

# Genome Projects and the Functional-Genomic Era

Sascha Sauer\*, Zoltán Konthur and Hans Lehrach

Max Planck Institute for Molecular Genetics, Department of Vertebrate Genomics, Ihnestrasse 73, D-14195 Berlin, Germany

**Abstract:** The problems we face today in public health as a result of the - fortunately - increasing age of people and the requirements of developing countries create an urgent need for new and innovative approaches in medicine and in agronomics. Genomic and functional genomic approaches have a great potential to at least partially solve these problems in the future. Important progress has been made by procedures to decode genomic information of humans, but also of other key organisms. The basic comprehension of genomic information (and its transfer) should now give us the possibility to pursue the next important step in life science eventually leading to a basic understanding of biological information flow; the elucidation of the function of all genes and correlative products encoded in the genome, as well as the discovery of their interactions in a molecular context and the response to environmental factors. As a result of the sequencing projects, we are now able to ask important questions about sequence variation and can start to comprehensively study the function of expressed genes on different levels such as RNA, protein or the cell in a systematic context including underlying networks. In this article we review and comment on current trends in large-scale systematic biological research. A particular emphasis is put on technology developments that can provide means to accomplish the tasks of future lines of functional genomics.

**Keywords:** Functional genomics, genotyping, expression analysis, proteomics, RNAi, cell, microarray.

## INTRODUCTION

All processes in biology can be seen as the computation of the information encoded in the genome of an organism into its phenotype, modified by environmental factors. It is known for decades, for example, that subtle differences in the responses of patients to medication are partly due to inherited variation [1]. Since the identification of the structure of DNA in 1953 [2], progress in understanding these processes has largely been driven by methodologies such as cloning, amplification and sequencing of nucleic acids, the application of robotic devices for high-throughput analyses and the use of bioinformatic tools for sequence and gene analysis. The human genome initiative - sequencing the entire human genome and deciphering the sequence of all (including unknown) genes - has been the most important step in this development [3]. There are still many interesting genome sequencing and sequence analysis projects going on, such as sequencing evolutionary important organisms like oikopleura, including functional annotation and prediction of recently unidentified genes by bioinformatic approaches [4]. Further sequencing projects are concerned with comparative genomic studies for shedding light on the evolution of crucial proteins, such as the major histocompatibility complex (MHC) [5] as well as on evolutionary processes like gene duplication [6].

The analysis of the genomic sequence and its variation are first steps towards the ultimate goal of the genome projects; the systematic analysis of all levels and all components executing the flow of information from the genome to the phenotype of the organism. We will have to

understand this 'computational' process to have any chance to make real progress in the treatment of most common diseases we are faced with. In the next phase of the work - the phase of functional genomics - we will essentially have to generate all types of information, which we are currently generating by hand on a few interesting genes on many genes. We will have to focus on gene products and their roles in the complex networks of biological processes on a much larger scale and in a high-throughput manner. This will have to be complemented by the development of new informatics tools, which enable predictions for processes based on complex data sets [7]. Nevertheless, this will still not be sufficient to make ultimate predictions about human biology in general. We will have to individualise this knowledge based essentially on genomic data (both structural and functional) to enable, for example, optimised treatment of individual patients.

We have just, in a large, longterm international effort, completed and refined the sequence of the human genome. It is very likely, that the systematic analysis of the components found in the genome (protein coding genes, non-coding RNA, regulatory elements etc.) will require an effort, which is again likely to be orders of magnitude above the generation of the sequence. Individualising this information will then, again, require an increase in the amount of necessary data of many orders of magnitude. This process will, therefore, to a large extent require a consistent program in technology development. New technologies are required to generate the enormous amounts of data (and insights) mandatory to understand the full process of the genome project - a major contribution to improving the health of the individuals.

## TRENDS IN TECHNOLOGY DEVELOPMENT FOR THE BIOLOGY OF THE FUTURE

At present, we are able to sequence entire genomes, to study DNA variation such as single nucleotide

\*Address correspondence to this author at the Max Planck Institute for Molecular Genetics, Department of Vertebrate Genomics, Ihnestrasse 73, D-14195 Berlin, Germany; Tel: +49-30-8413-1565; Fax: +49-30-8413-1365; E-mail: sauer@molgen.mpg.de

polymorphisms (SNPs) covering the whole human genome [8], as well as to identify and analyse many proteins [9,10]. However, all these efforts require immense (financial) resources and can in general not yet be applied to compare variation between organisms, or between different somatic cells within one individual. Hence, there is a growing demand of more (cost-) efficient technologies in the future.

Recent years have seen a rapid growth of techniques for high-throughput analyses of genes, transcripts, proteins, and cells using highly parallel array and microarray technologies [11], but current methods still capture only a small fraction of the information embodied in these molecules. Several consortia, as for example the MolTools consortium funded by the EU FP6 research program (<http://www.moltools.org>), are presently involved in the development of novel procedures, which should enable systematic approaches to study the molecules that make up genomes and their molecular correlatives. The purpose of such technology development projects is to spawn next-generation technologies and establish an infrastructure for large-scale molecular analyses, which should allow the investigation of the genome, transcriptome and proteome of individuals, and even of single cells of an individual.

With the successive availability of databases storing molecular complements for an increasing number of organisms in a more or less systematic manner and in addition to much progress on bioinformatic analysis tools, there is an urgent need for methods to use this information by investigating large sets of molecules in large numbers of samples. This requires novel integrated databases, such as the GenomeMatrix (<http://www.genome-matrix.org>) established in a collaboration between the Max Planck Institute for Molecular Genetics (MPI-MG) and the German Resource Center for Genome Research (RZPD), which provides a cross species overview over many functionally relevant data on many genes in parallel and gives a more complete picture of single genes, their functions and interactions in molecular networks.

It is furthermore important to go one step further from the mere descriptive approach of complex biological phenomena. We have to develop (algorithmic) models reflecting molecular pathways and have to look deeper into cellular networks to understand data from comprehensive large-scale genomics or proteomics experiments [12,13]. Progress in understanding cell biological mechanisms will facilitate drug development for the treatment of complex diseases, and will eventually lead to personalised medicine. This will almost certainly have to be based on detailed mathematical modeling of complex networks of processes, since it is highly unlikely that many complex networks underlying for example diseases, such as cancer will ever be understandable on a single gene basis.

On the other hand, significantly improved molecular technologies will be important to define new drug targets, to diagnose and treat complex diseases, as well as to select improved breeds of domestic animals and plants. Moreover, these methods are essential for arising research areas such as biodiversity monitoring at the molecular level [14]. In this article, we present a number of different trends of current genome research projects with a particular focus on technology development. However, due to the vast number

of different lines of genome research, we are fully aware that not each topic could be satisfactorily presented in this review. Therefore we mainly focus on specific projects that we consider as particularly valuable and apologise for the omissions of other approaches, which we have been unable to include, e.g. use of model organisms etc. [15]. Here, we concentrate on technologies that were initially developed and very successfully applied for nucleic acid detection and can potentially be modified for the efficient large-scale analysis of proteins.

In our view the following six interrelated analytical modes are of particular interest:

- 1) Scanning parts or whole genomes for sequence alterations (re-sequencing);
- 2) Tracing of extensive sets of genetic variants in large numbers of individuals (typing of DNA markers);
- 3) Analysis of gene expression patterns at high quantitative and sequence resolution (transcriptome analysis);
- 4) Detection of expression and interactions of large sets of proteins (protein analysis);
- 5) Monitoring at the ultimate level of single DNA, RNA, or protein molecules (single molecule analysis);
- 6) Functional characterisations of complex processes in viable cells at high throughput (cell analysis).

