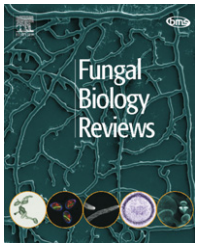
ELSEVIER

British Mycological Society promoting fungal science

journal homepage: www.elsevier.com/locate/fbr



Review

Of patterns and pathways: microarray technologies for the analysis of filamentous fungi

*Minou NOWROUSIAN**

Lehrstuhl für Allgemeine und Molekulare Botanik, Ruhr-Universität Bochum ND 6/165, Universitätsstrasse 150, DE-44801, Bochum

Keywords:
Expression pattern
Gene cluster
Microarray
Molecular phenotype
Regulatory pathway

ABSTRACT

During recent years, microarrays have been firmly established as valuable tools for the discovery of novel biological phenomena. Especially in combination with whole genome sequences, microarray data can help unravel the dynamics of the expressed genome. For filamentous fungi, microarray studies have already been performed with more than 20 different species; these investigations have explored a variety of different aspects of fungal biology. In this review, I will give an overview of some of the key questions that have been addressed using microarray hybridizations with filamentous fungi, with particular focus on the analysis of co-regulated pathways and physically clustered genes, as well as on the use of microarray data to determine a molecular phenotype. Additionally, a number of useful, freely available software tools for the analysis of fungal microarray data will be discussed.

1. Introduction

Microrarray technologies are high-throughput applications that allow the parallel hybridization of hundreds to thousands of carrier-bound DNA probes. In principle, microarrays are large-scale variations of reverse dot blots, meaning that the probe (usually a DNA fragment or oligonucleotide) is bound to or synthesized on a carrier (often glass slides or nylon membranes) and then a labeled DNA or RNA, the so-called target, is used for hybridization. Microarray hybridizations are widely used for expression analyses in which case the targets consist of reverse transcribed RNAs; however, there are also a number of other applications (Ehrenreich 2006; Nowrousian *et al.* 2004; Shannon & Rao 2002; Snijders *et al.* 2003). Although one of the first organisms to be studied using microarray techniques was a fungus, namely the yeast *Saccharomyces cerevisiae*, it took a while longer for microarray-based studies of other fungi, especially filamentous fungi, to get started. However, microarray technologies for filamentous fungi have now come of age, as probably best demonstrated by a recent review summarizing the first fifty microarray studies in filamentous fungi covering publications from 2002 to 2006 (Breakspear & Momany 2007). These studies deal with a variety of different aspects of fungal life, e.g. metabolism, pathogenesis, development, etc., and make use of array platforms for more than 20 different fungal species (Breakspear & Momany 2007).

Nowadays, there are a growing number of microarray data on filamentous fungi available in microarray-specific databases. These databases can be accessed and the information used by other researchers in their future experiments, similar to DNA sequence data that are now routinely used for sequence comparisons. Thus, data mining of microarrays might become a standard method performed by many more researchers than those actually doing hybridizations. Therefore, in this review, after a brief overview about microarray technologies, I will focus on the types of question about fungal

* Tel.: +49 234 3224588; fax: +49 234 3214184.
E-mail address: minou.nowrousian@ruhr-uni-bochum.de

1749-4613/$ – see front matter © 2007 The British Mycological Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.fbr.2007.09.002

biology that can be addressed with this technology as well as providing information on software resources for data mining that are publicly available. Readers interested in a more detailed description of the pre-data mining phase of a microarray experiment are referred to two excellent reviews on the subject (e.g. Churchill 2002; Ehrenreich 2006).

2. Microarray technologies for the analysis of differential gene expression: an overview

With filamentous fungi, microarrays have been used mostly for the analysis of gene expression (Breakspear & Momany 2007). The outline of a typical array experiment is depicted in Fig 1: strains are grown under the desired conditions, and then RNA is extracted from the mycelia. Samples for biological replicates should be grown independently to guard against artifacts that might arise in a one-off experiment. Targets are prepared by reverse transcription and are hybridized to the arrays. In the case of two-colour experiments that are, for example possible with cDNA microarrays, one target is labelled with one fluorescent dye, the second target with another dye, and both are hybridized to the same array slide. Two-colour hybridizations can either be performed in the form of a direct design (loop design) or with a common reference target (Fig 1). With a direct design, the number of array slides needed is similar to the number of targets, whereas when using a common reference, more slides are necessary (Churchill 2002). Thus, a direct design saves material and can be easily analyzed e.g. with Bioconductor - see below; thus, direct design is the method of choice for larger two-color hybridization experiments. However, not all analysis software tools can accommodate direct designs very well. Therefore, if it is necessary to use a specific software package, one should ensure in advance that the hybridization design can be integrated into the software. Other types of microarrays such as Affymetrix arrays do not allow two-colour experiments; here, each target is hybridized to a unique array.

