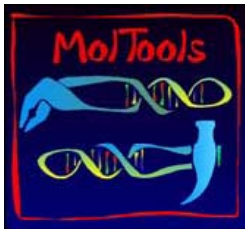


Project no: **503155**

Project acronym: **MolTools**

Project title: **Advanced molecular tools for array-based analyses of genomes, transcriptomes, proteomes and cells**

Integrated Project
Thematic Priority: LIFE Sciences, Genomics and Biotechnology for Health

Publishable final activity report

Period covered: **01/01/2006 – 30/06/2007**
Date of preparation: October 2007

Start date of project: 01/01/2004
Duration: 42 months

Project coordinator name: Professor Ulf Landegren
Project coordinator organisation name: **UPPSALA UNIVERSITET**

Partners and Group Leaders in MolTools

1. Uppsala Universitet (UU), Sweden – Represented by Professor Ulf Landegren, project coordinator
2. Karolinska Institutet (KI), Sweden - withdrawn, replaced by Partner 16
3. Royal College of Surgeons of Ireland (RCSI), Ireland
4. Centre National de Génotypage (CNG), France – Dr Ivo Gut
5. Deutsches Krebsforschungzentrum (DKFZ), Germany - Dr Joerg Hoheisel, Dr Peter Lichter, Dr Annemarie Poutska
6. Technical Research Centre (VTT), Finland – Professor Olli Kallioniemi
7. Aarhus University (UAAR), Denmark – Dr Jørn Koch
8. Max Planck Institute for Molecular Genetics (MPG), Germany – Professor Dr Hans Lehrach
9. Estonian Biocentre (EBC), Estonia – Professor Andres Metspalu
10. Oxford Gene Technologies (OGT), UK – Professor Sir Edwin Southern
11. Babraham Institute (BI), UK – Dr Michael Taussig
12. Withdrawn
13. Febit (former called Luma), Germany – Dr Marcus Beier
14. Fermentas UAB, Lithuania – Professor Arvydas Janulaitis
15. Åmic, Sweden – Dr Ove Öhman
16. University of Leicester (ULEICS), UK – Professor Anthony Brookes
17. University College Dublin (UCD), Ireland – Professor Dolores Cahill

Project web site: www.moltools.org

1. Project execution

The EU/FP6 Integrated Project MolTools was conceived as a means of linking some of Europe´s leading research groups active in molecular methods development, in order to bring about new efficient techniques for molecular analysis. Together we have aimed to develop a next generation toolbox that will make it affordable and feasible to analyse the genome, transcriptome and proteome of individuals and individual cells. Recording of the complete human genome sequences has provided for the first time the opportunity to characterise the flow of information from genetic variation at the DNA level, over messages expressed as RNA, and on to their protein products, and to the functions of the cell. However, enormous resources are required to analyse comprehensively molecular variation between humans, or between different somatic cells within one individual, in order to capture the full biomedical benefits of the sequence of the human genome. During the course of this 3.5 year project, a series of inter-related research problems has been attacked in collaboration between partners in academia and biotechnology companies. The partners have provided the complementary expertise required to establish an individualised genome analysis technology tool set, through a set of scientific workpackages dedicated to different technologies.

MolTools had a budget of 12.6 million €, of which 9 million € was contributed by the EU. The project was coordinated by Uppsala University and included 12 academic partners and four SMEs (listed on page 2).

General Project objectives

Analysis of genetic variation (Workpackages 1 and 2): Developing, evolving and applying methods of DNA analysis for DNA sequencing, genotyping, haplotyping and elucidation of duplications. During the first two years of the project, WP1 dealt with genome resequencing on microarrays and the analysis of previously unknown genetic variation, while WP2 developed and improved methods for the detection of known single nucleotide polymorphisms (SNP genotyping). In year 3, and following the mid-term review comments, it was felt appropriate to combine the efforts of WP1 and WP2 into a single workpackage (WP1/2), in order to reflect more demanding kinds of DNA analyses. Five areas of activity were defined: template preparation and multiplexing; reagent (oligonucleotide) production; support media; methods for readout/detection; and applications in sequencing and genotyping.

Contractors: 1 (UU), 4 (CNG), 5 (DKFZ), 8 (MPG), 9 (EBC), 16 (ULEICS), 13 (Febit)

Advanced array-based transcriptome analyses (Workpackage 3): Establishing methods for high-precision, high-throughput gene and low cost gene expression profiling. The ability to monitor the activity of a genome in action by measuring expression levels of transcripts provides important biological insights. Clinical studies of cancer and cardiovascular diseases demonstrate that gene expression profiles of large gene sets can identify molecular profiles correlated to diseases states, which can then be developed into diagnostic tools. Such tools are likely to enter

routine clinical laboratories in the near future. However, standard microarray technologies show inherent limitations and drawbacks in terms of cost and sensitivity. Very often the molecular characterisation of clinical samples is complicated and limited due to the available amount of samples. Three major strategies to overcome these problems were followed: development of alternative probing and amplification strategies for highly specific, parallel analyses using padlock probes and rolling-circle replication (RCA); development of protocols for signal amplification after probe hybridisation, either using RCA or via specific marker molecules of a branched structure; and a miniaturised expression profiling system based on nanowell technology.

Contractors: 1 (UU), 5 (DKFZ, Hoheisel and Lichter), 8 (MPG).

***Protein microarrays (Workpackage 4)*:** Analysing proteins and protein interactions at greatly increased sensitivity and specificity. The aim was to establish technologies for highly specific protein analysis in a manner that could be read out on microarrays. The technologies included the following. Firstly, antibody microarrays and procedures for their use; processes for selection of specific binding molecules and methods for the determination of their specificity, including cell-based array screening; and new methods for creating protein arrays *in situ* from DNA or printed from DNA template arrays, using cell free expression systems. A second focus was on new protein detection procedures based on proximity probes and mass spectrometry, respectively.

Contractors: 1 (UU), 4 (CNG), 5 (DKFZ Hoheisel), 6 (VTT), 8 (MPG), 11 (BI), 17 (UCD).

Single cell, single molecule analyses (Workpackage 5): Enabling analyses of individual cells for mutations in individual genes and for gene and protein expression. The aim was to develop a family of techniques for ultrasensitive analyses of proteins and nucleic acids in the complexity of biological samples. Procedures were developed for detection of single DNA, RNA and protein molecules, and the interaction between sets of molecules *in situ*. A unique set of probes including padlock and proximity probes was developed to react in arrayed samples and cells, and then trigger rolling circle replication reactions to allow visualisation of single molecules. These approaches permit individual cells to be assayed for multiple specific molecules at a molecular and spatial resolution enabling discrimination of single nucleotide variation and protein colocalisation within defined cellular domains.

Contractors: 1 (UU), 4 (CNG), 7 (UAAR), 10 (OGT), 14 (Fermentas), 15 (Åmic)

Functional cell microarray technologies (Workpackage 6): Using cell arrays to test functional roles of large numbers of genes and gene variants in high-throughput cell assays. This WP was to demonstrate the utility of the developed technologies in functional screening of gene knockdown / knock-in effects in various cell models. A high-throughput RNAi screen was established to reveal synthetic lethal interactions between genes that have been found to be deregulated or mutated in cancer. For gene knockdowns in the reverse transfection format of cell array, further optimisation and comparison of technologies (siRNA, esiRNA, shRNA) were performed. Novel, array-compatible apoptosis detection methods, with automatic image analysis, were

developed and validated. To enable multidimensional screening modes (gene/gene or gene/compound interactions), the consortium would manufacture permanent cell libraries expressing key gene targets or RNAi constructs.

Contractors: 5 (DKFZ Poustka), 6 (VTT), 8 (MPG),

Workpackage 7 was responsible for organisation of training courses.

Work performed and end results

WP1 and 2: Analysis of genetic variation

During the first two years of the project, **WP1** dealt with genome resequencing on microarrays and the analysis of previously unknown genetic variation. It included the establishment of software and molecular biology procedures for parallel amplification of selected parts of genomes as well as a nanowell based analysis platform, improvements of the oligonucleotide fingerprinting approach, mass spectrometry readout, and investigations of ligase-based approaches to DNA sequencing. Our third-party associate partner George Church published in *Science* a ligase-based resequencing approach applied in parallel on millions of PCR-clones attached to beads and proved the concept by sequencing an *E. coli* strain. **WP2** dealt with technological developments to facilitate, increase throughput, and decrease the cost of SNP genotyping. The notion emerged that copy-number variation is an important source of genetic variation, and efforts in WP2 gradually shifted to developing tools for DNA copy-number measurements as well. In year 3, it was felt appropriate to combine the efforts of WP1 and WP2 into a single work package (WP1/2), in order to reflect the more demanding kinds of DNA analyses. Five areas of activity were defined: template preparation and multiplexing; reagent (oligonucleotide) production; support media; methods for readout/detection; and applications in sequencing and genotyping.

Efforts to provide more effective template preparation were planned along three lines: selector probe technology, Megaplex, and on-array PCR. (UU, ULEICS, DKFZ, Febit.). For selector probe technology, specific parts of the genome are captured and cloned into a specific vector system (Figure 1). The universal vector insert sequence is amplified with a PCR using universal primers and conditions. In collaboration with Stanford University, UU applied the selector method to generate sequencing templates for parallel sequencing reactions. It was used to sequence 177 exons of 10 cancer genes, amplified and sequenced in multiplex. UU further developed a multiplex targeted copy number variation (CNV) assay useful for diagnostics, based on the selector method.