## 1. Re-Sequencing

Given sufficiently powerful technologies, it should be possible to identify changes of whole genomes, such as SNPs [16], or epigenetic phenomena such as 5'-methylation of cytosines. Bases that can be methylated in a genome are termed methylation variable position [17], short MVP, alluding to "most valuable player". SNPs and MVPs can predispose to disease or to other specific traits. While for SNP-detection the starting material is genomic DNA, the detection of MVPs requires the conversion of the methylation difference into a sequence difference by bisulphite treatment [18]. In the presence of bisulphite, the 5'-methylcytosine remains stable, while normal cytosines are converted to uracil and finally to thymine using PCR amplification. The resulting PCR products serve as templates for sequencing or any other suitable method to detect the methylation positions of a genomic region of interest.

For re-sequencing studies, methods are required for rapid and cheap analyses of partial and, ultimately, total genomes. Clearly, the technology development will have to aim for price reduction in orders of magnitude to be able to (re-) sequence the genome of individual patients at acceptable costs. The present gold standard method to detect unknown DNA sequence variation is still the Sanger sequencing procedure [19], significantly improved by modifications in reagents, instrumentation and software. For SNP discovery, re-sequencing of a DNA cohort representing a population of interest is performed. The obtained sequence traces are then analysed for variants using bioinformatic programs, such as PolyPhred [20]. For MVPs, modified sequencing protocols are used due to the occurrence of many thymine and adenine nucleobases. Cytosine nucleobases in the sequence traces of

bisulphite-treated DNA indicate a former methylated cytosine.

A large number of alternative techniques reflect the unsatisfied need for improved re-sequencing methods. These include denaturing HPLC [21], mass spectrometry [22-24], DNA microarrays [25], chemical or enzymatic cleavage of mismatched duplexes [26,27], denaturing gel electrophoresis [28], sequence-specific conformation polymorphism [29], and techniques based on bacterial mismatch repair systems [30].

An other procedure for re-sequencing is the polony technology [31]. This method integrates DNA isolation, amplification, and sequencing by the use of polymerase colonies (polonies). Small amounts of genomic DNA are integrated into a polyacrylamide gel covered on a microscope slide containing all reagents required for PCR. As the concentration of template DNA is very low, the chromosome fragments are efficiently separated from each other. Using a thermal cycler especially manufactured for slides, an in-gel PCR can be performed to amplify genomic loci of interest. One modified PCR primer is covalently attached to the polyacrylamide and hence, restricts the diffusion of the PCR products. This accumulation of double-stranded amplicons around chromosomes forms so-called polonies. After removing the unattached strands of the amplified DNA from the gel by heating and washing the slide [32], the resulting single-stranded PCR templates are used for subsequent sequencing using cycled nucleotide extension reactions with fluorescent dNTPs. High density of minimal overlapping polonies can be achieved allowing multiple polonies to be sequenced in parallel. Specific software for automated image alignment and sequence calling assists in subsequent data analysis. However, formats that achieve a higher density than the current gel matrices used, would make this procedure more powerful, at least in terms of throughput.

Recently, Church and co-workers have published a new DNA sequencing method based on emulsion PCR of single DNA molecules, the 'sequencing-by-ligation' concept and parallel four-color imaging in polyacrylamide gels, which was initially developed for the polony approach [33]. Simultaneously the company 454 Life Sciences has presented the development of an integrated platform that uses 'sequencing-by-synthesis' (pyrosequencing) in picoliter volumes and highly parallel luminescence detection. This technology is already commercially available and could provide a 100-fold throughput improvement over the fluorescence-based Sanger sequencing approach [34].

In our laboratory, we have developed an alternative approach for characterisation or signature profiling of genomes called oligonucleotide fingerprinting (OFP) [35]. During this process ~ 200 short synthetic oligonucleotides (8-mers) are hybridised to large numbers of cDNA or genomic DNA clones on an array. As a result, sequence-dependent fingerprints (vectors of hybridisation intensities) are derived for each individual clone. OFP has proven to be a powerful procedure for cDNA and genomic DNA library characterisation and normalisation. The method can be performed using either radioactive [36], fluorescence [37] or, recently, MALDI mass spectrometric detection [38]. While the fluorescence-based approach allows for experimentation

in miniaturised formats using induced fluorescence resonance energy transfer (iFRET) detection [39], currently under development, we have shown that highly parallel analysis can be performed using the MALDI approach.

## 2. Typing of DNA Markers

Once DNA variations, such as SNPs, have been detected and properly validated (in most cases by DNA sequencing), these data are commonly stored in public databases such as dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) [40]. Several million common SNPs are known for the human and other genomes that can facilitate tracing the inheritance of genetic factors associated to complex diseases such as cancer, hypertension, obesity or cardiovascular diseases. SNPs are used as high-density markers in whole-genome scanning approaches [41] or in more focused candidate gene association studies - to compare large DNA cohorts of well phenotyped patients with unaffected control groups [42]. Large cohorts are required to improve statistical power of genetic tests.

Moreover, SNPs should also be considered when studying RNA expression or proteins. They can, for example, influence promotor efficiency or pose problems in sequence tag RNA expression analysis techniques, such as SAGE (Serial Analysis of Gene Expression) [43] or MPSS (Massively Parallel Signature Sequencing) [44]. In case of protein-coding SNPs they can lead to proteins with diverse amino acid sequences resulting in functional variation, which can complicate protein analysis. In many cases particularly to improve power of genetic analysis several SNPs in close proximity have to be considered as one group defined as the haplotype [40], the phase of SNP alleles on a single chromosome, which can be more meaningful than the single SNP information. The haplotype can be either determined using prediction algorithms or when this is not accurately feasible due to large numbers and high heterozygosity of SNPs-by employing molecular haplotyping procedures, such as "clone-based systematic haplotyping" (CSH) [45].

For the purpose of SNP typing, many complementary wet-lab methods have been developed, each covering a different field of application. One can roughly define the (cost-) efficiency of the different methods by the number of SNPs and DNAs under investigation. While array-based methods [41,46,47] combined with high-multiplex generation of SNP products including molecular bond-coding are efficient when analysing several thousand SNPs at a time in hundreds of DNA samples, solution based approaches such as TaqMan using fluorescence detection [48], or mass spectrometry based procedures [49-53] are rather economical in candidate gene studies where several ten to hundred SNPs have to be typed in cohorts of several thousand DNAs. Many of the SNP typing methods are continuously optimised in terms of throughput, cost efficiency and applicability.

Once a SNP has been found as an indicator of a disease, it can be used for diagnostics. However, only few examples have been published by now. A highlight is the study of Hugot and co-workers who have discovered three DNA markers that cause predisposition to Crohn's disease, which could be used for diagnosis [54].

Similarly to SNPs, MVPs have a potential to be commercialised as diagnostics. Epigenetic phenomena are of high interest to understand gene regulation on the DNA level, since methylation patterns can cause up- or down-regulation [55]. Specific methylation patterns in certain genes represent early-stage indicators of upcoming diseases, such as prostate cancer. Epigenomics AG (<http://www.epigenomics.com>) in close alliance with a number of pharma companies keeps working intensively on this subject to develop early-stage cancer diagnostics based on DNA methylation [56,57].