After hybridization, the arrays are scanned and raw fluorescence data are obtained from the scanned images. These have to be background corrected and normalized to correct for errors due to unequal starting amounts of RNA, sample loss, dye bias, etc. (Huber *et al.* 2002; Quackenbush 2002). After the appropriate data transformations have taken place, it is finally possible to look for differentially expressed genes. In early array experiments, researchers often used two-fold up- or down-regulation, sometimes in combination with thresholds for mean and standard deviation of replicate experiments, as a criterion for differential expression. However, such an approach might lead to high false-discovery rates, especially at low intensities where the data are much more variable. Therefore, a number of statistical tests were developed that take into account the variabilities within the data structure (Cui & Churchill 2003; Quackenbush 2002; Smyth 2004). These statistical methods test the hypothesis that a gene is not differentially expressed (null hypothesis), and their output is not only a ratio or log ratio of differential expression, but also a p-value giving the probability that the null hypothesis was falsely rejected, i.e. that the gene is not differentially expressed. Thus, the genes with the lowest *p*-value have the highest probability of really being differentially expressed. Current software packages for microarray data analysis usually incorporate a number of different statistical tests (see below). These software tools usually deliver lists of differentially expressed genes that in themselves are not very informative, but can be used as a starting point for further analysis as described below.

3. Which questions can be addressed using microarray technologies?

Most microarray studies involve the comparison of more than two targets, and in such cases, it is not only of interest which genes are differentially expressed in any of the investigated samples, but also which genes show the same or different expression patterns in a number of samples. Thus, the next step in microarray data analysis usually involves the clustering of genes with similar expression patterns (D'haeseleer 2005). In addition, the targets themselves can also be clustered according to their overall expression patterns to determine samples (e.g. mutant strains or different growth conditions), that display similar expression on a genomic basis (Slonim 2002). These clusters of genes or targets with similar expression can be used to address several types of biological questions, and three aspects that have already been investigated in filamentous fungi will be the focus of this review (Fig 1, lower panels): Microarrays were used to: (1) determine regulatory or metabolic pathways that are differentially regulated in certain strains or under certain conditions, (2) identify genes that are not only co-regulated but also physically clustered within the genome, and (3) compare different targets (e.g. strains) with respect to their overall expression patterns in order to determine relationships among strains or conditions (i.e. to determine a molecular phenotype).

Microarray analyses to investigate regulatory or biochemical pathways

Genes encoding proteins that participate in common regulatory or biochemical pathways often display similar expression patterns at the transcriptional level. In filamentous fungi, a number of microarray studies have demonstrated a concerted transcriptional expression of metabolic genes from various pathways of primary and secondary metabolism (Breakspear & Momany 2007). Studies like these can be used to determine whether a specific pathway is essential under certain conditions. For example, the glyoxylate cycle was found to be necessary for full virulence in the dimorphic yeast *Candida albicans* (Lorenz & Fink 2001), and similar studies with filamentous fungi will certainly be forthcoming in the future.

Additionally, microarray studies can be used to identify novel genes that play a role in certain biological processes by virtue of their expression pattern. This was demonstrated in an analysis of glucose-regulated gene expression in *Neurospora crassa* where a previously uncharacterized putative transporter gene, *hgt-1*, was found to be glucose-repressed. Further analysis of *hgt-1* demonstrated that it encodes a high-affinity glucose transporter (Xie *et al.* 2004). Another example is the identification of a novel gene, CA747470, involved in the regulation of aflatoxin biosynthesis in *Aspergillus flavus*. CA747470