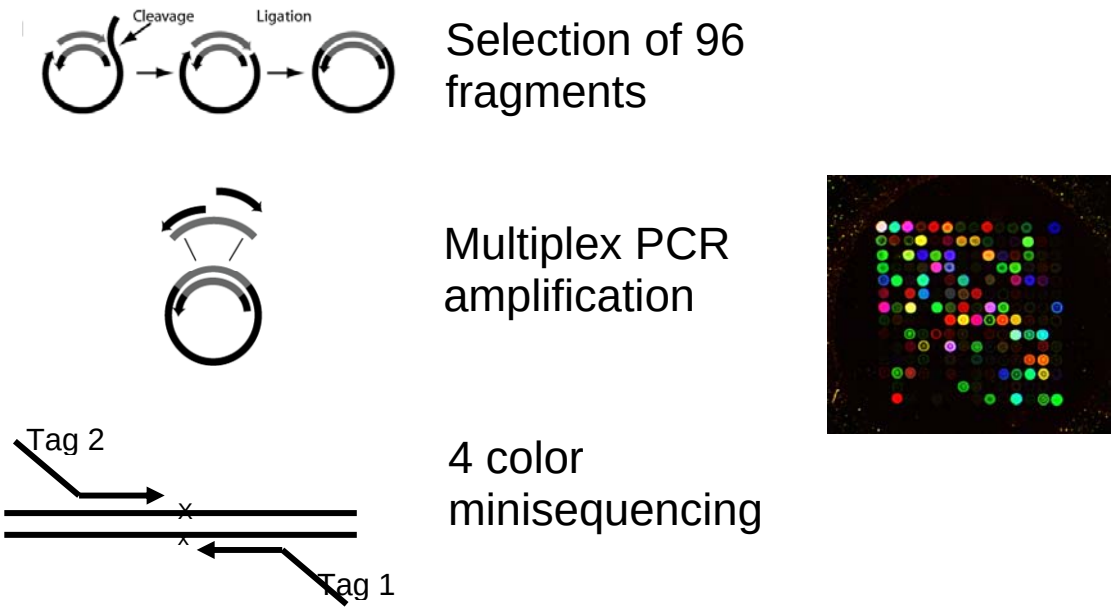


Figure 1: SNP genotyping with Selector probes (UU, Landegren)

Megaplex uses surface-bound primers for the initial cycles of a PCR, followed by PCR primed with universal primers (Figure 2). Primer pairs were spotted or synthesised with the Febir hardware in each spot or in neighbouring spots of an array. Upon hybridisation of genomic DNA, total RNA or cDNA, PCR was performed on the chip. Since the various primer pairs are separated physically, no competition between reactions takes place. To date, amplifications of up to 500 bp. are performed routinely. Products of >1 kb have also been obtained from less than 0.1 fmol DNA. After amplification, the molecules can be cleaved from the chip surface and eluted.

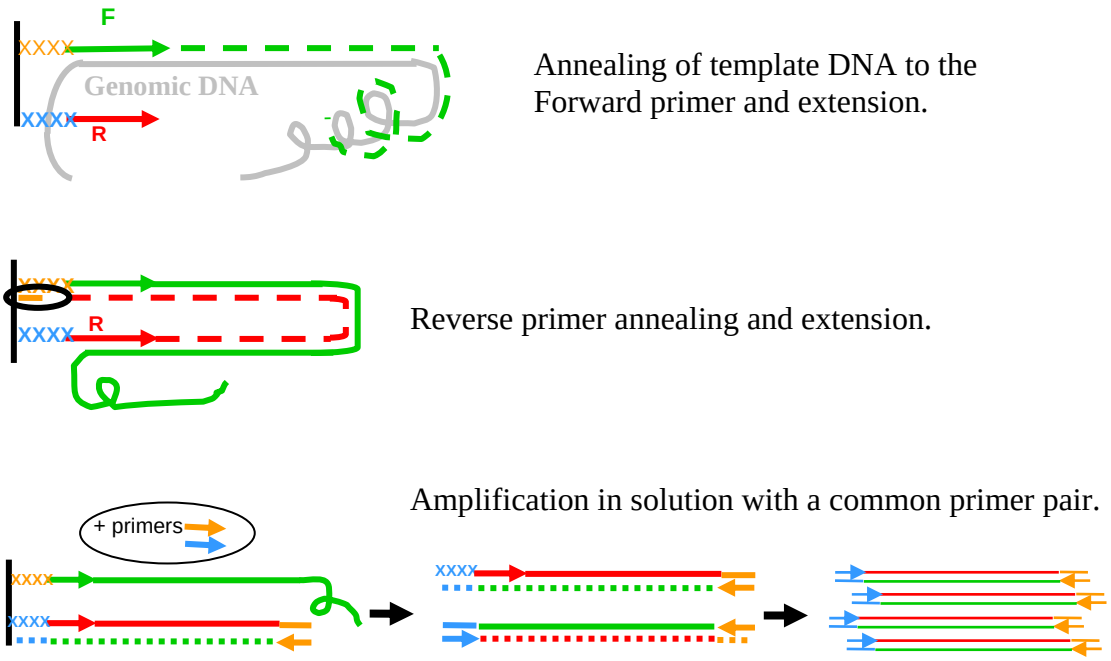


Figure 2: Mega-plex PCR

The preparation of oligonucleotides is still carried on individually and oligonucleotides remains one of the most expensive components in DNA analysis methods. In the Febit approach, large numbers of oligonucleotides are synthesized in parallel on an array surface. For this, array-prepared oligonucleotides are generated to support on-array PCR, on-array primer extension, and preparation of selectors. (Febit, DKFZ, EBC)

Several concepts of interrogation of DNA exist. Generally speaking these are a combination of a method such as hybridisation, ligation or primer extension with either direct readout or readout of a tag corresponding to the information to be captured. Oligonucleotide fingerprinting is a hybridisation-based interrogation. Here it was combined with a multiplex mass spectrometric (MS) readout using trityl-tagged oligonucleotides and with an optical readout with iFRET. Ligation was used with two readouts, first cycled ligation of a single 3'-tagged oligonucleotide followed by reading of the fluorescent tag and release, and secondly concatenation of many tag-identifiable oligonucleotides with a multiplexed readout of the tags by MS. On-array primer extension was combined with on-array PCR using an integrated approach where all nucleic acid components were synthesized *in situ*. The application of L-DNA capture was also combined with *in situ* synthesised L-DNA oligonucleotides. Microhaplotypes were analysed using a procedure where pools of extension primers were used and the readout deconvoluted by MS. (UU, MPI, DKFZ, CNG.)

With respect to SNP typing, methodology was advanced by DKFZ (Hoheisel), in collaboration with EBC and Febit, by a protocol for arrayed primer extension with dideoxynucleotide triphosphates. Febit developed an array system that allows the fully flexible design of oligonucleotide arrays by on-site *in situ* synthesis by means of a micromirror system and microfluidic platform. By combining this expertise and hardware with chemistry that permits a near quantitative, light-controlled oligonucleotide synthesis in the 5'-3' direction and thus produces chip-bound oligonucleotides that act as substrates for polymerase reactions, protocols for the analysis of very many selected polymorphisms were established. EBC developed specific software to identify suitable SNPs for APEX using the Febit Geniom system, allowing a user-defined number of SNPs to be selected from the human genome for certain chromosomes, certain chromosome regions, and around certain genes. MPG introduced a new microslide device for sample preparation, purification and MALDI MS detection, which was successfully established and applied for genotyping SNPs. The technique was patented and published in two articles.

A new approach to using spotted arrays for SNP genotyping involves the APEX-2 method (Figure 3). This has great potential to become a diagnostic and a research tool, enabling highly reproducible multiplex locus amplification in a single tube (up to 750-plex), with array based detection. It requires only ~250 ng of genomic DNA per patient. The multiplex PCR uses two 45-mer primers with two parts (SNP specific and universal) for each studied SNP/mutation. Thereafter PCR with universal primers is carried out to increase the amount of amplicons for visualisation and final detection on an array where the same primers have been spotted. Basically only one nucleotide of the unknown sequence (the SNP) ia amplified. APEX-2 primer extension and amplification is carried out in a single tube using multiplex reaction. 800-plex reactions have been performed, enormously reducing the cost. Primers bind genomic DNA just one nucleotide before the SNP or mutation and the product is further amplified with common primers. Amplification in liquid phase and final detection on

solid surface demands altogether two oligonucleotides per position - the same APEX-2 primers are used in amplification and in detection step using 4 color single-base extension (PCT/EE2007/000003). (EBC).

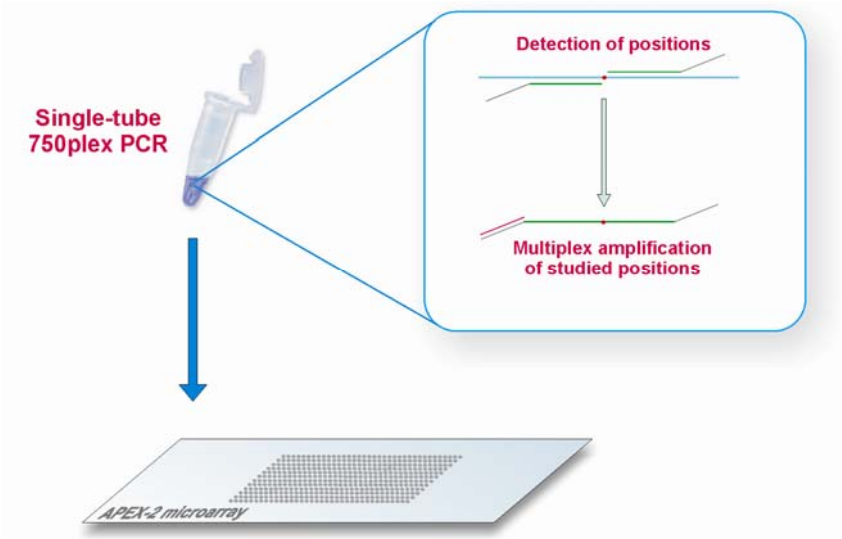


Figure 3: APEX-2 multiplex PCR and array-based detection (EBC, Metspalu)

ULEICS optimised conditions for running dynamic allele-specific hybridisation analyses in a probe-bound format on microarrays (Array-DASH). Extensive chemistry development was successfully undertaken to establish (i) optimised PCR to get single-stranded products routinely suitable for Array-DASH, (ii) the effects of PCR target length on the quality of Array-DASH signals, and (iii) alternative buffer and labelling options. These parameters were optimised, and a report passed to Febit for testing on their array platform. A second objective, extending Array-DASH towards a totally generic resequencing platform entailing custom array synthesis, was achieved in the format of membrane arrays and microplate wells, with an emphasis on diagnostics applications. An objective of inventing and validating a proof-of-principle mode for encapsulation of random array nanospaces in which sequencing can be performed, was very successful. Various microfabricated structures have been shown to be able to function as miniature ‘microplates’ with reaction volumes as small as 15 pL (Figure 4). This technology is currently being filed for patent protection. Method combinations with utility for the analysis of copy number variation (CNV) were explored and efforts centred on digital counting via single molecule analyses. The ideas are now being both multiplexed and transferred to the nanospace array systems.