### 3. Transcriptome Analysis

Efficient methods to investigate the expression of large numbers of genes are crucial to exploit the dramatically increased information contained within the configuration of genomes. Hence, methods significantly improving the convenience, precision and throughput are of great value in expression analysis. Studying the activity of a genome by measuring expression levels of transcripts provides important biological insights. For instance, it allows identifying factors that reflect or contribute to disease states or signal responses to environmental interference or drug treatment. One result of RNA expression studies could be the development of a very specific diagnostic or molecular phenotyping tool. This tool could improve the characterisation of patients in DNA marker based association studies thereby leading to more significant results and contribute to more specific, personalised medicine. However, today's popular microarray technologies for RNA expression analysis have limitations in terms of costs and sensitivity. Presently they capture only a fraction of the information represented by expression patterns in a tissue. Allelic and splice variants, as well as weakly expressed genes are frequently underrepresented. Careful quality control is a prerequisite to avoid false results [58]. Real-time reverse transcription (RT) PCR is a good alternative for quantitative detection of specific mRNAs and is frequently used for verification of array-based RNA monitoring. Once established, the procedure is rapid, inexpensive and requires low amounts of starting material. However, real-time RT-PCR involves significant efforts to optimise amplification conditions. Since the amplification of the mRNA can severely hamper expression profiling, this method has to be carefully assessed [59]. It seems currently unrealistic to combine the accuracy and dynamic range of alternative but sequential real-time PCR with the wide parallelism of hybridisation arrays. The combination of the advantageous features of microarrays and RT-PCR might be achieved by transferring RT-PCR in array-like 3-D formats such as nanowell plates with miniaturised wells having diameters between ~ 100 and ~ 500 micrometers [60]. However, novel technologies to measure gene expression with the ability to accurately analyse allele-specific and splice-variant-specific expression profiles at a dramatic cost reduction, and with a throughput in the range of at least 100,000 measurements per day, are essential for basic studies on RNA expression profiling. The technologies to be developed for this purpose will certainly require substantial modification for routine clinical tests of a limited number of diagnostic RNA markers.

Another important factor in the analysis of RNA expression is assay sensitivity. It occurs quite frequently that the molecular characterisation of clinical samples is complicated and limited by the amount of available sample material. To overcome this problem, two major strategies might be successful. (i) A representation of the starting material is amplified for example by means such as circularisable oligonucleotide combined with rolling circle amplification [61]; (ii) the assay is miniaturised to allow the analysis of very small samples, for example by using nanowells consisting of wells with a volume of about 10-50 nl [62]. The ultimate limit of such miniaturisation is the number of molecules of one particular type present in such a small volume of a complex mixture. Extreme miniaturisation of assays will therefore only be feasible, if the complexity of the input mixture can be reduced by a prefractionation or selective amplification step.

Since 1998 a new mode of RNA analysis termed RNA interference (RNAi) has become popular [63]. This procedure represents a powerful approach for reverse genetics where a gene of known- or unknown function is efficiently disrupted to determine the effects on a function or a molecular network. The principle of RNAi was discovered by Fire and Mello who injected double-stranded RNA into *C. elegans* inducing a sequence-specific degradation of cytoplasmic mRNAs containing the same sequence as the double-stranded RNA trigger. Later, it was discovered that RNAi could be efficiently applied to cell cultures as well as to knock out model organisms such as mouse and rat in a significantly shorter time than by traditional methods based on homologous recombination [64]. For these purposes, so-called small interfering RNAs (siRNAs) represent very exciting tools [65]. siRNA based screening can be carried out in *C. elegans* or *D. melanogaster* as well as in mammalian tissue cells [66]. With the availability of the genome sequences of the key organisms, siRNAs can be designed for silencing gene expression in a highly effective and sequence-specific manner. Cost effective methods for the production of siRNAs [67], as well as bioinformatic tools that aid in siRNA assay design [68] have been developed. The effects caused by RNAi can be studied on several levels addressed in this article, particularly by looking at the ensemble of the proteins and at reaction pathways of the entire cell. However, RNAi has also limitations that should be kept in mind. For example, RNAi ultimately knocks down complete proteins, whole protein families or untargeted proteins [69,70]. Hence, not only single functions of a targeted gene are knocked out. In other cases, phenotypic effects are not necessarily seen, as other protein family members might take over the function of a knocked-down protein that would otherwise not be favored. Further-more, RNAi can cause unpredictable secondary effects.

### 4. Protein Analysis

Proteins are the main effector-molecules in the cell, whose levels and modifications significantly mediate the cell's life cycle. Protein analysis would probably provide a more efficient approach for molecular phenotyping of patients than RNA profiling. The identification and characterisation of proteins, comprising for example post-translational modifications, protein-protein and protein-DNA

interactions go beyond mRNA expression analysis, revealing the functional roles of large sets of proteins. However, comprehensive protein studies would require novel, extremely sensitive and specific approaches for explicit protein detection and quantification.

The available genome sequences make it possible to clone and purify all their encoded proteins, which is, however, a very big challenge. DNA cloning techniques and cDNA clone resources established during the sequencing projects as well as automated systems for sample purification provided the basis to produce proteins in high-throughput, for example in structural genomic projects [71] that are devoted to the determination of the structure of all proteins in an organism. Ideally, the 3-D structures of the entire proteins of a cell, analysed by x-ray techniques or for proteins smaller than 20 kDa by nuclear magnetic resonance (NMR), could serve as a resource for functional analysis and drug development [72,73]. Clearly, this task is practically impossible to fulfill, since the number of different splice variants and changing levels of post-translational modifications exceed the number of genes by a minimum of factor 5. It can however be expected, that threading methods [74] might, in many cases, be able to successfully predict structures of such variants, once some of the structures have been determined experimentally. One of the major practical challenges in this endeavor is the preparation of proteins for structural determination producing high-quality crystals for x-ray analysis or proteins appropriate for NMR. In addition, these techniques are severely hampered, particularly in respect of the structural analysis of hydrophobic membrane proteins and protein complexes.

Protein arrays have evolved from more recent DNA arrays. However, the handling of proteins is not easy due to their 3-D structures and because they tend to change if not even denature when transferred from solution to a surface. Moreover, protein interaction can not be defined as easily as DNA strand interactions in nucleic acid hybridisation governed (in most cases) by the Watson-Crick rules. Protein arrays are distinguished in analytical and functional arrays. In general, analytical protein arrays build on protein detection by immunoassay strategies: These strategies include (i) direct immunoassays, in which a large protein collection is immobilised and specific proteins are detected by labeled antibodies; (ii) sandwich immunoassays, in which antibodies are attached to a solid phase and bound proteins are visualised by a second labeled antibody; (iii) antigen capture assays, in which proteins are captured by immobilised antibodies, and subsequently analysed by surface plasmon resonance (SPR), mass spectrometry or by being labeled in advance. For (i) and (ii) detection can be provided by enzyme-catalysed chemiluminescence, fluorophore or radiolabelling and rolling circle amplification detection by fluorescence [72].

Efforts are on the way to identify protein-ligand binding pairs that will act as a resource with broad applicability in research and diagnostics [75]. Potential binders include polyclonal [76], monoclonal [77] and phage display-derived antibodies [78], scaffold proteins [79], as well as DNA [80], RNA [81] or photo [82] aptamers. Efficient "evolutionary" procedures for the establishment of combinatorial libraries and selection of those molecules are in focus of current

research. High-throughput systems for aptamer selection have been developed [83,84] and represent - besides antibody production - the presently most promising platform for industrial binder generation.