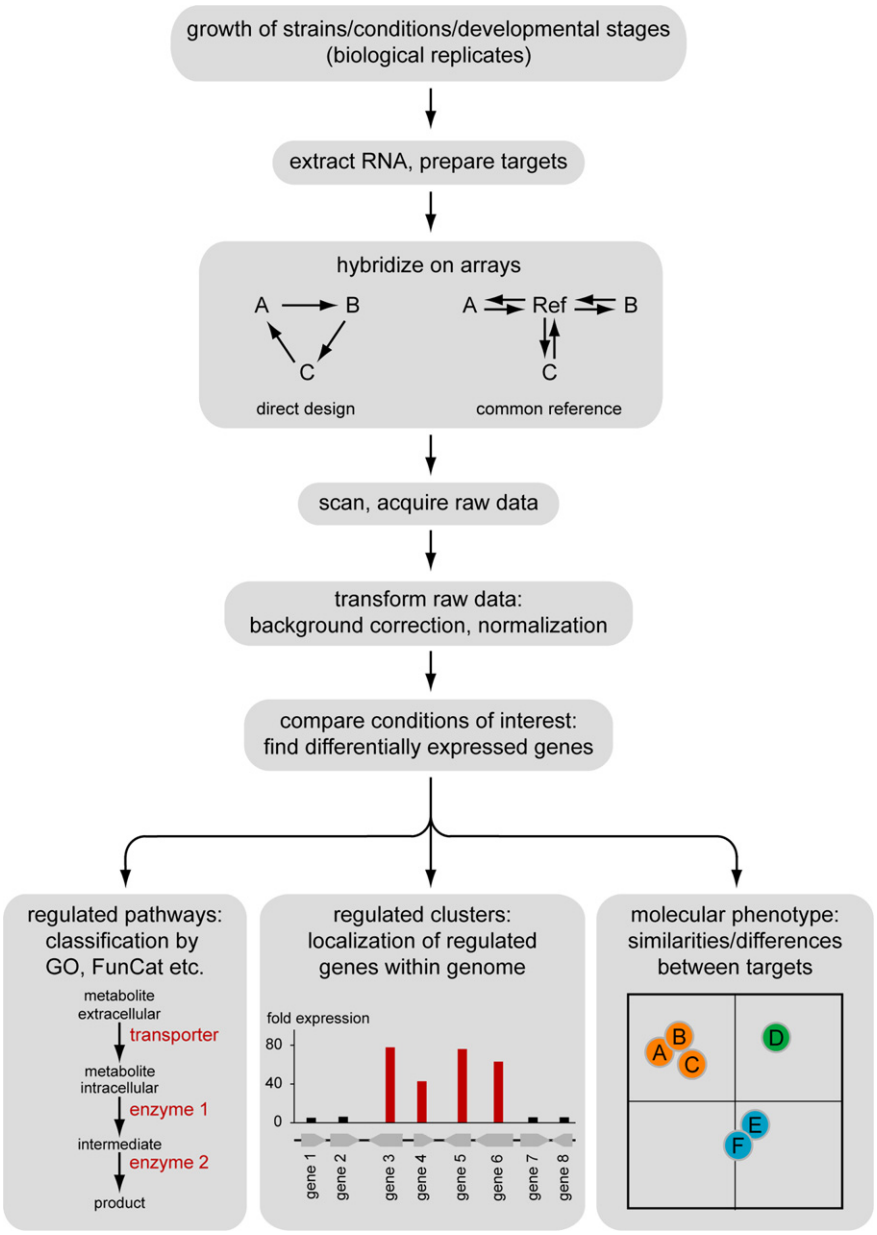


Fig. 1 – An overview of gene expression analyses with microarrays. Samples are grown under the conditions of interest with at least two biological replicates for each experimental setup. RNA is extracted and targets are prepared by reverse transcription and sometimes an amplification step. Targets are then hybridized to the arrays. For two-color hybridizations, experiments can be organized as direct design or common reference design; each arrow represents an array with the target (A, B, and C) at the base being labelled by one dye and the target at the tip by the second dye. After hybridization, arrays are scanned and the resulting images are processed to acquire raw data. These have to be background corrected and normalized before differentially expressed genes can be identified. The information about which genes are differentially expressed under which conditions can be used for downstream analyses: genes can be clustered according to (putative) function to identify biochemical or signaling pathways that are co-regulated (lower left panel). Furthermore, gene expression data can be linked with information about genome organization to analyze common regulation patterns of clustered genes (lower middle panel). Also, gene expression patterns can be compared across targets to identify conditions with similar gene expression patterns (lower right panel, targets A-F). In this graphical representation, targets with similar expression patterns cluster together in a two-dimensional "expression space". For more information, see text.

displays no significant homology to any known gene; however, its expression was found to be inversely correlated with aflatoxin production, and overexpression of CA747470 itself led to a reduction in the amount of aflatoxin produced by the transformants (Price *et al.* 2005). In addition to the analysis of different physiological conditions, the transcriptional profiling of mutant strains is a promising approach for the identification of genes that play a role in specific biological

processes. This was recently demonstrated with a mutant of the transcription factor-encoding gene *CON7* that is essential for appressorium formation in *Magnaporthe grisea* (Shi *et al.* 1998). In the mutant strain, about 100 genes were identified that were downregulated compared to the wild-type strain, and many of these encode proteins predicted to be involved in cell wall construction. One of these genes, the chitin synthase gene *CHS7*, was shown to be involved in appressorium formation itself, thereby linking the regulatory gene *CON7* and the downstream gene *CHS7* to the morphogenesis of infection structures in *M. grisea* (Odenbach *et al.* 2007).

The analysis of expression profiles from transcription factor mutants can also be used to search for putative *cis*-elements provided that the genomic regions upstream of open reading frames are known as is the case for a number of sequenced and annotated fungal genomes. One example for such a strategy is the analysis of putative binding sites for the transcription factor CPC1 that is involved in the regulation of amino acid biosynthesis genes in *N. crassa*. First, microarray analysis was used to identify putative target genes of CPC1, then the upstream regions of these genes were screened for enrichment of a known CPC1-binding site. Genes involved in amino acid biosynthesis were strongly enriched in the group of genes with a fully conserved consensus sequence in their upstream region making them likely targets for a direct regulation by CPC1 (Tian *et al.* 2007).

Microarray analyses to identify co-regulated, physically clustered genes

In contrast to prokaryotes where the physical linkage of genes that participate in the same biological process is common, this phenomenon is rarely found in eukaryotes. One exception are clusters containing genes for pathways of the secondary and sometimes primary metabolism of filamentous fungi (Keller & Hohn 1997). Such clusters can be identified by homology of their genes to other previously known metabolism genes; however, homology alone does not provide evidence for an actively expressed cluster. Furthermore, clusters without homologs will go undetected in sequence comparison analyses. Therefore, one way to find putative functional clusters is to determine whether the genes in the cluster are co-regulated at the level of transcription; and array hybridization experiments of whole genome microarrays are extremely well suited for this task. This was demonstrated in analyses of *Aspergillus nidulans* and *Aspergillus fumigatus* mutants lacking the transcriptional regulator LaeA, a methyltransferase that regulates a number of secondary metabolite clusters in several Aspergillus species (Bok & Keller 2004; Bok *et al.* 2006b). Microarray analyses of the *A. nidulans* wild type and an *laeA* deletion mutant were used to identify physically clustered genes that require *laeA* for their concerted expression. Among the clusters identified was one that encodes enzymes for the biosynthesis of terrequinon A, an antitumour compound that had not been previously identified in *A. nidulans* (Bok *et al.* 2006a).