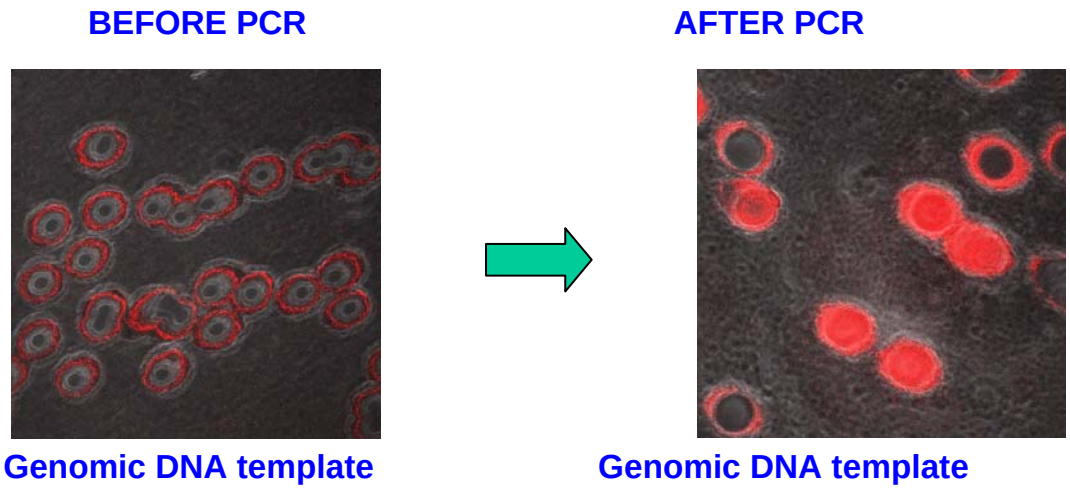


Figure 4: Single-Molecule PCR in Nanowells (Brookes, ULEICS)

Working toward the application of microarrays as a tool in routine applications, DKFZ (Hoheisel group) modified the approach of ZIP-code arrays by using L-DNA, which is the perfect mirror-image form of the naturally occurring D-conformation of DNA. L-DNA duplexes have the same chemical and physical characteristics but form a left-helical double-helix. Because of its chiral difference, L-DNA does not bind to its naturally occurring counterpart at all. DKFZ took advantage of the characteristics of L-DNA for the establishment of a universal microarray that permits the analysis of different kinds of diagnostic information at different molecular levels in a single experiment and a single platform, in various combinations (Figure 5). Apart from obtaining information from various molecular levels, the dimensionality problem of microarray experiments is markedly reduced, leading to superior predictions.

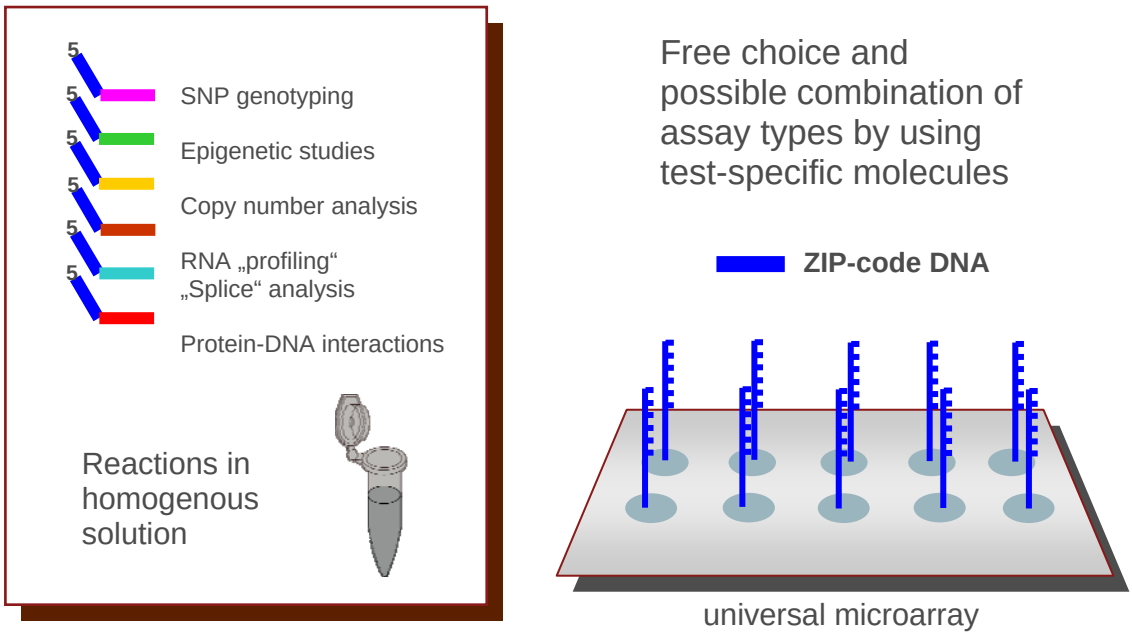


Figure 5: Universal ZIP-code Platform (Hoheisel, DKFZ)

WP3: *Advanced array-based transcriptome analyses*

In the course of WP3 new techniques and protocols for precise quantification of transcript monitoring of very few copies were developed and current limitations and bottlenecks of gene expression profiling were overcome. The strategies outlined in the objectives were followed successfully. Taken together, the newly established methods represent a portfolio of techniques for high-precision gene expression profiling at a high-throughput level obtained from limited sample material. The proof of concept and first implementation experiments provided in course of the project are the starting point for their routine use in the future.

Four complementary approaches were explored, two based on rolling circle amplification (RCA) in general, one using branched structures for signal amplification, and one miniaturisation of existing approaches. Rolling circle replication was used in two approaches, either primary amplification with the introduction of a sequence tag and subsequent use of tag-arrays (UU), or based on gene-specific arrays and secondary RCA amplification (DKFZ/Lichter). All tasks have their advantages and disadvantages and as in other technological fields, the approaches altogether will provide a package that supports different applications. As they were developed by combined effort, the power of the various approaches could be tested and compared.

UU was engaged in development of mRNA detection using padlock probes and RCA. A dual tag microarray platformfor expression profiling at the level of mRNA was developed (Figure 6). Homogenous-phase padlock probe ligation is followed by RCA of the circularised probes. After subsequent monomerisation of the RCA products, the fragments are re-circularised by enzymatic ligation, templated by specific tag oligonucleotides immobilised in an array format and used to prime a secondary RCA on the arrays. Compared to regular array hybridisation, this process results in greatly increased sensitivity, specificity, and improved sequence discrimination and dynamic range of detection. The technique provides highly sensitive detection to single molecule level (Figure 7) and ensures optimal specificity at the levels of target recognition and throughout the microarray readout procedure.

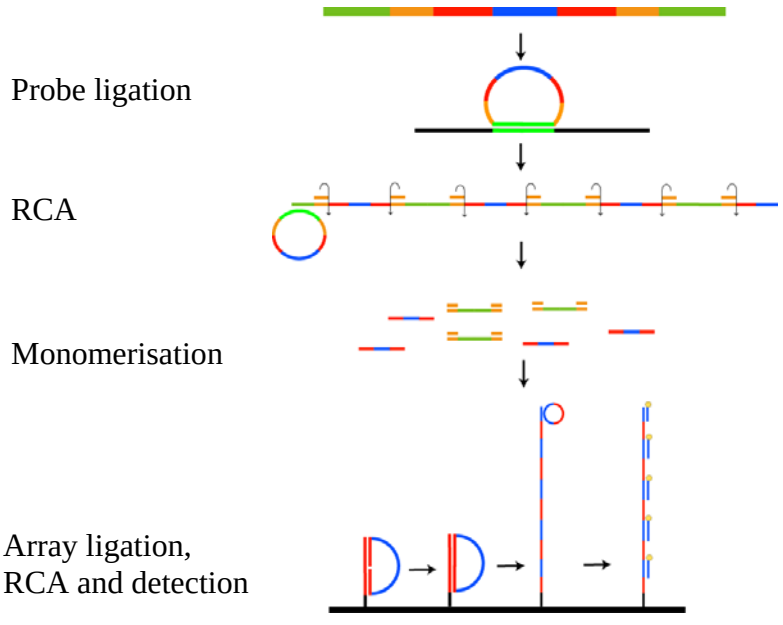


Figure 6: Padlock mRNA detection scheme (Landegren, UU)

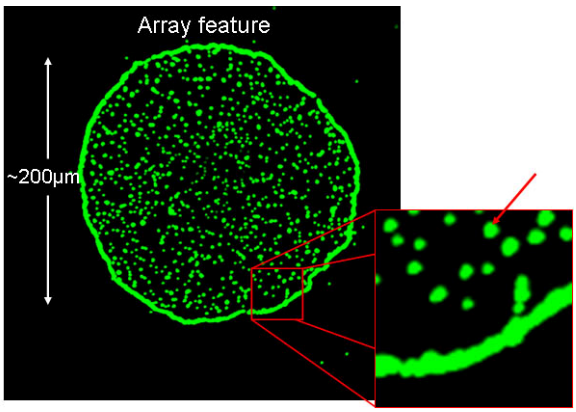


Figure 7: Single molecule detection: digital quantification of array RCA events for increased precision. (Landegren, UU)

The so-called TAcKLE technique developed by the Lichter group (DKFZ) generates amplified, antisense-oriented fluorescent representations of initial mRNA for the sensitive parallel detection of transcripts on oligonucleotide arrays. TAcKLE is successfully used in different academic and clinical research groups for clinical and tumour genetic projects. The Lichter group also developed a new robust signal amplification protocol after array hybridisation, based on prior target hybridisation and secondary amplification by RCA of preformed DNA circles, which was tested for RNA samples from human cell lines in comparison with TAcKLE. Gene-specific 70-mer DNA microarrays comprising 27,000 probes were used. Several systematic improvements were required, e.g., new protocol for generation of circular probes, immobilisation of RCA products via covalent cross-linking of modified duplexes of cDNA / 70mer probes on the array surface, control of mean length of RCA products, further chemical odifications of primers and probes, etc. All relevant steps of this multi-step protocol were successfully integrated into a robust signal amplification protocol. Proof of principle of data generation by this procedure of RNA expression profiling was obtained. However, the TAcKLE protocol showed a higher sensitivity, being a specific and robust protocol for precise routine applications.

A third approach to transcription profiling used target amplification by self-assembling marker molecules of branched structure (Hoheisel, DKFZ). Here, the sensitivity of on-chip detection was increased using a secondary probe hybridisation and avoiding direct labelling of the sample. The Y-shaped DNA-molecule clusters upon initial binding to the microarray by a self-assembly process. Sensitivities of less <1 zM were achieved.

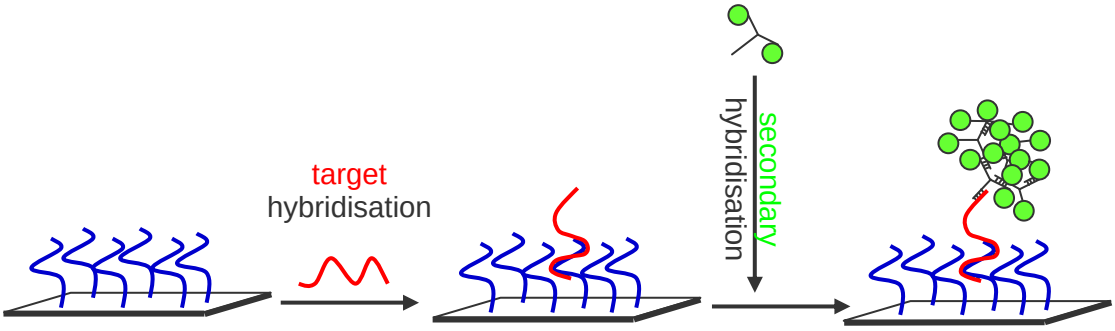


Figure 8: Target amplification using branched self-assembling marker molecules (DKFZ, Hoheisel).

