SMP: RNA

Project: On-chip Polymerase Reactions for Signal Amplification and Exon Identification; Simultaneous Analysis of Transcript Level and Splice Variation

Jörg Hoheisel - **German Cancer Research Center (DKFZ), Heidelberg - j.hoheisel@dkfz.de**

Introduction

The total number of genes in the human genome was found to be much lower than predicted. Therefore, it is to be expected that regulation of cellular function at the level of RNA is occurring by variation of both the actual transcript level and the splicing of the RNA. Simultaneous analysis of both is therefore crucial to the understanding of regulation at this level. It is very likely that a relatively large number of exons were not yet identified. Even in the much simpler structured sequence of *Saccharomyces cerevisiae*, several hundred open reading frames were initially missed altogether. In a recent study on *Drosophila melanogaster*, we could identify another about 2000 transcribed genes (**Hild *et al.*, 2003**), and there are probably more. In addition, transcribed sequences that are not coding for proteins have an even higher likelihood of being missed.

Current transcript analyses – irrespective of the system used – still select the probes based on the sequence annotation. PCR-product based microarrays are generally not able to discriminate between exons; the same is true for current arrays made of long oligonucleotides. Because of high capacity and relatively low cost for each probe, *in situ* synthesised oligonucleotide arrays can circumvent this. The mainly used Affymetrix system, however, is unable to change the chip design easily. Also, chemical oligonucleotide synthesis proceeds reversely to biological synthesis. Thus, no 3'-termini are available to act as substrates for polymerases.

We use the Geniom array platform of the company *febit biotech* (Heidelberg). It allows an entirely flexible design of each individual microarray – due to light-controlled but mask-free *in situ* synthesis – and sufficient capacity with up to 48,000 oligomers per array. At the same time, we developed a synthesis chemistry that conforms to biological synthesis direction: the resulting oligomers are attached via their 5'-ends, while the 3'-termini are freely accessible. Additionally, synthesis yields are near quantitative, so that also long oligomers can be synthesised in sufficient quantity (**Beier *et al.*, 1999; 2001; 2005**).

Project Status

On this basis, combining the febit hardware (Fig. 1) and our chemistry, we started to analyse the regulation at the level of RNA, both by detecting differential transcription and differential splicing. Technically, the issues of signal intensity and identification of new exons will be pursued. This is made possible by polymerase extension reactions on the chip. Signal intensity is increased by incorporation of fluorophor-labelled nucleotides, once an RNA/cDNA-template has bound an oligomer. Due to the length of the initial oligo-nucleotides, rather high temperatures can be applied. Consequently, longer fragments of RNA (or cDNA) can be used in the hybridisation, avoiding problems caused by intra-molecular folding. In addition, both the degree of extension and the percentage of extended oligomers can be increased by a cycling reaction. Since the *febit* chip is a contained system, no evaporation is taking place. Also, since there are only about 3 µl of liquid in each channel, cycling can be performed rather quickly. By this means, we expect to increase sensitivity of chip analysis by at least two orders of magnitude. Processes that are established in principle but still require some refinement will be used for a quantification of the absolute amount of molecules present at each individual spot on the array.

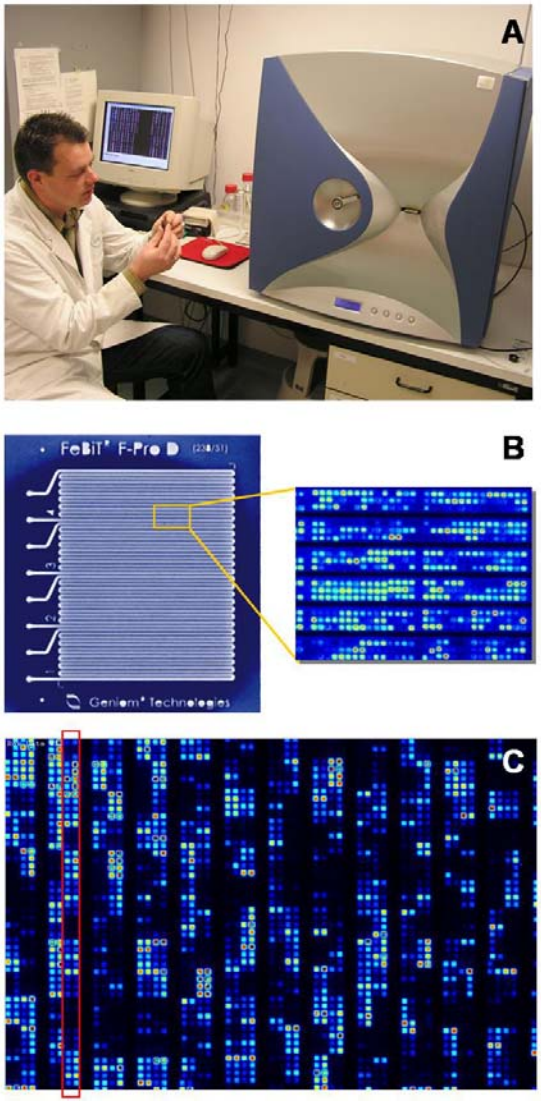

Fig 1: *In (A) a Geniom system of* febit biotech *is shown. On chips that consist of four or eight separate microfluidic channels (B) up to 48,000 oligonucleotides of independent sequence can be synthesised. The inlet and outlet are visible on the left side. On the chip presented in (C), oligonucleotides of identical sequence were synthesised side by side, with synthesis proceeding 3' to 5' or 5' to 3', respectively. Upon hybridisation of a sample, similar signal intensities were obtained.*