Our department was one of the first to introduce protein arrays based on the experience with DNA arrays and using cDNA clone libraries for protein expression [85]. A major field of application of these arrays in our hands is the screening with sera from autoimmune patients. For example, by comparing the signal patterns of antibodies from diseased versus healthy sera by dense protein arrays, we were able to identify antibody/autoantigen pairs previously unknown to be involved in these particular diseases [86]. Also, we were particularly active in establishing antibody arrays and in studying several aspects such as surface chemistry, spotting procedures, buffer and storage conditions [87]. Only recently we have set up a method, which allows the qualitative analysis of hundreds of recombinant antibodies on a single chip by the multiple spotting technique (MIST) [88,89]. Since the beginning of protein array technology, the design of the arrays is steadily improving, specifically with respect to designing controls and standards on these surfaces to aid the analysis and interpretation of results obtained. While so far protein arrays have been used for qualitative control assays in the evaluation of binding partners, the main objective in current developments is to use protein chips at high-throughput for a wider variety of research applications and specifically to evaluate serum samples from patients suffering from specific diseases.

The current minority of protein arrays are functional arrays, on which a set of proteins or the entire proteome of a cell is immobilised and can be functionally tested using for example biochemical assays. Procedures have recently been introduced for the systematic investigation of biochemical function, such as the determination of substrate specificity of predicted yeast kinases or the screening for protein-protein interactions [90]. It is obviously desirable to have more of these kinds of arrays available in the future and currently major efforts for providing arrays for functional protein analysis are on the way.

A promising new technology to provide large protein arrays is for example PISA ('protein *in situ* array') [91], a procedure for cell free protein expression introduced by Taussig and co-workers. PISA is very interesting because many proteins can hardly be expressed in *E. coli* or other cell based expression systems; for example, proteins with intrachain disulphide bonds tend to denature and form inclusion bodies, and some proteins are toxic to the bacterial host cell. In the PISA procedure, proteins are expressed in a cell-free environment from PCR products. The starting point can be any DNA source providing that information for primer design is available. The PCR product, with appropriate initiation sequence and stop codon, is used as a template to produce the protein in a coupled transcription/translation system. For the immobilisation of the protein on a surface, it carries a tag sequence and the translation is performed on arrays with agents that bind the tag. This way one can efficiently carry out single step production of protein arrays from DNA without the need for cumbersome purification and protein handling.

Available, well-established proteomic methods, most notably those building on two-dimensional (2-D) gel electrophoresis [92] or liquid-chromatography [93] coupled with electrospray ionisation (ESI) [94] and/or matrix-assisted laser desorption ionisation (MALDI) mass spectrometry [95], permit a detailed overview of the protein complements of cells and tissues. Particularly, protein variants and secondary modifications can be well characterised by such methods, provided that sufficient amounts of sample material is available [96]. However, due to the inherent limitations of the mass spectrometers, quantification and detection of low-abundance effectors, such as cytokines, is not always feasible. Recently, we have studied 8,767 proteins from the mouse brain proteome using large 2-D gels and have found 1,324 polymorphic proteins from the European collaborative interspecific backcross [97]. We subsequently mapped 665 of these proteins genetically and could identify 466 proteins by mass spectrometry. Qualitatively polymorphic proteins reflected to about 96 % changes in conformation and/or mass while quantitatively polymorphic proteins showed 73 % frequency of allele-specific transmission in codominant heterozygotes. We concluded that single proteins might act as polygenic traits due to variations in protein isoforms and varying quantities. A further conclusion of our study was that genetic analysis of proteomes could reveal the types of polymorphism, which are most relevant in association studies of complex diseases.

Recently, Cellzome (<http://www.cellzome.com>) introduced a mass spectrometry based technology that could be used as a very powerful tool for proteome-wide screening of protein-protein interaction [98] based on a mild co-precipitation technique called tandem affinity purification (TAP) [99]. Applying the Cellzome technology, the protein interaction network of around 32 known and candidate TNF-alpha/NF-kappa B pathway components was studied [100]. Additional information was gained by directed functional perturbation studies using RNAi. 221 molecular associations and 80 previously unknown interactors were identified by this route, including 10 new functional modulators of the pathway. This approach provided valuable insight into the mechanism of the TNF-alpha/NF-kappa B pathway and could be generally useful to study any pathway relevant to disease.

An alternative, genetic procedure, termed yeast-two hybrid screening (Y2H) [101], was introduced and further developed into an automated *in vivo* screening system of protein-protein interaction networks. Y2H was successfully applied to human disorders such as Huntington's disease [102]. The method relies on the use of a split transcription factor divided into two functionally distinct domains: one, mediating transcriptional activation, and the other responsible for complex building with DNA. Fusing one of the proteins in a putative binding pair (called bait) to the DNA binding domain and fusing the other protein (called prey) to the activation domain, the transcription factor is reconstituted by interaction of the two proteins producing a measurable activity of the reporter gene. Y2H was employed to identify protein-protein interaction in the yeast proteome [103]. However, a problem with Y2H seems to be that different data sets show significantly different results. Comparing Y2H data from two different screens and with

data sets relying on different technologies such as mass spectrometry often yield substantial differences. One reason that restricts the Y2H procedure can be the use of gene fusions and their possible influence on the localisation and secondary modifications of the proteins, the differences between post-translational modification in yeast and e.g. the mammalian cell, as well as the fact, that only binary interactions can be detected. Further limitations concern proteins that are difficult to analyse, such as integral membrane proteins, transcription factors and (large) proteins that do not enter the nucleus of a cell. Despite these disadvantages Y2H can produce valuable results. To overcome some of these difficulties, alternative versions of the basic two-hybrid approach have been introduced [104].

To increase sensitivity and specificity of classical antibody-based detection, which is commonly used in diagnostics or for protein arrays, new molecular approaches such as proximity ligation [105] might be useful. This method is based on the padlock technology firstly introduced for SNP genotyping in a highly parallel fashion. In this technique, two or more protein binders such as aptamers or antibodies, which bind in close proximity on a single target molecule or molecular complex can be detected provided that the binders are attached to DNA strands. Using a complementary template nucleic acid the free ends of these two DNA strands can be enzymatically ligated in a proximity-dependent reaction. The ligation reaction yields a DNA molecule that encodes the identity of the detected protein by a tag sequence. Individual or sets of DNA molecules can subsequently be amplified by PCR or alternative methods, and either detected in solution or on a surface. In principle, proximity-ligation offers important advantages over procedures currently in use for protein detection such as the well-known enzyme linked immunosorbant assay (ELISA): The requirement for simultaneous detection of a target molecule by two or more binders provides a basis for highly specific detection. Cross-reactive detection of irrelevant targets is less likely compared to traditional single molecule binders. Furthermore, the presence of several epitopes on a target molecule can be independently identified, allowing for the analysis of secondary modifications of the target. Finally, reagents that do not bind a target molecule are unlikely to produce a non-specific background, since this would require proximity to other detection reagents.

## 5. Single Molecule Analyses

Sample material availability is an important issue particularly for those approaches covered in this article for which re-amplification is not possible. Miniaturisation reduces the amount of sample required for the analysis and enables increased data production at lower cost.

Moreover, molecular biological analyses yield statistical measures over time and averages over millions of molecules and/or cells, which might very well differ from each other. For many problems in biology, the analysis of single molecules/single cells is fundamentally important and therefore, procedures will be required for highly efficient, specific detection of individual molecules. Such methods will particularly be valuable when studying biological

variation in large numbers of individual samples on the DNA, RNA, and protein level. The detection of single DNA, RNA and proteins, and of the interaction between sets of molecules *in situ* or *in vivo* would permit a more detailed view of cellular processes and would overcome the limitations imposed by current procedures that average analyses across heterogeneous populations of cells and subcellular domains.