The Lehrach group (MPG) set-up Nanowell Technology (NWT) for miniaturised gene expression analysis for, in principle, existing detection methods such as TaqMan and SybrGreen (Fig. 9). The platform consists of a nanodispensing station for filling nanowell plates using a 96-channel nanodispensing system STACCATO with 20 nl droplet resolution. The nanodispensing workstation is interfaced with a plate handling and storage module for automated processing in large-scale formats. A high sensitivity camera reads out the fluorescent signal in synchrony with the cycling system. Instrumentation software to control the cycling system and camera was developed. TaqMan, SybrGreen, CR-real-Time PCR assays as well as RCA monitored in real-time have been tested for amplification in miniaturised volumes. All real-time PCR based assays are equally suitable for expression analysis on the nanoliter scale and proof of principle experiments were performed with a cardiovascular gene subset.

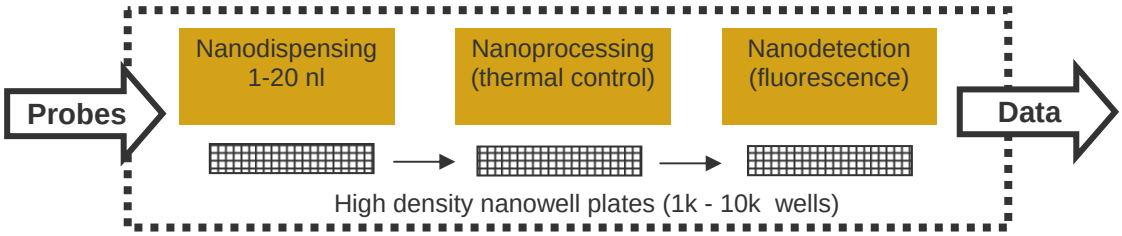


Figure 9: Nanowell technology for gene expression analysis (Lehrach, MPG)

WP4 *Protein microarrays*

The major projects and overall deliverables were:

- Establishment of complex antibody microarrays containing highly specific binders, for quantitative analysis of complex protein samples.
- Improved protein detection by utilisation of proximity probes
- Quantification of variation in protein expression by combining microarrays and mass spectrometry
- Identification of transcription factor binding sites
- Selection of disease-relevant marker proteins from body fluids
- Functional array-based screening of protein-protein interactions

- Functional array-based screening of antibodies
- New technologies expressing proteins directly onto array surfaces from DNA by cell free expression.

At DKFZ, Hoheisel's group developed an array of about 1000 antibodies and applied it to the quantitative analysis of complex protein extracts from cancer and normal tissues, in part within other EU-funded projects (e.g., MolDiagPaca).

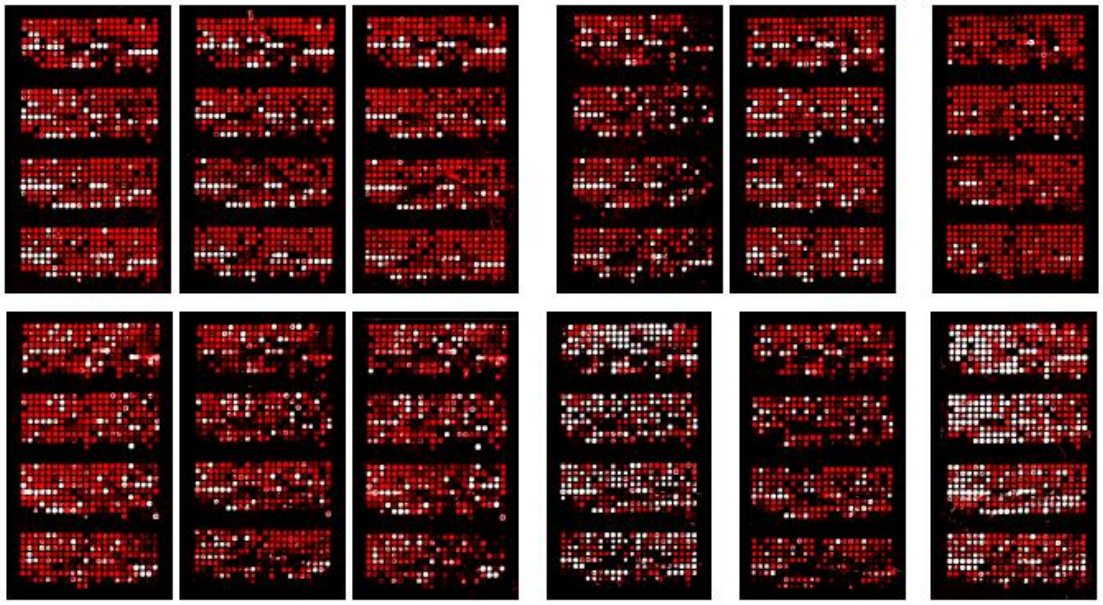


Figure 10: Signal patterns obtained on an antibody microarray of about 700 molecules upon incubation with a fluorescently labelled sample of serum proteins from different patients.

The group established processes for a reliable analysis of differences in actual protein expression by complex antibody microarrays. Next to issues such as surface coating, reproducible protein isolation and labelling, various aspects of kinetics and mass transport were addressed, which are essential to achieve reliable and reproducible results. The group established extensive collaborations with several companies, e.g. Eurogentec (Belgium), Advalytix (Germany), *Kreatec* (Netherlands), BASF (Germany) and Merck (Germany) on issues related but not covered by the MolTools project.

The Landegren group (UU) focused on developing highly specific protein analyses using the proximity ligation mechanism, and scaling to smaller amounts and larger numbers of proteins and other targets. Both solution phase and solid phase versions of the assays have been established, and widely published, and conditions defined that allow the assays to be performed in various biological samples with minimal effects on nonspecific background compared to assays performed in buffer. This is of great importance in enabling analyses of serum and plasma samples, as well as of cell lysates. The triple-binder mechanism (Figure 11) enhances specificity and sensitivity, and permits analyses of complex constellations of interacting proteins and of post-translational modifications. Improved means of producing the protein-DNA conjugates used as detection reagents in the proximity ligation assay were also established. Several other new variants of the basic assay mechanism were developed to further enhance assay sensitivity and specificity.

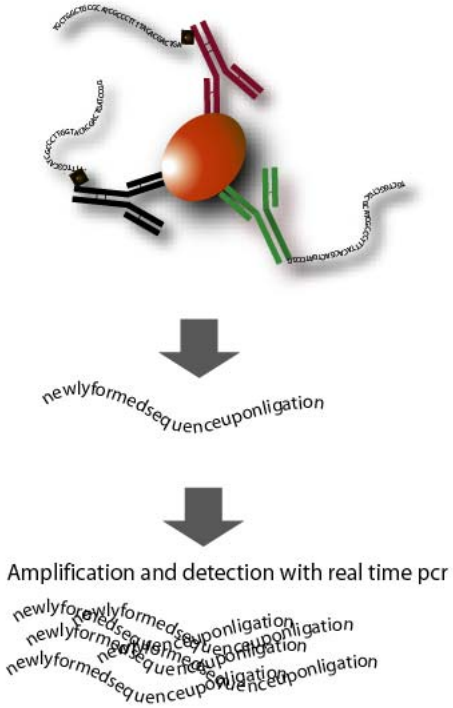


Figure 11: Triple Binder Proximity Ligation Protein Analysis via DNA (Schallmeiner, UU)

UU also established a so-called 'paired-tag' microarray, which potentially will be of great value for parallel protein analyses, and for monitoring interactions among large sets of proteins. Besides this technology development work, UU applied the methods for specific, demanding analyses, including measurement of protofibrils of Aβ in early Alzheimer's disease, and detection of an expanded repertoire of viral agents. The SME 'Olink' was spun out from the Uppsala group on the basis of work done during MolTools and earlier, in order to commercialise proximity ligation technology. Contracts for collaborations have been successfully negotiated with both large European and US companies, and Olink has launched its first commercial product.

CNG developed a protein profiling method for target proteins on an antibody array using a mass spectrometry (MS) readout, termed Affinity Arrays MALDI Mass Spectrometry (A2M2S). The underlying concept was that different samples would be treated with tags that convey mass differences. After mixing, the samples are applied to the array. By adapting procedures commonly used in MS-based protein analysis such as tryptic digestion, proteins are conditioned for MS and analysed. The peak heights or areas are used to determine the abundance of the target protein in each of the initial samples. Five different labels that perform sufficiently well were devised, all of which can be used to tag proteins directly on a crude serum or plasma sample. The sensitivity range of the method is surprisingly wide, with quantification over six orders of magnitude of protein expression in native samples. The sensitivity would be sufficient to detect down to the range of the most abundant tissue leakage proteins such as prostate specific antigen (PSA). The concept is shown below:

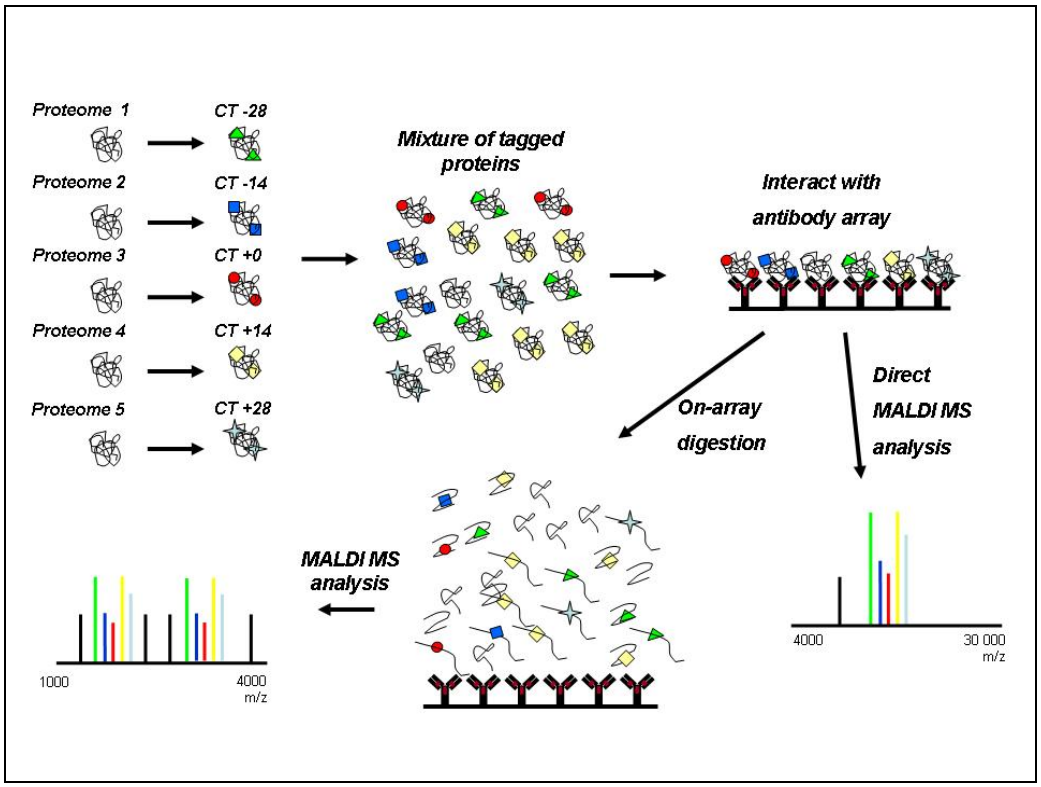


Figure 12: Principle of Affinity Arrays MALDI Mass Spectrometry (Gut, CNG)