While falsely positively annotated exons can be identified easily by missing signals on the respective oligomers, the identification of new exons requires extra steps. Oligomer primers are synthesised on the chip, which are located close to the exon-intron borders. Upon hybridisation of the com-

plementary RNA/cDNA, a polymerase extension reaction is performed. In the hybridisation solution, there is also a mix of primers, which are complementary to the next annotated exon, respectively. On-chip PCR then yields for each RNA/cDNA-oligomer duplex a defined fragment that is attached to the support. The length of the fragment can be determined by scoring the signal intensities obtained in a secondary hybridisation with a labelled heptamer library – subsequent to the removal of the initial hybridisation target – or merely by addition of a DNA-specific dye. Compared to given standards, the signal intensities are equivalent to DNA-length, thereby indicating the presence of additional DNA, if there is an unknown exon located between two exons used for priming.

Initially, analyses will be performed on pancreas carcinoma, since a large amount of transcriptional data has already been accumulated for this tumour, resulting from PCR-product arrays. Long oligonucleotides (50-70mers) will be synthesised that represent cancer-associated genes, of which about 3,500 have been identified for pancreas and are used routinely in other projects (e.g., **Esposito *et al.*, 2004**). Data analysis will be performed with existing software tools (www.mchips.org; Fellenberg *et al*., 2001; 2002; Fellenberg & Hauser, 2001; Busold *et al.*, 2005). In terms of transcript profiles, for example, our database currently holds the results of more than 6100 individual experiments. The relevant tissues are available from ongoing collaborations with clinical partners in Heidelberg and Ulm. Simultaneously, it is planned to compare this data with epigenetic variations determined within the *SMP Epigenetic Profiling* (Mund *et al*., 2005). Also, parallel analyses will be conducted on the actual level of protein expression in a project funded within the BMBF Proteome Network.

*Lit.: **1.** Beier, M. & Hoheisel, J.D. (1999). Nukleoside derivates with photolabile protection groups. PCT/DE00/001148 **2.** Beier, M., Stephan, A. and Hoheisel, J.D. (2001). Synthesis of photolabile 5'-O-phosphor-amidites for the production of microarrays of inversely oriented oligonucleotides. Hel. Chim. Acta 84, 2089-2095. **3.** Beier, M. & Hoheisel, J.D. (2005). DNA-microarray preparation by light-controlled in situ synthesis. Curr. Protocols Nucleic Acid Chem., Unit 12.5. **4.** Busold, C., Winter, S., Hauser, N., Bauer, A., Dippon, J., Hoheisel, J.D. & Fellenberg, K. (2005). GO-annotations in correspondence analysis facilitate interpretation of microarray data. Bioinformatics 21, 2424-2429. **5.** Esposito, I., Bauer, A., Hoheisel, J.D., Kleeff, J., Friess, H., Büchler, M.W., Bergmann, F., Rieker, R.J., Otto, H.F., Klöppel, G, & Penzel, R. (2004). Microcystic tubulopapillary carcinoma of the pancreas: a new tumour entity. Virchows Arch. 444, 447-453. **6.** Fellenberg, K., Hauser, N.C., Brors, B., Neutzner, A., Hoheisel, J.D. & Vingron, M. (2001). Correspondence analysis applied to microarray data. Proc. Natl. Acad. Sci. USA 98, 10781-10786. **7.** Fellenberg, K. & Hauser, N. (2001). Microarray data warehouse allowing for inclusion of experimental and clinical data. European Patent Appl. 01123732.8. **8.** Fellenberg, K., Hauser, N.C., Brors, B., Hoheisel, J.D. & Vingron, M. (2002). Microarray data warehouse allowing for the statistical analysis of experiment annotations. Bioinformatics 18, 423-433. **9.** Hild, M., Beckmann, B., Haas, S.A., Koch, B., Solovjev, V., Busold, C., Fellenberg, K, Boutros, M., Vingron, M., Sauer, F., Hoheisel, J.D. & Paro, R. (2003). An integrated gene annotation and transcriptional profiling approach towards the full gene content of the Drosophila genome. Genome Biol. 5, R3. **10.** Mund, C., Beier, V., Bewerunge, P., Dahms, M., Lyko, F. & Hoheisel, J.D. (2005). Array-based analysis of genomic DNA methylation patterns of the tumour suppressor gene p16 promoter in colon carcinoma cell lines. Nucleic Acids Res. 33, e73.*