SMP: RNA

Project: Analysis of Murine Tumor Models and Human Tumors

Meinhard Hahn - German Cancer Research Center (DKFZ), Heidelberg - m.hahn@dkfz.de

Introduction

Today, researchers can choose from a broad variety of methods for global transcriptional profiling. Among the different technical approaches, microarray technology has gained a premier position. In principle, microarrays can be produced either by robotic printing ("spotting") of DNA on a chemically modified glass surface1, or by *in situ* synthesis of oligonucleotides on a silane-reacted quartz substrate2.

Spotted arrays usually contain cDNA-specific PCR amplicons (cDNA arrays), ranging from several hundred to a few thousand base pairs in size. Generally, no more than one amplicon is used to probe a given gene. Although they are technically challenging and require both optimized protocols and workflow, cDNA arrays are typically produced by individual research groups or core facilities. Alternatively, they can be purchased from several commercial suppliers. But after the discovery of frequent discrepancies in the annotation of cDNA clones, investigators began to realize potential drawbacks of this highly advocated technology. *In situ* synthesis of oligonucleotide probes requires sophisticated equipment for photolithography and solid phase chemistry, which is usually too complex and elaborate for an academic environment. A widespread commercial implementation of this technology is the Affymetrix GeneChip platform, which currently uses 11-16 pairs (11 for the arrays used in this study) of perfect-match and single-base-mismatch 25-mer oligonucleotides for each gene. Recently, large collections of longer oligonucleotides (50-80 bases), produced by established suppliers using conventional phosphoramidite chemistry, have become increasingly popular as probes for spotted DNA arrays. Technical advantages of oligonucleotide arrays include a constant DNA concentration across all spots and biophysically optimized sequences, reducing secondary structures, avoiding repetitive sequence motives and providing a fixed range for both T_m and length. All this accounts for more uniform, stable and predictable hybridization conditions. The overall costs for long oligonucleotide arrays will often be lower when labor and other costs associated with cDNA libraries, such as replication, amplification or sequence verification, are regarded.

Considering this diversity of approaches and the resulting technical differences, researchers are highly interested in the general accuracy and reliability of microarray data and the cross-platform comparability. Several independent methods like Northern blotting or real-time quantitative PCR (RQ-PCR) have been used to validate microarray results for a small number of transcripts. Generally, there was good agreement between the corresponding values, affirming the ability to accurately profile gene expression with array-based approaches. Former studies also compared global expression measurements between cDNA arrays and short oligonucleotide arrays or SSH. Recently, Barczak et al. compared results between spotted arrays of 70-mer oligonucleotides and in situ synthesized Affymetrix GeneChip arrays. Using RNA of a cell line and a commercial reference RNA, they found strong correlations of the corresponding data sets. Despite these studies clarifying some fundamental questions, there still remains considerable uncertainty regarding the comparability of data from clinical specimens. As this lack of understanding constitutes a barrier, which keeps researchers from an immense amount of potentially valuable information (via efficient integration of microarray data generated on different array platforms), we conducted a comparison with tumor samples from clinical practice, which evaluates cross-platform reproducibility in a practical setting [1,2].

Results/Project Status

cDNA- and 70mer oligonucleotide microarrays

Human cDNA-microarrays: containing replicate spots of 4211 different gene-specific fragments, representing 2600 different genes with relevance to mitosis, cell cycle control, oncogenesis, or apoptosis **[3]**.

Custom-made cDNA-microarrays: cDNA microarrays which represented approximately 7,000 distinct human gene-specific fragments, with particular enrichment for cancer-relevant genes. In addition, each microarray carried gene-specific cDNA fragments representing 215 genes located within the chromosomal regions 1p36.13-p36.31 and 19q13.2-q13.33.

Human 70mer oligonucleotide microarrays (version 2 plus upgrade): 26,791 gene-specific 70mere oligonucleotides (Human Oligo set 2.1 plus upgrade; Operon)

Human 70mer oligonucleotide microarrays (version 4.0): 35,035 oligonucleotide probes, representing approximately 25,100 unique genes and 39,600 transcripts (Human genome oligo set 4.0).