A lot of progress has been made in miniaturizing assays. A number of procedures are available now, in which only a single molecule is detected by nucleic acid amplification based methods such as single molecule PCR [106], single molecule RT-PCR or the amplification of single molecules using the polony approach [31,32]. Another strategy emphasises direct detection using for example optical devices that allow for the analysis of single molecules such as fluorescence microscopy [107], atomic force microscopy [108], or fluorescence correlation spectroscopy (FCS) [109]. FCS, for example, allows detection and quantification of interacting biomolecules at very low concentrations. Binding of a small fluorescently labeled molecule to a larger ligand results in a change in diffusion time. FCS can detect fluorescence fluctuations, which are caused by a single molecule diffusing in a focal detection volume. A more recent extension of FCS to dual-colour fluorescence cross-correlation spectroscopy circumvents the precondition for a mass difference between the binding molecules [110].

Alternatively, amplification-based approaches building, for example, on circularisable probes and rolling circle amplification can allow individual cells to be screened for multiple specific molecules. The analysis can be performed at a molecular and spatial resolution, and permits discrimination of single nucleotide variation and protein co-localisation within defined cellular domains [111]. Mentioned proximity-ligation probes might be efficient reagents to trigger rolling circle replication reactions allowing single molecules to be visualised by microscopy. These probes have been shown to allow specific detection of RNA molecules at single-nucleotide resolution [112] and can potentially be used to quantitate single or large sets of transcripts with sufficient specificity to distinguish splice or allelic variants, either in solution or *in situ*.

## 6. Cell Analysis

Assessing gene function in living cells is becoming increasingly important in life science. In the past years, large-scale functional genomic and proteomic approaches have fundamentally changed our view of the possibilities to study the molecular basis of cells and tissues in health and disease. As a result, for example in cancer research, new diagnostic microarray methods for tumor classification of tissue biopsies and for prognostication based on gene expression monitoring have been developed [113]. However, functional genomics and proteomic profiling procedures produce results that are in general correlative and descriptive in nature. Logically, the next challenge in this branch of life science lies in the identification of direct causalities, such as the effects of genes and signaling pathways in living cells. For example, the impact of genes introduced into living cells in a microarray format can be analysed [114], or RNAi-based procedures can be applied [115].

The ability to explore the effects of genes on cellular phenotypes and signaling in high-throughput is of fundamental importance in molecular biology. However, most cell biology studies to date are not carried out systematically in large-scale, but rely on approaches that target one gene, pathway, phenotype, or cell type at a time. Procedures enabling cell biology to be performed on a global scale are needed. Such methods can push functional genomics and biological system theories as well as provide new insights into key therapeutic targets. A promising technology in this direction represents transfected cell microarrays that have been established and optimised for high-throughput functional characterisation of genes on the level of the cell [114]. In this method, full-length genes are inserted in expression vectors and spotted at high density on a glass slide along with a transfection reagent using a robot. When the microarray of DNA constructs is covered with a layer of adherent cells, only the cells growing on top of the DNA spots become transfected. This results in the expression of specific proteins in spatially distinctive groups of cells and the phenotypic effects of the “reverse transfection” of up to thousands of genes can be detected using appropriate cell-based assays in parallel. Transfected cell arrays have been used to silence genes in high-throughput. Fluorescence microscopy combined with cell-by-cell image analysis enabled to visualise and quantitate the effect of the siRNAs on cell function. Drawbacks of the procedure, however, are that (i) only adherent cells can be analysed, (ii) the cationic transfection agent restricts usage to a subset of specific cells, and (iii) modified methods to enable long-term storage of siRNA microarrays still have to be developed [64].

## CONCLUDING REMARKS

The technologies we have described, as well as many others under development offer a new way to look at biological processes. Applying these technologies to fundamental questions concerning life and death, nutrition and even renewable energies, we might have a chance to solve many of the real world problems all of us are facing in public health as well as in agronomics. This message does however not seem to fully reach the scientific community and the policy makers, who should be very aware of the limitations of last-centuries approaches. We can just hope that in the interest of science and society as a whole this attitude will change rapidly and will allow providing resources and a livable environment for the next generations.

## ACKNOWLEDGEMENTS

We would like to acknowledge the European Union (grant LSHG-CT-2004-503155), the German Ministry for Research and Education (NGFN2, grant 01GR0414 and 01GR0427), and the Max-Planck Society for support.

## REFERENCES

- [1] Meyer, U. A. *Nat. Rev. Genet.* **2004**, *5*, 669.
- [2] Watson, J. D.; Crick, F. H. *Nature* **1953**, *171*, 737.
- [3] Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann,