Aspergillus fumigatus is a saprotrophic fungus that can lead to invasive aspergillosis in immunocompromised patients, and secondary metabolites like mycotoxins and melanins have been implicated in the virulence of this Aspergillus species. In a recent microarray study, it was found that LaeA regulates more than half of the 22 secondary metabolite clusters that are present in the *A. fumigatus* genome (Perrin *et al.* 2007). Additionally, *laeA* itself was shown to be necessary for pathogenicity (Bok *et al.* 2005). Thus, one can conclude that the combination of secondary metabolites that are produced by *A. fumigatus* in the presence of LaeA might be a prerequisite for virulence, making LaeA a promising target for the development of antifungal drugs (Perrin *et al.* 2007).

Another case of clustered, co-expressed genes was recently found in the dimorphic basidiomycete *Ustilago maydis* (Kämper *et al.* 2006). The genome of this phytopathogenic fungus contains a large number of genes that are predicted to encode secreted proteins, and of these genes, nearly 20 % are arranged within 12 clusters within the genome. Microarray analyses showed that most of the genes in all the clusters are upregulated during tumour formation *in planta* indicating that they might be involved in plant infection by *U. maydis*. Deletion of each cluster revealed several clusters that affect fungal virulence, with four cluster mutants being less pathogenic and one showing a phenotype of hypervirulence (Kämper *et al.* 2006). The molecular function of the products of these co-expressed genes is as yet unknown; however, the results confirm the value of screening for clustered, co-regulated genes as a means for identifying biologically relevant factors in fungi.

Microarray analyses to determine a molecular phenotype

As mentioned above, microarray data can not only be used to identify regulated genes, but also to compare overall expression patterns of targets (e.g. different strains or conditions). Gene expression patterns constitute a phenotype, and similar to morphogenetic or physiological traits they can be used to determine genetic relationships between genes (e.g. epistasis, Zupan *et al.* 2003). For this purpose, overall gene expression of mutant and wild-type strains under the conditions of interest is compared, and targets are clustered according to their expression patterns (Slonim 2002). This approach has been used to investigate developmental mutants of the slime mould *Dictyostelium discoideum* as well as mutants of the mediator complex from *Saccharomyces cerevisiae* (van de Peppel *et al.* 2005; van Driessche *et al.* 2005). In filamentous fungi, this approach was used to study lovastatin-producing strains of *Aspergillus terreus*. Here, transcriptional profiling was combined with secondary metabolite analysis, and it turned out that strains with similar metabolite profiles also tended to have similar transcriptional profiles. The information gained by this integrated approach was then applied to develop strain improvement strategies for lovastatin production (Askenazi *et al.* 2003). In the ascomycete *Sordaria macrospora*, the clustering of targets was used succesfully to determine the genetic relationship between the developmental genes *pro1* and *pro41*. The corresponding mutants pro1 and pro41 as well as the double mutant pro1/41 are morphologically similar; therefore no conclusion with respect to a possible epistatic relationship could be reached in an analysis of morphological phenotypes. Thus, gene expression patterns of the single mutants and the double mutant pro1/41 were used as a molecular phenotype. The expression pattern of the double mutant was more similar to that of the pro41 mutant than that of the pro1 mutant, indicating that *pro41* most likely acts downstream of

pro1 in a genetic network (Nowrousian *et al.* 2007). Approaches like these will certainly become increasingly common, especially with the advent of whole genome microarrays for a number of fungal species, because expression profiling across a whole genome results in a molecular phenotype that comprises expression data for thousands of genes and as a whole is less likely to suffer from random variation.

Other questions that can be answered using microarray analyses

Apart from the experimental approaches discussed above, microarrays offer a number of additional possibilities to address problems in fungal biology. One aspect worth mentioning is that microarrays can not only be used for hybridizations with targets derived from the same species as the probes on the arrays, but that hybridization with targets from closely related species is often possible. The success of this so-called cross-species hybridization depends on (1) the phylogenetic distance and thus the remaining sequence identity of the two species under investigation, and (2) on the type of array that is used (Bar-Or *et al.* 2007). cDNA microarrays, where the probes are usually several hundred nucleotides in length, yield better results than arrays with short oligonucleotide probes. However, given closely related species that share a high degree of sequence similarity, good results can even be achieved using oligonucleotide arrays as was demonstrated in cross-species hybridization experiments of *N. crassa* whole genome oligonucleotide arrays with targets derived from *S. macrospora* (Nowrousian *et al.* 2007).