DKFZ and BI established processes that permit the *in situ* production on microarrays of proteins from DNA templates, using either full-length cDNAs or artificially produced molecules (Figure 13). They developed the production of protein microarrays from full-length cDNA libraries. For cell-free expression, the cDNA is amplified with primer pairs harbouring promoter, ribosomal binding site and terminator sequences. The PCR-products are spotted onto a microarray. Proteins are expressed by an *in situ* cell free transcription and translation extract on the same slide. An array carrying several hundred proteins has been produced and published (DKFZ). BI showed that the PISA (protein arrays *in situ*) method can be miniaturised and carried out on a slide, similar to that of DKFZ.

In a further development, BI designed another system (DAPA, or DNA Array to Protein Array) by which DNA arrays can be converted into pure protein arrays. In DAPA, a DNA slide is sandwiched with a protein-capture slide, with an interposed membrane carrying the cell free transcription/translation lysate. DAPA was shown to work well for a number of different proteins and at least 20 identical replicate arrays could be produced from a single DNA array. This development will make further cell free protein interaction studies possible. An array of transcription factor DNA-binding domains was produced for future interaction studies.

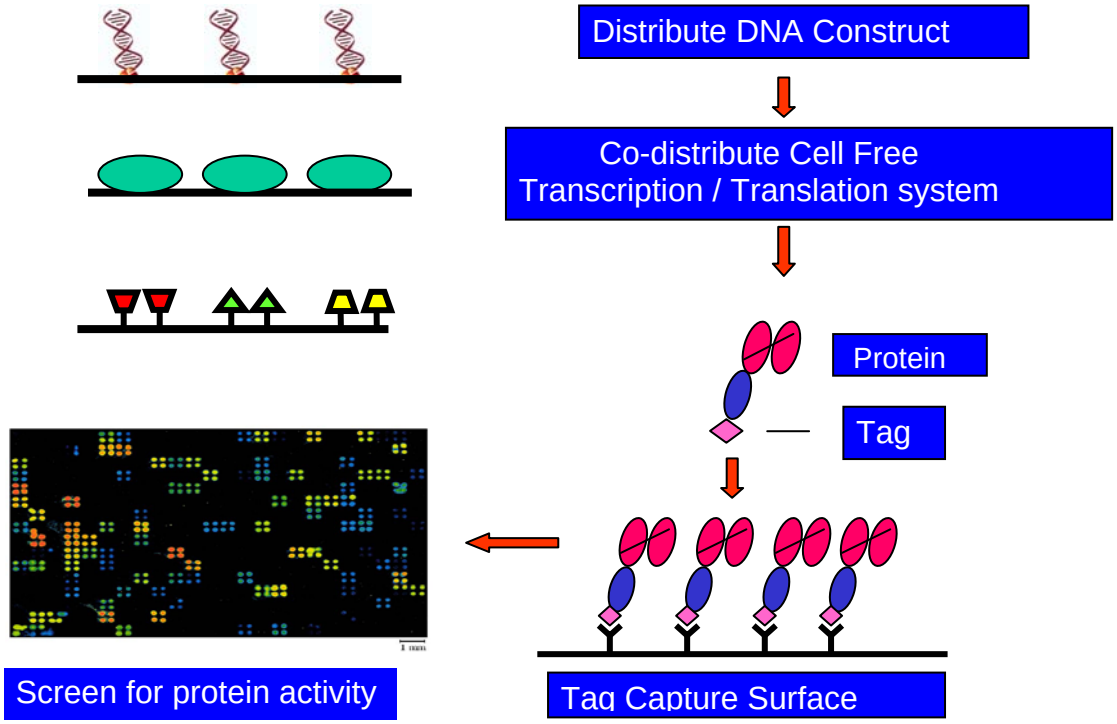


Figure 13: Protein in situ arrays High Density Protein Microarrays by Cell-free in situ Expression (Hoheisel DKFZ, Taussig BI).

UCD provided purified proteins to other partners and developed reliable and versatile nondenatured protein microarrays using a new directional immobilisation strategy that has allowed efficient and reproducible binding of proteins while keeping their overall structure, under one standard mild buffer condition. They also developed a new label-free detection technology to measure K_{on} and K_{off} rates of antibody binding to their immobilised protein targets in a microarray format. Using imaging Surface Plasmon Resonance, this is able to measure accurately and reproducibly kinetic parameters for hundreds of proteins in parallel. One of the major improvements and applications of this technology is to rank label-free binders based on their kinetic parameters in an array format.

WP5: *Single cell, single molecule analyses*

WP5 was characterised by establishing novel methods rather than doing them faster, bigger or cheaper. Throughout the project the results obtained not only met but exceeded expectations. Within the first year, a collaborative paper in *Nature Methods* from UU and UAAR reported the establishment of padlock / rolling circle technology that successfully genotyped individual mitochondria for single nucleotide variation *in situ*. Two cell lines, homoplastic for a single nucleotide variant in position 3243 of the mitochondrial genome, were used. In Figure 14, red spots represent amplification products from single padlock probes having detected mitochondra with an A at this position, the wild type sequence, while green spots signal a G in the same position.

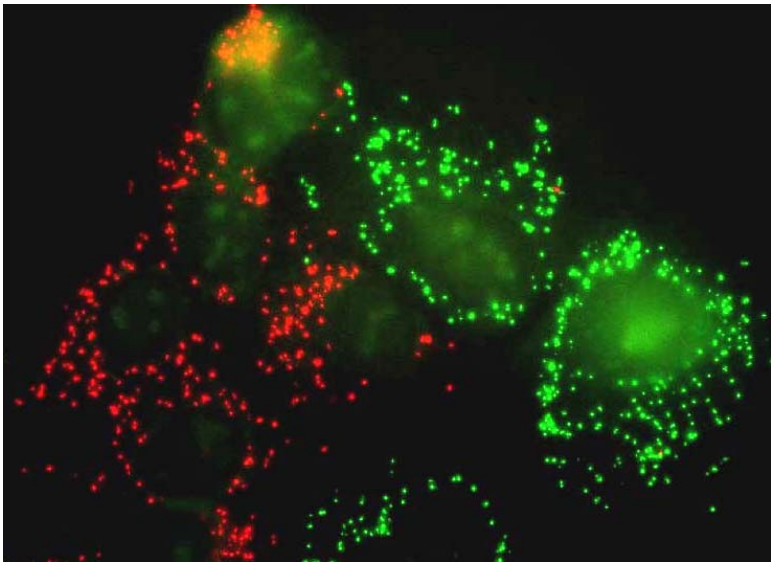


Figure 14. *In situ* genotyping of single-nucleotide variation in single mitochondrial genomes, using a combination of padlock probes and rolling circle amplification. (Larsson et al. (2004) Nature Meth., 1. 227-232).

Fermentas provided enzymatic expertise to facilitate and improve the molecular strategies, testing and optimising the quintessential Phi29 DNA polymerase, providing enzyme variants and test results in addition to a stable supply of enzyme at an affordable price.

In addition, Fermentas provided a method for RCA using padlock probes which can be circularised on the RNA target sequence in any position of RNA molecule, but which does not require an additional DNA primer for RCA initiation. The novel detection method was successfully applied in individual transcript detection *in vitro* and *in situ*. Based on our results we have filed patent application covering this technology.

Åmic developed and delivered plastic microstructures, which have proven useful for a number of subsequent applications at UU. These include (a) designs for proximity probe ligation (cultivation of cells exposed to gradients of different reagents and capillary driven microfluidics for ultra-sensitive reading of proteins in serum) and (b) the platform used by UU for studying the presence of single molecules by padlock or proximity probing, increasing the signal by RCA for detection in a LIF microscope, where the design provides fluidic means to ensure that the microscope securely sees the amplified single-molecules in the field of view.

UU established the proximity ligation mechanism for protein detection, patented and published the technology and established a new SME based on it (Olink AB). Proteins and protein-protein interactions have been detected in cultured cells, frozen tissue sections and paraffin embedded formalin fixed tissue sections with the *in situ* proximity ligation assay. Protocols were developed for multiplexed detection of interacting protein pairs and protein DNA/RNA interactions. *In situ* padlock probing assays have been developed for detection of individual molecules. Detection of larger protein complexes (more than 2 interacting protein) has also been achieved and a

generalized protocol using secondary proximity reagents has been established. An example of protein detection *in situ* is shown in Figure 15.

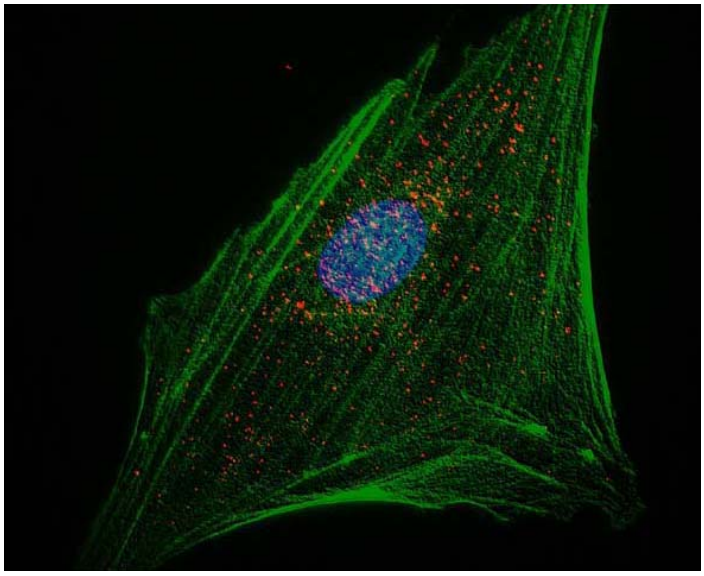


Figure 15: c-Myc/Max interactions in an *in vitro* cultured human fibroblast by proximity ligation *in situ* (red dots). Endogenous protein interactions were detected with single-molecule resolution. The cell was counterstained with FITC-conjugated anti-actin mAb labeling the actin filaments (green), and the nucleus visualised by Hoechst staining of the DNA (blue).