Murine 20k cDNA microarray: 20k ArrayTAG™ collection (LION Bioscience), The 20k ArrayTAGTM included 20,172 sequence-verified cDNA clones (comprising about 10,000 annotated genes) with an average insert size of 200-600 bp. All clones were proved for the absence of repetitive elements and low-complexity regions.

Murine 23k cDNA microarray (NIA): The microarray contained "unique" cDNA clones that were derived from pre- and peri-implantation embryos, E12.5 female gonad / mesonephros, and newborn ovary. Up to 50% of these clones were originated from novel genes expressed during mouse embryogenesis with an average insert size of 1.5 kb

Patient-based cross-platform comparison of oligonucleotide microarray expression profiles

The comparison of gene expression measurements obtained with different technical approaches is of substantial interest in order to clarify whether inter-platform differences may conceal biologically significant information. To address this concern, we analyzed gene expression in a set of human head and neck squamous cell carcinoma (HNSCC) samples, using both spotted oligonucleotide microarrays made from a large collection of 70-mer probes and commercial arrays produced by *in situ* synthesis of sets of multiple 25-mer oligonucleotides per gene. Expression measurements were compared for 4,425 genes represented on both platforms, which revealed strong correlations between the corresponding data sets. Of note, a global tendency towards smaller absolute ratios was observed when using the 70-mer probes. RQ-PCR measurements were conducted to verify expression ratios for a subset of genes and achieved good agreement regarding both array platforms. In conclusion, similar profiles of relative gene expression were obtained using arrays of either single 70-mer or multiple short 25-mer oligonucleotide probes per gene. Although qualitative assessments of the expression of individual genes have to be made with caution, our results indicate that the comparison of gene expression profiles generated on these platforms will help to discover disease related gene signatures in general. For a subset of genes, we verified microarray-derived expression ratios by real-time quantitative PCR (RQ-PCR) and found good qualitative agreement between the two array platforms and the PCR-based method (see Figure 1).

We have shown that, overall, expression profiles obtained with either long (Operon) or multiple short (Affymetrix) oligonucleotide microarrays display a reasonable correlation, with variable concordance of individual genes. Based on patient samples, we obtained results which are in good agreement with previous studies that utilized cell line-derived RNA. Projecting these findings to a larger series of array experiments, one could expect to obtain similar albeit not identical results, concerning, e.g., a hierarchical clustering or a gene expression signature, with either of the two investigated platforms. On the level of individual genes and quantitative precision, however, our results reaffirm that microarrays have to be considered a screening technology and that their data should be regarded with caution. This should be kept in mind particularly when comparing data from different array platforms.

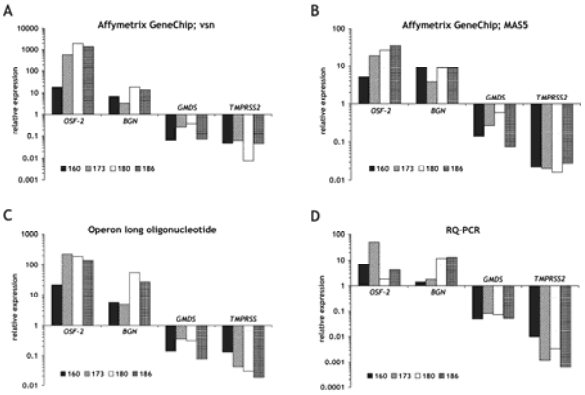

Fig 1: *Comparison of relative gene expression for the genes OSF2, GMDS, TMPRSS2 and BGN. Expression ratios were determined for tumour versus control tissue of the indicated patients, using either Affymetrix GeneChip arrays (A-B), Operon long oligonucleotide arrays (C) or real-time quantitative PCR analysis (D). Affymetrix ratios were either normalized by variance stabilization (A) or the MAS5 algorithm (B).*