- L.; Lehoczy, J.; LeVine, R.; McEwan, P.; McKernan, K.; Meldrim, J.; Mesirov, J. P.; Miranda, C.; Morris, W.; Naylor, J.; Raymond, C.; Rosetti, M.; Santos, R.; Sheridan, A.; Sougnez, C.; Stange-Thomann, N.; Stojanovic, N.; Subramanian, A.; Wyman, D.; Rogers, J.; Sulston, J.; Ainscough, R.; Beck, S.; Bentley, D.; Burton, J.; Clee, C.; Carter, N.; Coulson, A.; Deadman, R.; Deloukas, P.; Dunham, A.; Dunham, I.; Durbin, R.; French, L.; Grafham, D.; Gregory, S.; Hubbard, T.; Humphray, S.; Hunt, A.; Jones, M.; Lloyd, C.; McMurray, A.; Matthews, L.; Mercer, S.; Milne, S.; Mullikin, J. C.; Mungall, A.; Plumb, R.; Ross, M.; Showkeen, R.; Sims, S.; Waterston, R. H.; Wilson, R. K.; Hillier, L. W.; McPherson, J. D.; Marra, M. A.; Mardis, E. R.; Fulton, L. A.; Chinwalla, A. T.; Pepin, K. H.; Gish, W. R.; Chissoe, S. L.; Wendt, M. C.; Delehaunty, K. D.; Miner, T. L.; Delehaunty, A.; Kramer, J. B.; Cook, L. L.; Fulton, R. S.; Johnson, D. L.; Minx, P. J.; Clifton, S. W.; Hawkins, T.; Branscomb, E.; Predki, P.; Richardson, P.; Wenning, S.; Slezak, T.; Doggett, N.; Cheng, J. F.; Olsen, A.; Lucas, S.; Elkin, C.; Uberbacher, E.; Frazier, M. *Nature* **2001**, *409*, 860.
- [4] Seo, H. C.; Kube, M.; Edvardson, R. B.; Jensen, M. F.; Beck, A.; Spriet, E.; Gorsky, G.; Thompson, E. M.; Lehrach, H.; Reinhardt, R.; Chourrout, D. *Science* **2001**, *294*, 2506.
- [5] Hurt, P.; Walter, L.; Sudbrak, R.; Klages, S.; Muller, I.; Shiina, T.; Inoko, H.; Lehrach, H.; Gunther, E.; Reinhardt, R.; Himmelbauer, H. *Genome Res.* **2004**, *14*, 631.
- [6] Panopoulou, G.; Hennig, S.; Groth, D.; Krause, A.; Poustka, A. J.; Herwig, R.; Vingron, M.; Lehrach, H. *Genome Res.* **2003**, *13*, 1056.
- [7] Klipp, E.; Herwig, R.; Kowald, A.; Wierling, C.; Lehrach, H. *Systems Biology in Practice. Concepts, Implementation and Application*, Wiley-VCH, Weinheim, **2005**.
- [8] The International HapMap Consortium *Nature* **2003**, *426*, 789.
- [9] Aebersold, R.; Mann, M. *Nature* **2003**, *422*, 198.
- [10] Reboul, J.; Vaglio, P.; Rual, J. F.; Lamesch, P.; Martinez, M.; Armstrong, C. M.; Li, S.; Jacotot, L.; Bertin, N.; Janky, R.; Moore, T.; Hudson, J. R., Jr.; Hartley, J. L.; Brasch, M. A.; Vandenhoute, J.; Boulton, S.; Endress, G. A.; Jenna, S.; Chevet, E.; Papatotipoulos, V.; Tolia, P. P.; Ptacek, J.; Snyder, M.; Huang, R.; Chance, M. R.; Lee, H.; Doucette-Stamm, L.; Hill, D. E.; Vidal, M. *Nat. Genet.* **2003**, *34*, 35.
- [11] Grigorenko, E. V. *DNA Arrays: Technologies and Experimental Strategies*, CRC Press, Boca Raton, **2001**.
- [12] Norberg, J.; Nilsson, L. *Q Rev. Biophys.* **2003**, *36*, 257.
- [13] Kell, D. B. *Curr. Opin. Microbiol.* **2004**, *7*, 296.
- [14] Pollock, D. D. *Appl. Bioinformatics* **2002**, *1*, 81.
- [15] Beckers, J.; Hrade de Angelis, M. *Curr. Opin. Chem. Biol.* **2002**, *6*, 17.
- [16] Brookes, A. J. *Gene* **1999**, *234*, 177.
- [17] Eckhardt, F.; Beck, S.; Gut, I. G.; Berlin, K. *Expert Rev. Mol. Diagn.* **2004**, *4*, 609.
- [18] Olek, A.; Oswald, J.; Walter, J. *Nucleic Acids Res.* **1996**, *24*, 5064.
- [19] Sanger, F.; Nickens, S.; Coulson, A. R. *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5463.
- [20] Nickerson, D. A.; Tobe, V. O.; Taylor, S. L. *Nucleic Acids Res.* **1997**, *25*, 2745.
- [21] Underhill, P. A.; Jin, L.; Lin, A. A.; Mehdi, S. Q.; Jenkins, T.; Vollrath, D.; Davis, R. W.; Cavalli-Sforza, L. L.; Oefner, P. J. *Genome Res.* **1997**, *7*, 996.
- [22] Nordhoff, E.; Luebbert, C.; Thiele, G.; Heiser, V.; Lehrach, H. *Nucleic Acids Res.* **2000**, *28*, e86.
- [23] Krebs, S.; Medugorac, I.; Seichter, D.; Forster, M. *Nucleic Acids Res.* **2003**, *31*, e37.
- [24] Stanssens, P.; Zabeau, M.; Meersseman, G.; Remes, G.; Gansemans, Y.; Storm, N.; Hartmer, R.; Honisch, C.; Rodi, C. P.; Bocker, S.; van den Boom, D. *Genome Res.* **2004**, *14*, 126.
- [25] Hacia, J. G.; Fan, J. B.; Ryder, O.; Jin, L.; Edgemon, K.; Ghandour, G.; Mayer, R. A.; Sun, B.; Hsie, L.; Robbins, C. M.; Brody, L. C.; Wang, D.; Lander, E. S.; Lipshutz, R.; Fodor, S. P.; Collins, F. S. *Nat. Genet.* **1999**, *22*, 164.
- [26] Roberts, E.; Deeb, V. J.; Woods, C. G.; Taylor, G. R. *Nucleic Acids Res.* **1997**, *25*, 3377.
- [27] Youil, R.; Kemper, B.; Cotton, R. G. *Genomics* **1996**, *32*, 431.
- [28] Fodde, R.; Losekoot, M. *Hum. Mutat.* **1994**, *3*, 83.
- [29] Hayashi, K. *PCR Methods Applications* **1991**, *1*, 34.
- [30] Faham, M.; Baharloo, S.; Tomitaka, S.; DeYoung, J.; Freimer, N. B. *Hum. Mol. Genet.* **2001**, *10*, 1657.