UAAR developed and filed patent applications on two new formats of circle probes for RNA analysis *in situ* (so-called Turtle Probes and Slicer Probes). In parallel, an enzymatic synthesis system for the production of Padlock, Turtle, and Slicer Probes with improved performance, primarily stemming from improved ligation efficiency of the probes, was developed and a patent application filed. This amplification system is applicable to oligonucleotide synthesis in general and was suggested by the C2C-system derived by UU. Based on the presumably upcoming patents, as well as older patents on target primed rolling circle reactions (Jörn Koch), UAAR has now established a new SME – In Situ RCP A/S – to exploit the technology. Qualitative and semi-quantitative detection of RNA targets in model systems has been achieved and RNA processing in individual cells demonstrated in the form of human telomerase (hTR) RNA accumulation in Cajal Bodies of cancer cells (Figure 16). Working enzyme combinations for the detection of single RNA targets have been derived. A novel rolling circle assay for DNA processing enzymes in individual cells has been developed to expand the range of coverage from the detection of molecules to the functions and activities of the molecules. Two patent applications were filed on this new assay (May and July 2007) and are being transferred to In Situ RCP.

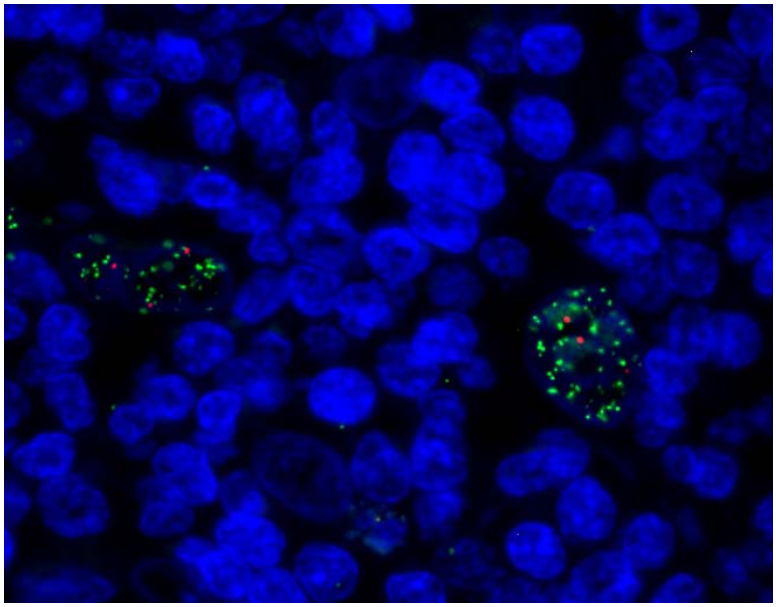


Figure 16: RNA detection in Formalin-Fixed Paraffin-Embedded (FFPE) human tissue from a patient with Epstein-Barr Virus (EBV)-associated Hodgkin's lymphoma. EBER1 RNA appears in the neoplastic Reed-Sternberg cells and not in the surrounding lymphocytes. Combined detection of EBER1 and hTR with the probes TP-EBER1-id33 (green) and TPhTR-id16 (red). The counterstain is DAPI, producing blue cell nuclei.

OGT have been pushing array-analysis to the level of single-molecule counting within array spots, creating and building a high-speed high-resolution optical fluorescence microscope with automatic image stitching for large fields of view for the detection of single fluorescent molecules (Cy3-labeled dNTP) (Figure 17). OGT has also supported the work of CNG, developing and providing photo-cleavable tags that are easily detected in a MALDI mass spectrometer. CNG has tested molecular procedures involving combined ligation, hybridisation and readout of a tag attached to the hybridised oligonucleotide, and imaging (scanning) of a slide in the MALDI mass spectrometer. This concept would allow a far higher degree of multiplexing than conventional immunohistochemical analysis using optical detection, as the number of tags resolvable by mass spectrometry is far higher than the number of fluorescent tags that can be resolved by fluorescence detection.

The integration of the efforts in WP5 has worked smoothly, both within academia and between academia and the SMEs, which were particularly well represented in this WP (three of six partners). Whereas the SMEs present in WP5 may also assist in the exploitation of the produced IP, that dimension of the consortium was further ensured by one of the academic partners (UU) starting an SME early on (Olink AB), and by UAAR starting another in 2006 (In Situ RCP A/S). To further enhance the integration process, all partners gathered at UAAR for a halfway meeting (August 2005), and a follow-up meeting at Fermentas in Lithuania in September 2006. Several follow-up projects involving various combinations of WP5 partners, as well as partners from the other WPs are expected.

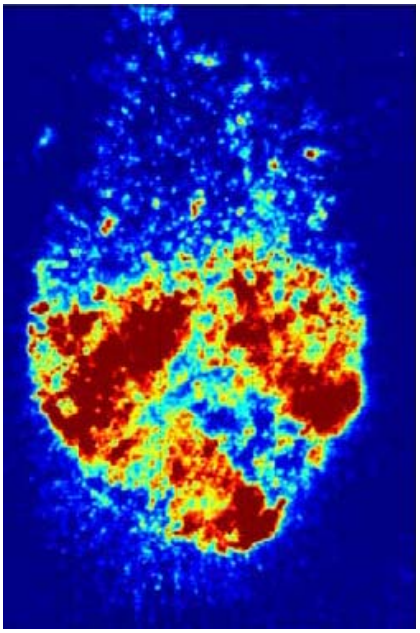


Figure 17. An image of the total mRNA content of an individual cell. Cells were spread on a slide that had been derivatized with dT30. After lysis, mRNA was captured to the oligo dT by hybridisation and reverse transcribed, incorporating Cy-labeled dCTP. (OGT, Southern).

WP6: *Functional cell microarray technologies*

The WP aimed at the functional interrogation of individual gene activities in living cells in a highly parallel fashion. The aims was to develop and test high-throughput platforms for the introduction of cDNA expression vectors and RNAi constructs into cells, as well as recording the cell-biological consequences of overexpression / silencing of the individual genes in various cell models. In line with the aspirations of MolTools, the ultimate goal was to extend this parallel analysis capability to a whole genome scale. Objectives were broken down into the following technological categories:

- *Platform development* involved both plate- and microarray-based transfection methods, and the utilisation of compatible vector-based or chemically / enzymatically prepared nucleic acid constructs for gene over-expression/silencing.
- *Multiplexing and synthetic lethality* objectives encompassed technical solutions for the combinatorial overexpression and/or silencing of two or more genes, or for studying gene/compound interactions in a high troughput mode.
- *Image-based cell phenotyping* comprised the application of computer image analysis methods to microsope/scanner images of reverse transfection cell arrays in order to characterise the transfected cells for cell survival, cell cycle phase, or other phenotypic traits, as well as subcellular localisation of reporters.
- As an important complement to the most often used transient transfection methods, the WP also aimed at the construction of *standardised, stable cell lines with inducible gene knockdowns/knock-ins* for the controlled, tunable regulation of individual target genes.
- A common theme in the WP was to apply the high throughput platforms for the *identification of genes which have a role in promoting/preventing the induction of*

apoptosis, as these genes have recurrently been shown to be involved in the initiation and progression of cancer.

Plate-based transfection platforms for high throughput functional cell-based screening were developed and used by DKFZ/Lichter (shRNA and cDNA transfection) and VTT (siRNA). VTT integrated the transfections into a general robotic cell-based screening platform with automated reverse transfection, compound addition, cell culture and endpoint assay, having a capacity of over 10,000 assays per day. Reverse transfection cell arrays were developed by partners DKFZ/Lichter (cDNA, shRNA), VTT (siRNA), and MPG (cDNA, siRNA). VTT developed an improved cell array technology in which the cells are in distinct, separate spots instead of a continuous lawn.

For multiplexing and synthetic lethality, MPG developed a multiplexed siRNA transfection method in cell array format and used it to reveal a previously unknown interplay of particular genes in apoptosis induction. VTT used a synthetic lethal approach involving isogenic cell lines (below) and siRNA screening to identify genes that are particularly effectively targeted in cells which over/underexpress other genes, such as oncogenes or tumor suppressor genes. Isogenic cell line technology for the overexpression or silencing of single genes was developed by partners DKFZ/Mollenhauer and DKFZ/Lichter.

Two novel approaches were used for apoptosis detection in reverse transfection cell arrays using image-based cell phenotyping. DKFZ/Lichter used overexpression of the EYFP protein, which made the apoptotic bodies easily visible by fluorescence microscopy. Software was developed for automated recognition of apoptotic cells. MPG developed transfected cell arrays for high-throughput evaluation of subcellular protein localisation, protein-protein interactions and measurement of promoter activities. In addition, siRNA-based cell arrays were produced and applied for apoptosis studies. In all, over 1,000 different cDNA constructs and siRNAs were spotted for reverse transfection. A new apoptosis detection method was established based on a combination of three existing assays (TUNEL, caspase 3 and Annexin V), achieving a a dramatic enhancement of apoptotic signal. Using this approach over 400 different genes previously suggested to be involved in apoptosis were screened, and by multiplexing some of those siRNAs, novel interplay of particular genes in apoptosis induction was demonstrated.

DKFZ/Mollenhauer developed a technology to create cell line pairs (and libraries) differing in the expression level of a single gene at a time. Moreover, expression/silencing are inducible by tetracycline, making the system highly controllable and tunable for accurate measurements of graded responses to individual gene expression levels. The system was also used to create isogenic cell lines pairs for synthetic lethal screens for VTT. Another stable cell line system was developed by DKFZ/Lichter as an addition to the work programme.