Another application of cross-species microarray hybridizations are comparative genomic hybridization (CGH) experiments. For array-CGH, fragments of genomic DNA from different strains or species are hybridized to microarrays, and the results can be used to measure genomic divergence or to identify conserved *versus* rapidly evolving genes. For filamentous fungi, CGH has been used to compare several *A. fumigatus* strains and closely related species as well as for an analysis of strains from the genus Paxillus (Le Quéré *et al.* 2006; Nierman *et al.* 2005). In the first study, array-CGH was performed with *A. fumigatus* strain Af293 that was also used for genome sequencing, as well as two additional *A. fumigatus* strains and the closely related species *Aspergillus clavatus*, *Neosartorya fischeri*, and *Neosartorya fenneliae*. In particular, the comparison with *N. fischeri*, a nonpathogenic relative of *A. fumigatus*, revealed about 700 genes that are absent or diverged compared to *A. fumigatus*. Such genes are candidates for future searches for pathogenicity-related factors in *A. fumigatus* (Nierman *et al.* 2005). In another array-CGH analysis, several strains from the ectomycorrhizal basidiomycete *Paxillus involutus* as well as its close relative *Paxillus filamentosus* were compared using a *P. involutus* cDNA microarray. The strains were selected as they have different host plant specificities. Among the genes that showed a high rate of divergence, orphan genes and genes whose products are involved in stress response/defense reactions or are localized at membranes were enriched, probably indicating host-specific adaptation processes of ectomycorrhizal fungi (Le Quéré *et al.* 2006).

One further aspect that will become even more important with the increasing amount of microarray data is the comparison of expression profiles between species (comparative functional genomics). Genes that display conserved expression patterns in different species can help to define a core group of factors that might be involved in the process under investigation. Such genes would be candidates of choice for further analysis, because evolutionary conservation of expression patterns is a powerful criterion to identify genes that might be functionally important (Stuart *et al.* 2003). In the first studies involving filamentous fungi, microarray data on glucose metabolism from *N. crassa* and *Trichoderma reesei*, respectively, were compared to the corresponding results from *S. cerevisiae* (Chambergo *et al.* 2002; Xie *et al.* 2004). These data showed some similarities in expression patterns of genes involved in glycolysis and respiration; however, it was also possible to uncover species-specific differences that correlated well with known physiological reactions of the different fungi to glucose availability.

In a more recent study, a three-way comparison of expression data from *N. crassa*, *C. albicans*, and *S. cerevisiae* was performed (Tian *et al.* 2007). Here, the authors used array results from wild-type strains and mutants in the transcription factor gene *cpc-1* and its yeast orthologs that were grown under conditions of amino acid starvation. This analysis identified a group of 32 orthologous genes that define a core group of amino acid starvation response genes. When looking not only at the regulation of orthologous genes but whole functional groups of genes such as genes involved in amino acid or sulfur metabolism, a much larger overlap was found in the expression patterns of the three species. This might indicate that while the overall expression of functional categories has to be maintained, functionality can be achieved by a wide variety of regulation patterns at the level of the single gene (Tian *et al.* 2007). Further comparative functional genomics studies including larger numbers of fungal species will soon be possible and will lead to even more robust data as microarrays gain their strength by numbers (i.e. the more experiments are used for data mining, the more significant the analysis usually gets).

There are several more technologies that make use of microarrays. For example, the so-called ChIP-chip assays combine chromatin immunoprecipitation with microarray analyses, thereby enabling genome-wide screens for sequences that are bound by DNA-binding proteins (Shannon & Rao 2002). However, such methods often require specialized arrays that contain not only sequences representing open reading frames but intergenic regions as well. As most of these methods have not yet been applied to filamentous fungi, they will not be covered in this review. Interested readers are refered to the yeast literature where these approaches are described in more detail (e.g. Clark *et al.* 2002; Iyer *et al.* 2001; Ren *et al.* 2000).

4. Software tools for microarray analyses of filamentous fungi

Due to the growing number of microarray data available in public databases, efficient data mining is becoming an increasingly important task for many researchers. There are several programs commercially available that allow the organization and analysis of array data; however, most of these

software packages are relatively expensive. Fortunately, there are also quite a few freely available software tools that have been developed mostly by groups from research institutes, and that often perform as well as or even better than commerially available ones. An overview of some of the freely available software tools is given in Fig 2. Especially two software packages have already been used for the analysis of transcriptional profiling experiments with whole genome-microarrays of filamentous fungi, namely Bioconductor and TM4 (Gentleman *et al.* 2005; Saeed *et al.* 2003). Both allow full-scale array data analyses including preprocessing and normalization procedures, the analysis of differentially expressed genes, and tools for the analysis of regulated pathways etc. (Fig 2). The TM4 suite also includes a program to analyze array images (Spotfinder) and a database tool (MADAM) to store and organize microarray data (Saeed *et al.* 2003). For filamentous fungi, TM4 has already been used in several analyses of whole genome-microarray data from *A. fumigatus* (Nierman *et al.* 2005; Perrin *et al.* 2007).