Identification of novel tumour-associated genes differentially expressed in the process of squamous cell cancer development

Chemically induced mouse skin carcinogenesis represents the most extensively utilized animal model to unravel the multistage nature of tumour development and to design novel therapeutic concepts of human epithelial neoplasia. We combined this tumour model with comprehensive gene expression analysis (see Figure 2) and could identify a large set of novel tumour-associated genes that have not been associated with epithelial skin cancer development yet **[4]**. Expression data of selected genes were confirmed by semi-quantitative and quantitative RT-PCR as well as in situ hybridisation and immunofluorescence analysis on mouse tumour sections. Enhanced expression of genes identified in our screen was also demonstrated in mouse keratinocyte cell lines that form tumours in vivo. Self-organizing map clustering was performed to identify different kinetics of gene expression and co-regulation during skin cancer progression. Detailed analysis of differential expressed genes according to their functional annotation confirmed the involvement of

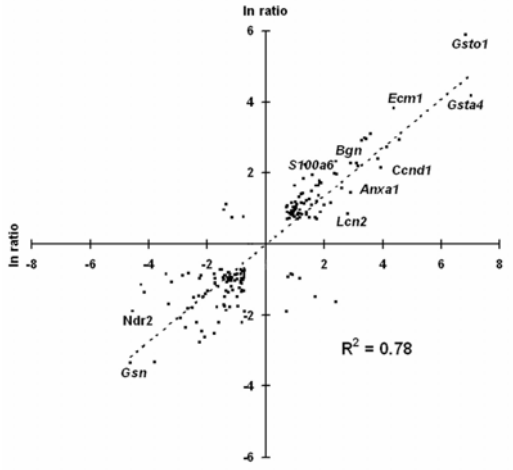

Fig 2: *Correlation between expression ratios measured on NIA and ArrayTAG microarrays. The expression ratios for 181 genes, which show significantly altered expression levels (linear ratio greater or smaller than 2.71) in SCC and are present on both chips, were plotted in a 2D graph. The x-axis represents the logarithmic (ln) ratio values of the NIA microarray, whereas the y-axis indicates the logarithmic (ln) ratio values of the ArrayTAG microarray. The square of the correlation coefficient, R2, is 0.78, which indicates a high correlation between the expression ratios obtained from both microarrays.*

several biological processes, such as regulation of cell cycle, apoptosis, extracellular proteolysis and cell adhesion, during skin malignancy. Finally, we detected high transcript levels of ANXA1, LCN2, and S100A8 as well as reduced levels for NDR2 protein in human skin tumour specimens demonstrating that tumour-associated genes identified in the chemically induced tumour model might be of great relevance for the understanding of human epithelial malignancies as well.

Outlook

We produce high-quality cDNA and 70mer oligonucleotide microarrays using standardized conditions (quality management) and will apply them for comprehensive RNA expression studies. Our focus will be both the analysis of murine model systems of medulloblastoma, squamous cell carcinoma and mantle cell lymphoma but also the analysis of defined human tumor entities, *e.g.*, of the central nervous system.

Lit.: ***1.*** *Schlingemann J et al. Patient-based cross-platform comparison of oligonucleotide microarray expression profiles. Lab Invest. 2005 Aug;85(8):1024-39.* ***2.*** *Schlingemann J et al. Effective transcriptome amplification for expression profiling on sense-oriented oligonucleotide microarrays. Nucleic Acids Res. 2005 Feb 17;33(3):e29.* ***3.*** *Wrobel G et al. Optimization of high-density cDNA-microarray protocols by 'design of experiments'. Nucleic Acids Res. 2003 Jun 15;31(12):e67. Erratum in: Nucleic Acids Res. 2003 Dec;31(23):7057.* ***4.*** *Hummerich L et al. Identification of novel tumour-associated genes differentially expressed in the process of squamous cell cancer development. Oncogene. 2005 Oct 17; [Epub ahead of print].*