VTT developed and implemented a novel, HTS cell spot microarray platform that will significantly enhance RNAi technology applications (Figure 18). The array scale - up to 30,000 siRNAs per microplate area - makes it possible to carry out genome-scale functional analysis of genes in living cells. This invention provides up to 100-fold, enhancement of screening capacity and a corresponding reduction of reagent

consumption compared to existing 384-well plate based screening. While previous technologies have been available for this purpose, the scale, ease of imaging, practical implementation and possibilities for entirely new assays and applications are advantages of the technology. For example, cell array technology with distinct spots is particularly suitable for automated recognition and analysis. We optimised conditions for over 20 cell lines (including breast, ovarian, prostate, melanoma and lung cancer) using an siRNA library against 1183 cancer-related genes (2 siRNAs per gene). Fifteen different assays (antibody or dye staining) were set up and optimised for cell array.

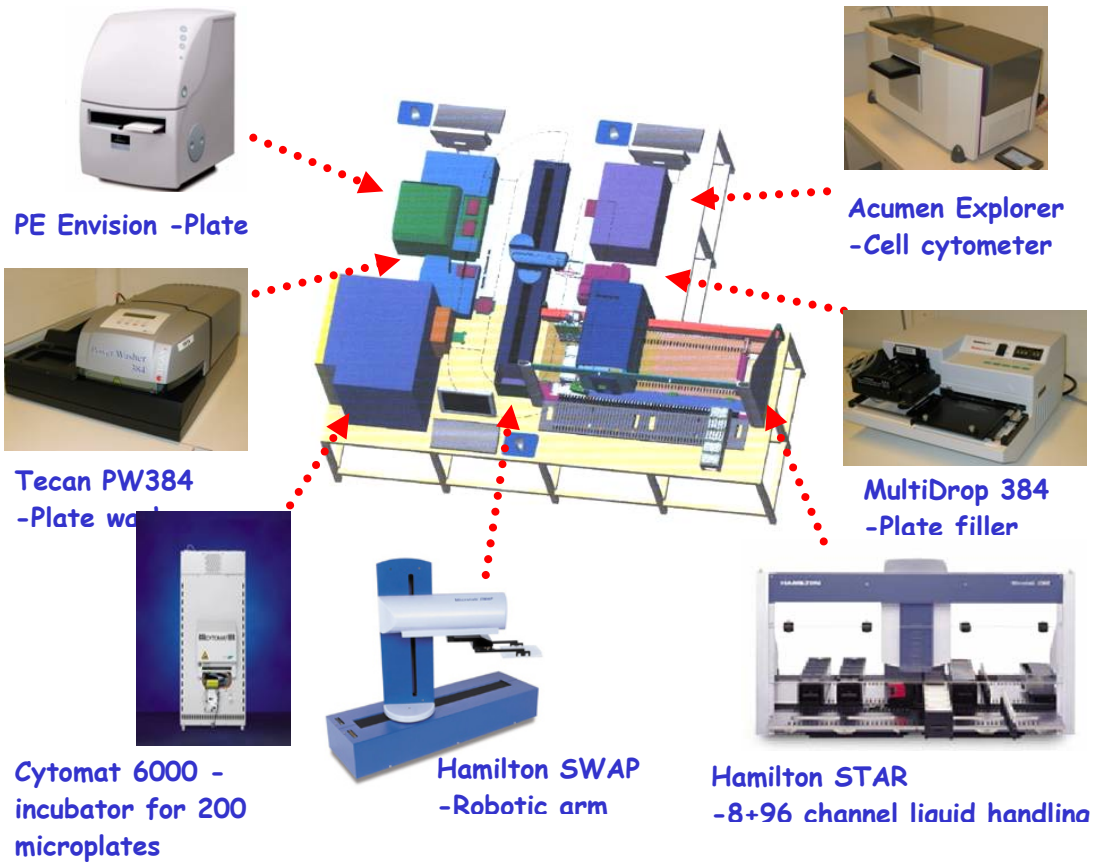


Figure 18 High throughput screening of 50,000 RNAis and compounds (VTT)

WP7

This was concerned with the organisation of training activities. Details of the 9 training events organised over the 3.5 years of the programme can be found in the projects Final plan for using and disseminating the knowledge.

Impact of the Moltools project

The overarching goal of the project has been to enable scientists in academia, healthcare and industry to follow the flow of genomic information from its repositories in genomes of individuals and cells, over expressed transcripts and proteins, to its manifestations in cellular phenotypes. Among the technology developments which Moltools partners have been particularly concerned with and which will continue for the forseeable future are high throughput arrays for analysis of genome variation, genome resequencing, protein detection and cellular analysis, and nanoscale methods and devices to study the interactions and dynamics of single molecules. A whole range of entities and material can now be analysed in the array format, as reflected in the MolTools work programme: DNA arrays for genotyping and expression analysis at the RNA level, antibody arrays for expression analysis at the protein level, other protein arrays for protein interaction analysis, and cell arrays enabling protein function to be analysed in an *in vivo* setting. The following describes the impact of activities in the project; the references are all papers published by MolTools partners during the course of the project.

DNA- and RNA-level analyses

New techniques for DNA sequence analysis have vastly accelerated sequence acquisition. Improved methods of recording and typing nucleotide sequences with respect to SNPs and other sequence variants (1, 2*)*, copy number variation (3), CpG methylation (4), or miRNA levels (5) that could explain individual phenotypic variation, particularly that underlying disease, were all important elements in MolTools (WPs1-3). The new parallel DNA sequencing methods (6) also create a need for procedures to prepare abridged editions of genomes for priority sequencing, by selectively isolating exons, promoters, or other segments of special interest e.g. in disease causation. Selector probes (7) and MegaPlex amplification (Brookes et al, submitted) are tools to single out many parts of genomes for amplification to useful amounts, while avoiding cross-reactivity problems that normally arise in multiplex PCR.

Another goal pursued along several lines of research was to adapt DNA microarray technology for extremely sensitive, parallel analyses of all classes of biomolecules (9). Tag sequence oligonucleotides, arranged in a microarray format, have been constructed with the L-form of DNA, allowing these mirror-image strands to hybridise exclusively to the complementary L-form sequences included in probes, but not to natural DNA in other parts of the probes or in the samples, enhancing performance compared to standard tag arrays (10). Arrays of nanoliter-size reaction wells represent yet another means of combining specificity and wide dynamic range with parallel analyses by engineering large numbers of real-time amplification reaction volumes (11,12) (Brookes et al in prepn). A new array reader with the sensitivity and throughput to scan for individual fluorophores in array features promises to greatly impact assay sensitivity and dynamic range in RNA expression analyses (Milner patent pending, Southern unpub.). In another array approach to single-molecule detection, reacted DNA probes are specifically circularised via DNA ligation reactions on arrays so that they can be locally amplified via rolling-circle replication, allowing individual amplified molecules to be easily detected, thereby ensuring high detection specificity and sensitivity (Landegren et al. in preparation).

Improved protein detection

As developed in WP4, proximity ligation enables highly sensitive and specific protein detection, by using pairs of antibodies with attached DNA strands that form amplifiable sequences upon target protein recognition and oligonucleotide ligation (13). By extending the assays to require binding by sets of three binding agents, as little as a few hundred molecules of a target protein can be detected in cell lysates and serum samples (14). The need for parallel analyses of many different target molecules, or conversely, using many different binding agents, is met by approaches involving protein or antibody arrays for analysis of protein expression or protein interactions (15, 16). The heterogeneity and relative instability of proteins make the construction and use of immobilised protein microarrays particularly demanding. Rather than simply spotting proteins onto a surface, protein arrays can now advantageously be produced by depositing DNA molecules that are transcribed and translated *in situ* (17,18), and even printed repeatedly on new surfaces from one DNA template array (He et al in preparation). The work in MolTools towards improved protein assays complements effectively the EU-FP6 sponsored project ProteomeBinders (www.proteomebinders.org), which plans to establish a comprehensive resource of binding agents against all human proteins, including splice variants and modifications, for proteome-wide analysis (19).

Pushing detection to the level of single biomolecules

A new generation of methods has been developed that can identify and locate individual macromolecules and their interacting partners directly in cells for a fuller understanding of the subcellular modules that execute cellular functions. WP5 was designed to meet the need for *in situ* detection methods with the requisite sensitivity to analyze single nucleotide variants of individual DNA (20) or RNA molecules (Janulaitis et al, in prepn), or individual and interacting protein molecules (21) in cells and tissues. In these approaches, specific detection by circularisable padlock and proximity DNA probes, followed seamlessly by a localized amplification reaction, permits single-molecule detection. Each anchored spool of amplified DNA can be easily observed against any background arising from nonspecifically bound detection reactions and autofluorescence *in situ*, and they are also suitable for digital recording of individual amplified probes traveling in a microfluidic device (22). In future, these procedures will be extended to simultaneous analysis of large sets of molecules, allowing ther relative levels and positions to be ascertained. One interesting approach for monitoring numerous distinct reaction products *in situ* involves mass spectrometers as powerful and accurate multiplex detectors. By scanning tissue sections for reagents modified with mass tags using targeted multiplex mass spectrometry imaging (TAMSIM), several different proteins detected by affinity probes can be imaged in histological sections (23).

Analyzing the function of genes using arrays of cells

WP6 has exploited the ability of cells to integrate internal and external molecular processes and signals to a phenotypic output. Using cancer as a model disease, series of gain- or loss-of-function cell lines have been prepared and arrayed for parallel analyses of the effects on their responses to external agents by defined genetic alterations (24). The detected responses can be integrated with simultaneous or sequential measurements of levels of transcripts and proteins for a fuller understanding of cellular reactions. In a proof-of-principle experiment a set of 13