The particular strength of Bioconductor lies in its versatility that allows the incorporation of all types of microarray data and includes a huge number of different statistical tests and analyses (Gentleman *et al.* 2005). Its ''limma package'' is especially suited for the analysis of complex array hybridization designs (e.g. direct designs, Smyth 2004) and has been successfully used in experiments with whole genome arrays from *N. crassa* and *F. graminearum* (Hallen *et al.* 2007; Nowrousian *et al.* 2007). Bioconductor was also used for downstream analyses of fungal array data e.g. to determine the statistical significance of the enrichment of *cis*-elements in the upstream regions of differentially regulated genes from *N. crassa* Tian *et al.* (2007). Another example is the application of correspondence analysis to

A. Array image analysis software

ScanAlyze (http://rana.lbl.gov/EisenSoftware.htm)

Spotfinder (http://www.tm4.org/)

B. Array data analysis software

Software	backgr., norm.	diff. expr. genes	clustering of genes/ targets	pathway analysis etc.	graphics	pros	cons
Bioconductor [1]	+	+ complex designs	+	+	+	very versatile, powerful statistics	not intuitive, takes time to learn
Cluster/Treeview [2]	-	-	+	-	+	straightforward	limited analysis/graphics
Expression Profiler [3]	+	+ simple designs	+	+	+	online tool, no installation	slow, not easily adapted for complex experiment designs
TM4 suite [4]	+	+ simple designs	+	+	+	windows-based design	not easily adapted for complex experiment designs

[1] http://www.bioconductor.org, [2] http://rana.lbl.gov/EisenSoftware.htm, [3] http://www.ebi.ac.uk/expressionprofiler, [4] http://www.tm4.org

C. Array databases

ArrayExpress (EBI, http://www.ebi.ac.uk/arrayexpress/)

GEO (Gene Expression Omnibus, NCBI, http://www.ncbi.nlm.nih.gov/projects/geo/)

MGOS (*Magnaporthe grisea Oryza sativa* interaction database, http://www.mgosdb.org/microarray)

Neurospora Functional Genomics Microarray database (http://www.yale.edu/townsend/Links/ffdatabase/introduction.htm)

PLEXdb (Plant expression database, contains Fusarium data, http://www.plexdb.org)

D. Other useful software tools / web sites

Blastlocal (http://www.ncbi.nlm.nih.gov/blast/download.shtml)
Blastcl3 (netblast, http://www.ncbi.nlm.nih.gov/blast/download.shtml)

Tools to run batch sequence comparisons of multiple sequences with BLAST against local databases, e.g. a fungal genome (blastlocal), or against NCBI-based databases (blastcl3).

FunCatDB (The Functional Catalogue, http://mips.gsf.de/genre/proj/ncrassa/Search/Catalogs/catalog.jsp)
GO (Gene Ontology, http://www.geneontology.org/)
KEGG (Kyoto encyclopedia of genes and genomes, http://www.genome.jp/kegg)

Annotation databases, contain information about functional categories of genes, can be searched using gene identification numbers (may be database specific, FunCatDB can be searched with gene numbers for *N. crassa* and *F. graminearum*) or sequences

MatInspector (http://www.genomatix.de/products/MatInspector/MatInspector1.html)

Tool for screening putative promotor sequences for (common) binding sites, limited free access for academic users

Fig. 2 – Some freely available software tools for microarray analysis. This (non-comprehensive) list gives some publicly available tools for different stages of microarray analysis. A. Programs that can be used for the analysis of images from two-color arrays to obtain raw data for downstream applications. B. Raw data can be processed by several freely available software packages, many of which allow a complete analysis from background correction and normalization to meta-analyses like pathway annotation. C. Databases that contain publicly available microarray data. D. Software tools for batch sequence comparisons, for databases that contain annotation data like (putative) cellular functions of genes, and for databases of putative transcription factor binding sites.

expression data from *S. macrospora* mutants (Nowrousian *et al.* 2007). Correspondence analysis is a computational method that allows the study of associations between variables, in this case between genes and targets. It groups both targets and genes according to the similarity of their expression patterns. Thus, the results not only show which samples (targets) are similar to each other, but also which genes ''make the difference'', i.e. which genes are most differentially expressed between samples (Culhane *et al.* 2005; Fellenberg *et al.* 2001). Bioconductor offers this and many other downstream analyses; however, its command line-based structure is not intuitive for most biologists. Nevertheless, its great versatility and statistical power make the effort of learning it worthwhile.

Another software tool that is of interest for the analysis of fungal microarrays is FunCat (Ruepp *et al.* 2004). Similar to some other annotation systems and databases (Fig 2D), it classifies genes in a hierarchical system according to their (putative) cellular function. In addition, it offers the possibility to annotate lists of genes from annotated fungal genomes (at present *S. cerevisiae*, *N. crassa*, *F. graminearum*, and *U. maydis*). Thus, a list of genes that are differentially regulated can be used as input on the FunCat web site, and the resulting output associates each gene with its predicted functional category. Additionally, FunCat gives information about whether a group of functionally related genes is overrepresented in the input list (i.e. among the differentially regulated genes vs. the whole genome). FunCat is a very useful tool that has been applied to array data from *N. crassa* and *F. graminearum* in two recent studies (Hallen *et al.* 2007; Tian *et al.* 2007).