- [31] Mitra, R. D.; Shendure, J.; Olejnik, J.; Edyta Krzymanska, O.; Church, G. M. *Anal. Biochem.* **2003**, *320*, 55.
- [32] Mitra, R. D.; Church, G. M. *Nucleic Acids Res.* **1999**, *27*, e34.
- [33] Shendure, J.; Porreca, G. J.; Reppas, N. B.; Lin, X.; McCutcheon, J. P.; Rosenbaum, A. M.; Wang, M. D.; Zhang, K.; Mitra, R. D.; Church, G. M. *Science* **2005**, *309*, 1728.
- [34] Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y. J.; Chen, Z.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Godwin, B. C.; He, Y.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L.; Jarvie, M. T. P.; Kirage, K. B.; Kim, J. B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S. M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.; Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. R.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P.; Begley, R. F.; Rothberg, J. M. *Nature* **2005**, in press.
- [35] Poustka, A.; Pohl, T.; Barlow, D. P.; Zehetner, G.; Craig, A.; Michiels, F.; Ehrlich, E.; Frischauf, A. M.; Lehrach, H. *Cold Spring Harb. Symp. Quant. Biol.* **1986**, *51*, 131.
- [36] Meier-Ewert, S.; Lange, J.; Gerst, H.; Herwig, R.; Schmitt, A.; Freund, J.; Elge, T.; Mott, R.; Herrmann, B.; Lehrach, H. *Nucleic Acids Res.* **1998**, *26*, 2216.
- [37] Guerasimova, A.; Nyarsik, L.; Girmus, I.; Steinfath, M.; Wruck, W.; Griffiths, H.; Herwig, R.; Wierling, C.; O'Brien, J.; Eickhoff, H.; Lehrach, H.; Radelof, U. *Biotechniques* **2001**, *31*, 490.
- [38] Bauer, O.; Guerasimova, A.; Sauer, S.; Thamm, S.; Steinfath, M.; Herwig, R.; Janitz, M.; Lehrach, H.; Radelof, U. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 1821.
- [39] Howell, W. M.; Jobs, M.; Brookes, A. J. *Genome Res.* **2002**, *12*, 1401.
- [40] Smigielski, E. M.; Sirotkin, K.; Ward, M.; Sherry, S. T. *Nucleic Acids Res.* **2000**, *28*, 352.
- [41] Matsuzaki, H.; Loi, H.; Dong, S.; Tsai, Y. Y.; Fang, J.; Law, J.; Di, X.; Liu, W. M.; Yang, G.; Liu, G.; Huang, J.; Kennedy, G. C.; Ryder, T. B.; Marcus, G. A.; Walsh, P. S.; Shriver, M. D.; Puck, J. M.; Jones, K. W.; Mei, R. *Genome Res.* **2004**, *14*, 414.
- [42] Hoehe, M. R. *Pharmacogenomics* **2003**, *4*, 547.
- [43] Velculescu, V. E.; Zhang, L.; Vogelstein, B.; Kinzler, K. W. *Science* **1995**, *270*, 484.
- [44] Brenner, S.; Johnson, M.; Bridgman, J.; Golda, G.; Lloyd, D. H.; Johnson, D.; Luo, S.; McCurdy, S.; Foy, M.; Ewan, M.; Roth, R.; George, D.; Eletr, S.; Albrecht, G.; Vermaas, E.; Williams, S. R.; Moon, K.; Burcham, T.; Pallas, M.; DuBridge, R. B.; Kirchner, J.; Fearon, K.; Mao, J.; Corcoran, K. *Nat. Biotechnol.* **2000**, *18*, 630.
- [45] Burgtorf, C.; Kepper, P.; Hoehe, M.; Schmitt, C.; Reinhardt, R.; Lehrach, H.; Sauer, S. *Genome Res.* **2003**, *13*, 2717.
- [46] Hardenbol, P.; Baner, J.; Jain, M.; Nilsson, M.; Namsaraev, E. A.; Karlin-Neumann, G. A.; Fakhrai-Rad, H.; Ronaghi, M.; Willis, T. D.; Landegren, U.; Davis, R. W. *Nat. Biotechnol.* **2003**, *5*, 1.
- [47] Oliphant, A.; Barker, D. L.; Stuelpnagel, J. R.; Chee, M. S. *Biotechniques* **2002**, *Suppl.*, 56-8, 60.
- [48] Holland, P. M.; Abramson, R. D.; Watson, R.; Gelfand, D. H. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 7276.
- [49] Sauer, S.; Gut, I. G. *J. Chrom. B Analyt. Technol. Biomed. Life Sci.* **2002**, *782*, 73.
- [50] Haff, L. A.; Smirnov, I. P. *Nucleic Acids Res.* **1997**, *25*, 3749.
- [51] Buetow, K. H.; Edmonson, M.; MacDonald, R.; Clifford, R.; Yip, P.; Kelley, J.; Little, D. P.; Strausberg, R.; Koester, H.; Cantor, C. R.; Braun, A. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 581.
- [52] Sauer, S.; Lehrach, H.; Reinhardt, R. *Nucleic Acids Res.* **2003**, *31*, e63.
- [53] Sauer, S.; Kepper, P.; Smyra, A.; Dahl, A.; Ferse, F-T; Lehrach, H.; Reinhardt, R. *J. Chrom. A* **2004**, *1049*, 9.
- [54] Lesage, S.; Zouali, H.; Cezard, J. P.; Colombel, J. F.; Belaiche, J.; Almer, S.; Tysk, C.; O'Morain, C.; Gassull, M.; Binder, V.; Finkel, Y.; Modigliani, R.; Gower-Rousseau, C.; Macry, J.; Merlin, F.; Chamailard, M.; Jannot, A. S.; Thomas, G.; Hugot, J. P. *Am. J. Hum. Genet.* **2002**, *70*, 845.
- [55] Walter, J.; Paulsen, M. *Semin. Cell Dev. Biol.* **2003**, *14*, 101.
- [56] Adorjan, P.; Distler, J.; Lipscher, E.; Model, F.; Muller, J.; Pelet, C.; Braun, A.; Florl, A. R.; Gutig, D.; Grabs, G.; Howe, A.; Kursar, M.; Lesche, R.; Leu, E.; Lewin, A.; Maier, S.; Muller, V.; Otto, T.; Scholz, C.; Schulz, W. A.; Seifert, H. H.; Schwöpe, I.; Ziebarth, H.; Berlin, K.; Piepenbrock, C.; Olek, A. *Nucleic Acids Res.* **2002**, *30*, e21.