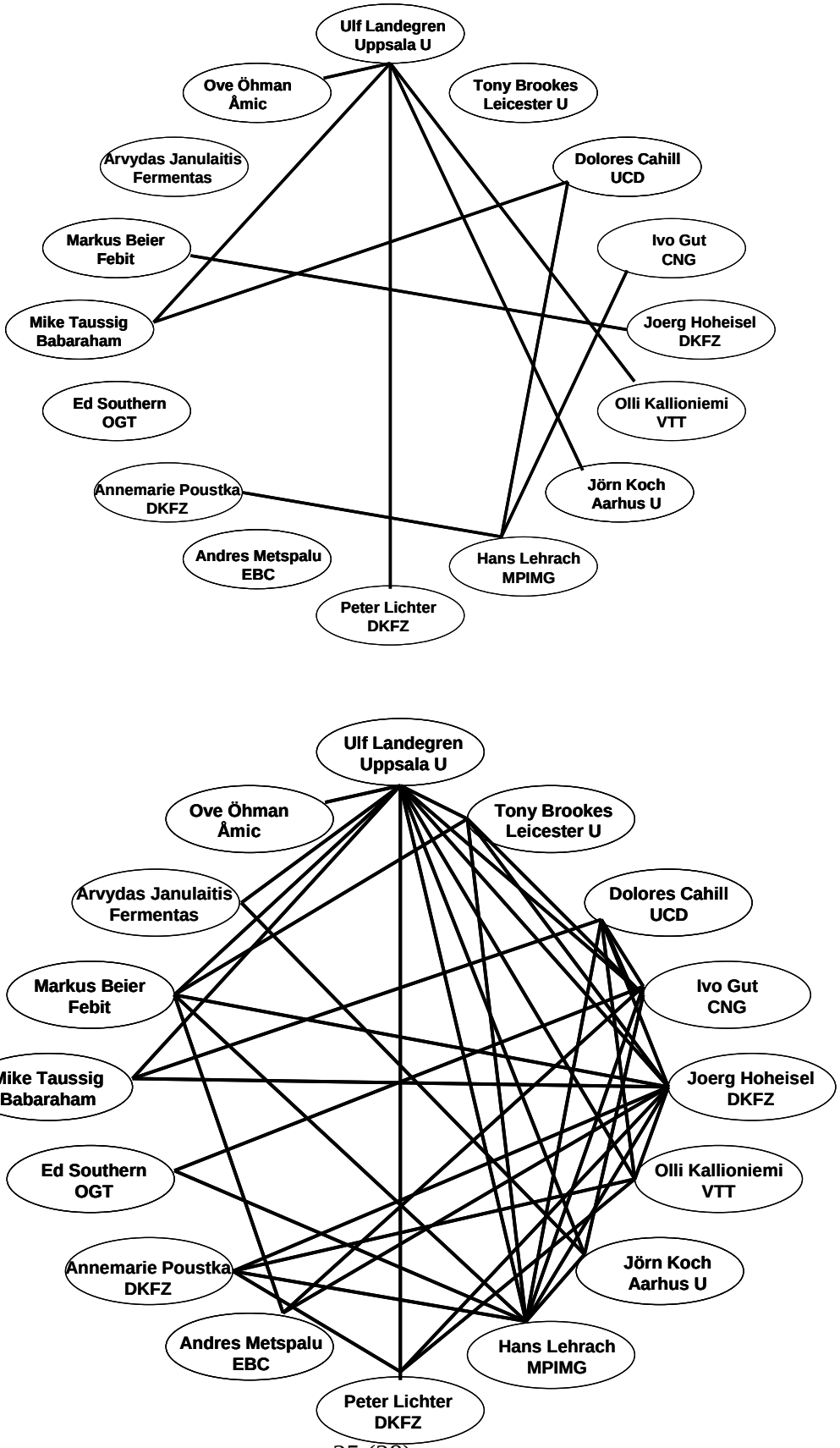
novel genes that affect tumor growth and/or progression was recovered on a genome-wide scale (25). Cell array techniques provide ideal substrates for systems level biological investigations, synthetic lethal screens, drug target discovery and validation, drug discovery, as well as for construction of novel reporter systems.

The MolTools project in context

The project has seen a dramatic increase in interactions between partners (Figure 19 below). Apart from advancing science and developing new techniques, we were also involved in the dissemination of the results. This ranged from discussions with students at schools, colleges and universities *via* training courses for graduate students and post-docs to scientific workshops and conferences that were organised as part of MolTools. Spin-out companies have been formed around the technologies, including one (Olink) to commercialise the proximity ligation technology (UU), another (In Situ RCP A/S) for exploiting the IPR generated by the Aarhus group, and a third which will develop *in situ* protein arrays and ribosome display technology (Discerna, BI). The SME partners have also honed their products in collaboration with other partners.

The projects underway in MolTools were also coordinated with other research efforts at the European level, including the integration of human sample collections – biobanks, a perceived strength of European science – to determine how these can best be explored using comprehensive reagent sets and advanced molecular assays. In combination with technology standards like MIAME and methods databases under development, this will enhance the use of sample collections and support metastudies combining results from separate investigations. Moreover, activities such as EMERALD – an action to promote the development of standards and quality metrics for microarray technology – or ProteomeBinders – a network of groups that is aiming at the establishment of reliable ligand binding molecules against the human proteome (19) – were an immediate consequence of MolTools. MolTools has provided technologies with improved precision and throughput for systems biology, allowing individual cells and individual molecules to be investigated without confounding effects of tissue heterogeneity to permit ultimately computation of biological processes. Finally, the techniques pursued in MolTools may also find application in diagnostic analyses in order to promote a molecular definition of disease and detecting diseases while still at treatable stages. Collaborative projects such as MolTools will continue to link academic and commercial interests in FP7, combining the participants with users and customers as well as the relevant regulatory bodies to accelerate development and uptake of new technologies.

Figure 19: **The three years in MolTools have triggered many new research collaborations, as illustrated by identifying interactions at project onset (upper panel) compared to the status at the close (lower panel). Each line represents a collaboration that has or is expected to result in one or more joint publication.**



References

1. Sauer, S., Reinhardt, R., Lehrach, H. & Gut, I.G. *Nature Protocols* **1**, 1661-1671 (2006).
2. Mauger, F. et al. *Nucleic Acids Res.* **in press** (2007).
3. Fredman, D. et al. *Nat. Genet.* **36**, 861-6 (2004).
4. Pfister, S. et al. *Nucleic Acids Res.* (2007).
5. Jonstrup, S.P., Koch, J. & Kjems, J. *Rna* **12**, 1747-52 (2006).
6. Shendure, J. et al. *Science* **309**, 1728-32 (2005).
7. Dahl, F., Gullberg, M., Stenberg, J., Landegren, U. & Nilsson, M. *Nucleic Acids Res.* **33**, e71 (2005).
8. Tian, J. et al. *Nature* **432**, 1050-4 (2004).
9. Hoheisel, J.D. *Nat. Rev. Genet.* **7**, 200-10 (2006).
10. Hauser, N.C. et al. *Nucleic Acids Res.* **34**, 5101-11 (2006).
11. Sauer, S. et al. *Nat. Rev. Genet.* **6**, 465-76 (2005).
12. Dahl, A. et al. *Biomed. Microdevices* (2007).
13. Gullberg, M. et al. *Proc. Natl. Acad. Sci. U S A* **101**, 8420-4 (2004).
14. Schallmeiner, E. et al. *Nat. Methods* **4**, 135-7 (2007).
15. Kusnezow, W. et al. *Mol. Cell. Proteomics* **5**, 1681-96 (2006).
16. Kusnezow, W. et al. *Proteomics* **6**, 794-803 (2006).
17. He, M. & Taussig, M.J. *Nat. Methods* **4**, 281-8 (2007).
18. Angenendt, P., Kreutzberger, J., Glokler, J. & Hoheisel, J.D. *Mol. Cell. Proteomics* **5**, 1658-66 (2006).
19. Taussig, M.J. et al. *Nat. Methods* **4**, 13-7 (2007).
20. Larsson, C. et al. *Nat. Methods* **1**, 227-32 (2004).
21. Soderberg, O. et al. *Nat. Methods* **3**, 995-1000 (2006).
22. Jarvius, J. et al. *Nat. Methods* **3**, 725-7 (2006).
23. Thiery, G. et al. *Rapid Commun. Mass. Spectrom.* **21**, 823-829 (2007).
24. Pscherer, A. et al. *Faseb J.* **20**, 1188-90 (2006).
25. Mannherz, O., Mertens, D., Hahn, M. & Lichter, P. *Genomics* **87**, 665-72 (2006).

2. Dissemination and use

The EU FP6 project MolTools was designed to establish key tools for research and diagnostic analyses of DNA, RNA, proteins and cells, by coordinating a wide range of expertise in some of Europe's leading groups in academic or industrial technology development. The success of the MolTools project has had important consequences by providing a wide range of valuable new research tools in academia and industry, a prerequisite for progress in all biomedicine. Moreover, the output of the research programme has already given rise to important applications in industry and will continue to do so. A few of these are summarised below.

MolTools partners rapidly seized upon the need for isolating all sequences of special interest from genomic DNA samples for priority sequencing using emerging DNA sequencing techniques. Between them, they have pioneered and patented several of the major approaches considered by DNA sequencing companies, including selector probes, MegaPlex PCR and padlock probes. Similarly, techniques invented by the partners and perfected during MolTools have proven valuable for measuring and distinguishing gene sequences, as exemplified by the padlock probes licensed by the Uppsala group to the company Affymetrix for their targeted genotyping platform.

Regarding medical applications of nucleic acid technologies, the TAcKLE RNA amplification protocol is used in clinical breast cancer studies and aPRIMES is used in genetic studies of medulloblastoma (partner 5, DKFZ). Furthermore, the nanowell real-time PCR platform (partner 8, MPG) will be used for miniaturized gene expression profiling in research settings in particular for clinical samples available at very low amount and the need of a flexible assay set-up.

Numerous contributions were also made to the central problem of measuring proteins in biological specimens with ever-higher sensitivity and precision. The *in situ* protein arrays established by partners 11 (BI) and 5 (DKFZ) provide an efficient means of constructing planar arrays of many different proteins for parallel analyses of their properties. Proximity ligation (partner 1, UU) represents a new and quite general test mechanism that offers unprecedented sensitivity for protein measurements in patient samples. Some aspects of the technology have been licensed to Applied Biosystems, while others will be commercialized by the start-up Olink, spun out by the partner 1 during MolTools.

Another very important trend where MolTools has made contributions that will prove important in biomedicine concerns the measurement of even single nucleic acid or protein molecules in individual cells. A proximity ligation based strategy to detect interacting proteins in patient samples was established and published during MolTools, and is now commercially available from Olink. In a related fashion, partner 10 (OGT) developed technology allowing single molecules to be imaged on arrays.

As a final example, several partners of MolTools contributed to establishing technologies that permit parallel analyses of many cell populations modified to influence their susceptibility to drug treatment. Partner 6 (VTT) has used their

miniature cell arrays to successfully identify potential targets for drug therapy in cancer, of great interest to the drug industry.

The groups brought together in this programme are experienced inventors of tools for molecular analysis. An excellent indicator of the scientific novelty of the MolTools project is the large number of patents and patent applications that have been filed during the course of the project so far, many of which are being exploited commercially. As far as generation of start-up companies is concerned, Olink, founded by partner 1 (UU) to exploit proximity ligation technology, has already been mentioned. Similarly, partner 7 (UAAR) founded a new company (In Situ RCP A/S) for exploiting the IPR generated by this group, and BI (partner 11) is in the process of establishing Discerna Ltd, in order to commercialise ribosome display and in situ protein array systems.

Results of the work in the MolTools project have been disseminated through our public website, www.moltools.org, which has been accessed on average by 200 visitors a day. The work has also been presented for the scientific community through research articles, conference presentations, and through 14 MolTools symposia and workshops that have been arranged by the consortium. Hands-on practical training in our techniques has been made available through the MolTools training courses arranged in Uppsala, Berlin, Paris, and Heidelberg.