There is an ever increasing number of programs to deal with the vast amount of data generated by microarray hybridizations. In this review, I have covered only a small fraction of them; however, that does not mean that others programs not mentioned are not equally valuable for the analysis of fungal array data. Finally, it is important for future analyses to have full sets of raw and processed microarray data publicly available. To this end, there are several general microarray databases like ArrayExpress and GEO as well as fungal-specific databases (Fig 2C) that will be valuable resources for functional genomics and systems biology approaches with filamentous fungi in the future.

Acknowledgements

I would like to thank Prof. Dr. U. Kück for his support, Dr. Birgit Hoff, Ines Engh, and Jens Kamerewerd for helpful comments on the manuscript, and the European Molecular Biology Organization for giving me the opportunity to participate in the EMBO course ''Advanced practical course on analysis and bioinformatics of microarray data''.

REFERENCES

Askenazi M, Driggers EM, Holtzman DA, Norman TC, Iverson S, Zimmer DP, Boers ME, Blomquist PR, Martinez EJ, Monreal AW, Feibelman TP, Mayorga ME, Maxon ME, Sykes K, Tobin JV, Cordero E, Salama SR, Trueheart J, Royer JC, Madden KT, 2003. Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. *Nature Biotechnology* **21**: 150–156.

Bar-Or C, Czosnek H, Koltai H, 2007. Cross-species microarray hybridizations: a developing tool for studying species diversity. *Trends in Genetics* **23**: 200–207.

Bok JW, Balajee SA, Marr KA, Andes D, Nielsen KF, Frisvad JC, Keller NP, 2005. LaeA, a regulator of morphogenetic fungal virulence factors. *Eukaryotic Cell* **4**: 1574–1582.

Bok JW, Hoffmeister D, Maggio-Hall LA, Murillo R, Glasner JD, Keller NP, 2006a. Genomic mining for *Aspergillus* natural products. *Chemistry and Biology* **13**: 31–37.

Bok JW, Keller NP, 2004. LaeA, a regulator of secondary metabolism in *Aspergillus* spp. *Eukaryotic Cell* **3**: 527–535.

Bok JW, Noordermeer D, Kale SP, Keller NP, 2006b. Secondary metabolic gene cluster silencing in *Aspergillus nidulans*. *Molecular Microbiology* **61**: 1636–1645.

Breakspear A, Momany M, 2007. The first fifty microarray studies in filamentous fungi. *Microbiology* **153**: 7–15.

Chambergo FS, Bonaccorsi ED, Ferreira AJ, Ramos AS, Ferreira Junior JR, Abrahao-Neto J, Farah JP, El-Dorry H, 2002. Elucidation of the metabolic fate of glucose in the filamentous fungus *Trichoderma reesei* using expressed sequence tag (EST) analysis and cDNA microarrays. *Journal of Biological Chemistry* **277**: 13983–13988.

Churchill GA, 2002. Fundamentals of experimental design for cDNA microarrays. *Nature Genetics Supplement* **32**: 490–495.

Clark TA, Sugnet CW, Ares MJ, 2002. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* **296**: 907–910.

Cui X, Churchill GA, 2003. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biology* **4**: 210.

Culhane AC, Thioulouse J, Perriere G, Higgins DG, 2005. MADE4: an R package for multivariate analysis of gene expression data. *Bioinformatics* **21**: 2789–2790.

D'haeseleer P, 2005. How does gene expression clustering work? *Nature Biotechnology* **23**: 1499–1501.

Ehrenreich A, 2006. DNA microarray technology for the microbiologist: an overview. *Applied Microbiology and Biotechnology* **73**: 255–273.

Fellenberg K, Hauser NC, Brors B, Neutzner A, Hoheisel JD, Vingron M, 2001. Correspondence analysis applied to microarray data. *Proceedings of the National Academy of Sciences of the United States of America* **98**: 10781–10786.

Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S (eds), 2005. *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer, Heidelberg, New York.

Hallen HE, Huebner M, Shiu SH, Güldener U, Trail F, 2007. Gene expression shifts during perithecium development in *Gibberella zeae*(anamorph *Fusarium graminearum*), with particular emphasis on ion transport proteins. *Fungal Genetics and Biology* **44**: 1146–1156.

Huber W, von Heydebreck A, Sültmann H, 2002. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18**: S96–S104.

Iyer VR, Horak CE, Scafek CS, Botstein D, Snyder M, Brown PO, 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**: 533–538.

Kämper J, Kahmann R, Bölker M, Ma LJ, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Müller O, Perlin MH, Wösten HA, de VR, Ruiz-Herrera J, Reynaga-Pena CG, Snetselaar K, McCann M, Pérez-Martin J, Feldbrügge M, Basse CW, Steinberg G, Ibeas JI, Holloman W, Guzman P, Farman M, Stajich JE, Sentandreu R, González-Prieto JM,

Kennell JC, Molina L, Schirawski J, Mendoza-Mendoza A, Greilinger D, Münch K, Rössel N, Scherer M, Vranes M, Ladendorf O, Vincon V, Fuchs U, Sandrock B, Meng S, Ho EC, Cahill MJ, Boyce K, Klose J, Klosterman SJ, Deelstra HJ, Ortiz-Castellanos L, Li W, Sanchez-Alonso P, Schreier PH, Häuser-Hahn I, Vaupel M, Koopmann E, Friedrich G, Voss H, Schlüter T, Margolis J, Platt D, Swimmer C, Gnirke A, Chen F, Vysotskaia V, Mannhaupt G, Güldener U, Münsterkötter M, Haase D, Oesterheld M, Mewes HW, Mauceli EW, DeCaprio D, Wade CM, Butler J, Young S, Jaffe DB, Calvo S, Nusbaum C, Galagan J, Birren BW, 2006. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* **444**: 97–101.