- [57] Cottrell, S. E.; Distler, J.; Goodman, N. S.; Mooney, S. H.; Kluth, A.; Olek, A.; Schwoppe, I.; Tetzner, R.; Ziebarth, H.; Berlin, K. *Nucleic Acids Res.* **2004**, *32*, e10.
- [58] Chuaiqui, R. F.; Bonner, R. F.; Best, C. J.; Gillespie, J. W.; Flaig, M. J.; Hewitt, S. M.; Phillips, J. L.; Krizman, D. B.; Tangrea, M. A.; Ahram, M.; Linehan, W. M.; Knezevic, V.; Emmert-Buck, M. R. *Nat. Genet.* **2002**, *32*, 509.
- [59] Endrizzi, K.; Fischer, J.; Klein, K.; Schwab, M.; Nussler, A.; Neuhaus, P.; Eichelbaum, M.; Zanger, U. M. *Anal. Biochem.* **2002**, *300*, 121.
- [60] Biran, I.; Walt, D. R. *Anal. Chem.* **2002**, *74*, 3046.
- [61] Lizardi, P. M.; Huang, X.; Zhu, Z.; Bray-Ward, P.; Thomas, D. C.; Ward, D. C. *Nat. Genet.* **1998**, *19*, 225.
- [62] Angenendt, P.; Nyarsik, L.; Szaflarski, W.; Glokler, J.; Nierhaus, K. H.; Lehrach, H.; Cahill, D. J.; Lueking, A. *Anal. Chem.* **2004**, *76*, 1844.
- [63] Fire, A.; Xu, S.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C. *Nature* **1998**, *391*, 806.
- [64] Dorsett, Y.; Tuschl, T. *Nat. Rev. Drug Discov.* **2004**, *3*, 318.
- [65] Sohail, M.; Hohegger, H.; Klotzbucher, A.; Guellec, R. L.; Hunt, T.; Southern, E. M. *Nucleic Acids Res.* **2001**, *29*, 2041.
- [66] Elbashir, S. M.; Harborth, J.; Lendeckel, W.; Yalcin, A.; Weber, K.; Tuschl, T. *Nature* **2001**, *411*, 494.
- [67] Sohail, M.; Doran, G.; Riedemann, J.; Macaulay, V.; Southern, E. M. *Nucleic Acids Res.* **2003**, *31*, e38.
- [68] Yuan, B.; Latek, R.; Hossbach, M.; Tuschl, T.; Lewitter, F. *Nucleic Acids Res.* **2004**, *32*, W130.
- [69] Scacheri, P. C.; Rozenblatt-Rosen, O.; Caplen, N. J.; Wolfsberg, T. G.; Umayam, L.; Lee, J. C.; Hughes, C. M.; Shanmugam, K. S.; Bhattacharjee, A.; Meyerson, M.; Collins, F. S. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1892.
- [70] Persengiev, S. P.; Zhu, X.; Green, M. R. *RNA* **2004**, *10*, 12.
- [71] Bussow, K.; Quedenau, C.; Sievert, V.; Tischer, J.; Scheich, C.; Seitz, H.; Hieke, B.; Niesen, F. H.; Gotz, F.; Harttig, U.; Lehrach, H. *Genome Biol.* **2004**, *5*, 17.
- [72] Agaton, C.; Uhlen, M.; Hober, S. *Electrophoresis* **2004**, *25*, 1280.
- [73] Heinemann, U.; Frevert, J.; Hofmann, K.; Illing, G.; Maurer, C.; Oschkinat, H.; Saenger, W. *Prog. Biophys. Mol. Biol.* **2000**, *73*, 347.
- [74] Cherkasov, A.; Jones, S. J. *BMC Bioinformatics* **2004**, *5*, 37.
- [75] Gullberg, M.; Fredriksson, S.; Taussig, M.; Jarvius, J.; Gustafsdottir, S.; Landegren, U. *Curr. Opin. Biotechnol.* **2003**, *14*, 82.
- [76] Agaton, C.; Falk, R.; Hoiden Guthenberg, I.; Gostring, L.; Uhlen, M.; Hober, S. *J. Chrom. A* **2004**, *1043*, 33.
- [77] Kohler, G.; Milstein, C. *Nature* **1975**, *256*, 495.
- [78] Konthur, Z.; Walter, G. *Targets* **2002**, *1*, 30.
- [79] Binz, H. K.; Amstutz, P.; Kohl, A.; Stumpp, M. T.; Briand, C.; Forrer, P.; Grutter, M. G.; Pluckthun, A. *Nat. Biotechnol.* **2004**, *22*, 575.
- [80] Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. *Nature* **1992**, *355*, 564.
- [81] Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505.
- [82] Golden, M. C.; Collins, B. D.; Willis, M. C.; Koch, T. H. *J. Biotechnol.* **2000**, *81*, 167.
- [83] Cox, J. C.; Rajendran, M.; Riedel, T.; Davidson, E. A.; Sooter, L. J.; Bayer, T. S.; Schmitz-Brown, M.; Ellington, A. D. *Comb. Chem. High Throughput Screen.* **2002**, *5*, 289.
- [84] Drolet, D. W.; Jenison, R. D.; Smith, D. E.; Pratt, D.; Hicke, B. J. *Comb. Chem. High Throughput Screen.* **1999**, *2*, 271.
- [85] Bussow, K.; Cahill, D.; Nietfeld, W.; Bancroft, D.; Scherzinger, E.; Lehrach, H.; Walter, G. *Nucleic Acids Res.* **1998**, *26*, 5007.
- [86] Lueking, A.; Possling, A.; Huber, O.; Beveridge, A.; Horn, M.; Eickhoff, H.; Schuchardt, J.; Lehrach, H.; Cahill, D. J. *Mol. Cell Proteomics* **2003**, *2*, 1342.
- [87] Angenendt, P.; Glokler, J.; Murphy, D.; Lehrach, H.; Cahill, D. J. *Anal. Biochem.* **2002**, *309*, 253.
- [88] Angenendt, P.; Glokler, J.; Konthur, Z.; Lehrach, H.; Cahill, D. J. *Anal. Chem.* **2003**, *75*, 4368.
- [89] Angenendt, P.; Wilde, J.; Kijanka, G.; Baars, S.; Cahill, D. J.; Kreuzberger, J.; Lehrach, H.; Konthur, Z.; Glokler, J. *Anal. Chem.* **2004**, *76*, 2916.
- [90] Zhu, H.; Klemic, J. F.; Chang, S.; Bertone, P.; Casamayor, A.; Klemic, K. G.; Smith, D.; Gerstein, M.; Reed, M. A.; Snyder, M. *Nat. Genet.* **2000**, *26*, 283.
- [91] He, M.; Taussig, M. J. *Nucleic Acids Res.* **2001**, *29*, e73.
- [92] Giavalisco, P.; Nordhoff, E.; Lehrach, H.; Gobom, J.; Klose, J. *Electrophoresis* **2003**, *24*, 207.
- [93] Washburn, M. P.; Wolters, D.; Yates, J. R. 3rd. *Nat. Biotechnol.* **2001**, *19*, 242.
- [94] Fenn, J. B. *Angew Chem. Int. Ed. Engl.* **2003**, *42*, 3871.
- [95] Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299.
- [96] Mann, M.; Jensen, O. N. *Nat. Biotechnol.* **2003**, *21*, 255.
- [97] Klose, J.; Nock, C.; Herrmann, M.; Stuhler, K.; Marcus, K.; Bluggel, M.; Krause, E.; Schalkwyk, L. C.; Rastan, S.; Brown, S. D.; Bussow, K.; Himmelbauer, H.; Lehrach, H. *Nat. Genet.* **2002**, *30*, 385.
- [98] Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. *Nature* **2002**, *415*, 141.
- [99] Rigaut, G.; Shevchenko, A.; Rutz, B.; Wilm, M.; Mann, M.; Seraphin, B. *Nat. Biotechnol.* **1999**, *17*, 1030.
- [100] Bouwmeester, T.; Bauch, A.; Ruffner, H.; Angrand, P. O.; Bergamini, G.; Crougton, K.; Cruciat, C.; Eberhard, D.; Gagneur, J.; Ghidelli, S.; Hopf, C.; Huhse, B.; Mangano, R.; Michon, A. M.; Schirle, M.; Schlegl, J.; Schwab, M.; Stein, M. A.; Bauer, A.; Casari, G.; Drewes, G.; Gavin, A. C.; Jackson, D. B.; Joberty, G.; Neubauer, G.; Rick, J.; Kuster, B.; Superti-Furga, G. *Nat. Cell Biol.* **2004**, *6*, 97.
- [101] Fields, S.; Song, O. *Nature* **1989**, *340*, 245.
- [102] Goehler, H.; Lalowski, M.; Stelzl, U.; Waelder, S.; Stroedicke, M.; Worm, U.; Droege, A.; Lindenberg, K. S.; Knoblich, M.; Haenig, C.; Herbst, M.; Suopanki, J.; Scherzinger, E.; Abraham, C.; Bauer, B.; Hasenbank, R.; Fritzsche, A.; Ludewig, A. H.; Buessow, K.; Coleman, S. H.; Gutekunst, C. A.; Landwehrmeyer, B. G.; Lehrach, H.; Wanker, E. E. *Mol. Cell* **2004**, *15*, 853.
- [103] Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T. A.; Judson, R. S.; Knight, J. R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; Qureshi-Emili, A.; Li, Y.; Godwin, B.; Conover, D.; Kalbfleisch, T.; Vijayadamar, G.; Yang, M.; Johnston, M.; Fields, S.; Rothberg, J. M. *Nature* **2000**, *403*, 623.
- [104] Hu, J. C. *Trends Microbiol.* **2001**, *9*, 219.
- [105] Fredriksson, S.; Gullberg, M.; Jarvius, J.; Olsson, C.; Pietras, K.; Gustafsdottir, S. M.; Ostman, A.; Landegren, U. *Nat. Biotechnol.* **2002**, *20*, 473.
- [106] Ruano, G.; Kidd, K. K.; Stephens, J. C. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 6296.
- [107] Weiss, S. *Science* **1999**, *283*, 1676.
- [108] Woolley, A. T.; Guillemette, C.; Li Cheung, C.; Housman, D. E.; Lieber, C. M. *Nat. Biotechnol.* **2000**, *18*, 760.
- [109] Eigen, M.; Rigler, R. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 5740.
- [110] Schwille, P.; Meyer-Almes, F. J.; Rigler, R. *Biophys. J.* **1997**, *72*, 1878.
- [111] Landegren, U.; Schallmeiner, E.; Nilsson, M.; Fredriksson, S.; Baner, J.; Gullberg, M.; Jarvius, J.; Gustafsdottir, S.; Dahl, F.; Soderberg, O.; Ericsson, O.; Stenberg, J. *J. Mol. Recognit.* **2004**, *17*, 194.
- [112] Nilsson, M.; Antson, D. O.; Barbany, G.; Landegren, U. *Nucleic Acids Res.* **2001**, *29*, 578.
- [113] Kononen, J.; Bubendorf, L.; Kallioniemi, A.; Barlund, M.; Schraml, P.; Leighton, S.; Torhorst, J.; Mihatsch, M. J.; Sauter, G.; Kallioniemi, O. P. *Nat. Med.* **1998**, *4*, 844.
- [114] Ziauddin, J.; Sabatini, D. M. *Nature* **2001**, *411*, 107.
- [115] Mousses, S.; Caplen, N. J.; Cornelison, R.; Weaver, D.; Basik, M.; Hautaniemi, S.; Elkhoulou, A. G.; Lotufo, R. A.; Choudary, A.; Dougherty, E. R.; Suh, E.; Kallioniemi, O. *Genome Res.* **2003**, *13*, 2341.