Klein for her kind support with the murine DC culture. We also thank Jan Vijg for carefully reading the manuscript and Kerstin Möhr for excellent technical assistance. The work was funded by the BioFuture grant 0311884 from the German Federal Ministry for Education and Science, the Wilhelm-Sander Stiftung, and a grant from the Bavarian State Ministry of Sciences, Research and the Arts. Funding to pay the Open Access publication charges for this article was provided by the Bavarian State Ministry.

Conflict of interest statement. Stocks of Micromet AG.

REFERENCES

- Alberts, B. (2002) *Molecular Biology of the Cell*, 4th. Edn. Garland Publishing, New York.
- Rhodes, D.R. and Chinnaiyan, A.M. (2005) Integrative analysis of the cancer transcriptome. *Nature Genet.*, **37** (Suppl.), S31–S37.
- Segal, E., Friedman, N., Kaminski, N., Regev, A. and Koller, D. (2005) From signatures to models: understanding cancer using microarrays. *Nature Genet.*, **37**(Suppl.), S38–S45.
- Mahadevappa, M. and Warrington, J.A. (1999) A high-density probe array sample preparation method using 10- to 100-fold fewer cells. *Nat. Biotechnol.*, **17**, 1134–1136.
- Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Zettel, M. and Coleman, P. (1992) Analysis of gene expression in single live neurons. *Proc. Natl Acad. Sci. USA*, **89**, 3010–3014.
- Brady, G. and Iscove, N.N. (1993) Construction of cDNA libraries from single cells. *Methods Enzymol.*, **225**, 611–623.
- Klein, C.A., Seidl, S., Petat-Dutter, K., Offner, S., Geigl, J.B., Schmidt-Kittler, O., Wendler, N., Passlick, B., Huber, R.M., Schlimok, G. et al. (2002) Combined transcriptome and genome analysis of single micrometastatic cells. *Nat. Biotechnol.*, **20**, 387–392.
- Tietjen, I., Rihel, J.M., Cao, Y., Koentges, G., Zakhary, L. and Dulac, C. (2003) Single-cell transcriptional analysis of neuronal progenitors. *Neuron*, **38**, 161–175.
- Klein, C.A., Wilke, M., Pool, J., Vermeulen, C., Blokland, E., Burghart, E., Krostina, S., Wendler, N., Passlick, B., Riethmueller, G. et al. (2002) The hematopoietic system-specific minor histocompatibility antigen HA-1 shows aberrant expression in epithelial cancer cells. *J. Exp. Med.*, **196**, 359–368.
- Goodell, M.A., Brose, K., Paradis, G., Conner, A.S. and Mulligan, R.C. (1996) Isolation and functional properties of murine hematopoietic stem cells that are replicating *in vivo*. *J. Exp. Med.*, **183**, 1797–1806.
- Bengtsson, M., Stahlberg, A., Rorsman, P. and Kubista, M. (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. *Genome Res.*, **15**, 1388–1392.
- Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D. and Eberwine, J.H. (1990) Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad. Sci. USA*, **87**, 1663–1667.
- Feldman, A.L., Costouros, N.G., Wang, E., Qian, M., Marincola, F.M., Alexander, H.R. and Libutti, S.K. (2002) Advantages of mRNA amplification for microarray analysis. *Biotechniques*, **33**, 906–912914.
- Heil, S.G., Kluijtmans, L.A., Spiegelstein, O., Finnell, R.H. and Blom, H.J. (2003) Gene-specific monitoring of T7-based RNA amplification by real-time quantitative PCR. *Biotechniques*, **35**, 502–504, 506–508.
- Li, J., Adams, L., Schwartz, S.M. and Bumgarner, R.E. (2003) RNA amplification, fidelity and reproducibility of expression profiling. *C R Biol.*, **326**, 1021–1030.
- Dixon, A.K., Richardson, P.J., Pinnock, R.D. and Lee, K. (2000) Gene-expression analysis at the single-cell level. *Trends Pharmacol. Sci.*, **21**, 65–70.
- Freeman, T.C., Lee, K. and Richardson, P.J. (1999) Analysis of gene expression in single cells. *Curr. Opin. Biotechnol.*, **10**, 579–582.
- Saghizadeh, M., Brown, D.J., Tajbakhsh, J., Chen, Z., Kenney, M.C., Farber, D.B. and Nelson, S.F. (2003) Evaluation of techniques using amplified nucleic acid probes for gene expression profiling. *Biomol. Eng.*, **20**, 97–106.
- Park, P.J., Cao, Y.A., Lee, S.Y., Kim, J.W., Chang, M.S., Hart, R. and Choi, S. (2004) Current issues for DNA microarrays: platform comparison, double linear amplification, and universal RNA reference. *J. Biotechnol.*, **112**, 225–245.
- Iwasaki, A. and Medzhitov, R. (2004) Toll-like receptor control of the adaptive immune responses. *Nat. Immunol.*, **5**, 987–995.
- Poltorak, A., He, X., Smirnova, I., Liu, M.Y., Van Huffel, C., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C. et al. (1998) Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science*, **282**, 2085–2088.
- Kurimoto, K., Yabuta, Y., Ohinata, Y., Ono, Y., Uno, K.D., Yamada, R.G., Ueda, H.R. and Saitou, M. (2006) An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. *Nucleic Acids Res.*, **34**, e42.
- Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttill, R.A., Dolle, M.E., Calder, R.B., Chisholm, G.B., Pollock, B.H., Klein, C.A. et al. (2006) Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature*, **441**, 1011–1014.
- Blake, W.J., Kaern, M., Cantor, C.R. and Collins, J.J. (2003) Noise in eukaryotic gene expression. *Nature*, **422**, 633–637.
- Levsky, J.M., Shenoy, S.M., Pezo, R.C. and Singer, R.H. (2002) Single-cell gene expression profiling. *Science*, **297**, 836–840.
- Kaern, M., Elston, T.C., Blake, W.J. and Collins, J.J. (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.*, **6**, 451–464.
- Levsky, J.M. and Singer, R.H. (2003) Gene expression and the myth of the average cell. *Trends Cell Biol.*, **13**, 4–6.