Keller NP, Hohn TM, 1997. Metabolic pathway gene clusters in filamentous fungi. *Fungal Genetics and Biology* **21**: 17–29.

Le Quéré A, Eriksen KA, Rajashekar B, Schützendübel A, Canbäck B, Johansson T, Tunlid A, 2006. Screening for rapidly evolving genes in the ectomycorrhizal fungus *Paxillus involutus* using cDNA microarrays. *Molecular Ecology* **15**: 535–550.

Lorenz MC, Fink GR, 2001. The glyoxylate cycle is required for fungal virulence. *Nature* **412**: 83–86.

Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, Berriman M, Abe K, Archer DB, Bermejo C, Bennett J, Bowyer P, Chen D, Collins M, Coulsen R, Davies R, Dyer PS, Farman M, Fedorova N, Fedorova N, Feldblyum TV, Fischer R, Fosker N, Fraser A, Garcia JL, Garcia MJ, Goble A, Goldman GH, Gomi K, Griffith-Jones S, Gwilliam R, Haas B, Haas H, Harris D, Horiuchi H, Huang J, Humphray S, Jimenez J, Keller N, Khouri H, Kitamoto K, Kobayashi T, Konzack S, Kulkarni R, Kumagai T, Lafton A, Latge J-P, Li W, Lord A, Lu C, Majoros WH, May GS, Miller BL, Mohamoud Y, Molina M, Monod M, Mouyna I, Mulligan S, Murphy L, O'Neil S, Paulsen I, Penalva MA, Pertea M, Price C, Pritchard BL, Quail MA, Rabbinowitsch E, Rawlins N, Rajandream M-A, Reichard U, Renauld H, Robson GD, de Cordoba SR, Rodriguez-Pena JM, Ronning CM, Rutter S, Salzberg SL, Sanchez M, Sanchez-Ferrero JC, Saunders D, Seeger K, Squares R, Squares S, Takeuchi M, Tekaia F, Turner G, de Aldana CRV, Weidman J, White O, Woodward J, Yu J-H, Fraser C, Galagan JE, Asai K, Machida M, Hall N, Barrell B, Denning DW, 2005. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature* **438**: 1151–1156.

Nowrousian M, Dunlap JC, Nelson MA, 2004. Functional genomics in fungi. In: Kück U (ed), *The Mycota II*. Springer, Berlin, Heidelberg, pp. 115–128.

Nowrousian M, Frank S, Koers S, Strauch P, Weitner T, Ringelberg C, Dunlap JC, Loros JJ, Kück U, 2007. The novel ER membrane protein PRO41 is essential for sexual development in the filamentous fungus *Sordaria macrospora*. *Molecular Microbiology* **64**: 923–937.

Odenbach D, Breth B, Thines E, Weber RWS, Anke H, Foster AJ, 2007. The transcription factor Con7p is a central regulator of infection-related morphogenesis in the rice blast fungus *Magnaporthe grisea*. *Molecular Microbiology* **64**: 293–307.

Perrin RM, Fedorova ND, Bok JW, Cramer RA, Wortman JR, Kim HS, Nierman WC, Keller NP, 2007. Transcriptional regulation of chemical diversity in *Aspergillus fumigatus* by LaeA. *PLoS Pathogens* **3**: e50.

Price MS, Conners SB, Tachdjian S, Kelly RM, Payne GA, 2005. Aflatoxin conducive and non-conducive growth conditions reveal new gene associations with aflatoxin production. *Fungal Genetics and Biology* **42**: 506–518.

Quackenbush J, 2002. Microarray data normalization and transformation. *Nature Genetics Supplement* **32**: 496–501.

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA, 2000. Genome-wide location and function of DNA binding proteins. *Science* **290**: 2306–2309.

Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, Tetko I, Güldener U, Mannhaupt G, Münsterkötter M, Mewes HW, 2004. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Research* **32**: 5539–5545.

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J, 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**: 374–378.

Shannon MF, Rao S, 2002. Transcription. Of chips and ChIPs. *Science* **296**: 666–669.

Shi Z, Christian D, Leung H, 1998. Interactions between spore morphogenetic mutations affect cell types, sporulation, and pathogenesis in *Magnaporthe grisea*. *Molecular Plant-Microbe Interactions* **11**: 199–207.

Slonim DK, 2002. From patterns to pathways: gene expression analysis comes of age. *Nature Genetics Supplement* **32**: 502–508.

Smyth GK, 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3** Article 3.

Snijders AM, Pinkel D, Albertson DG, 2003. Current status and future prospects of array-based comparative genomic hybridisation. *Briefings in Functional Genomics and Proteomics* **2**: 37–45.

Stuart JM, Segal E, Koller D, Kim SK, 2003. A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**: 249–255.

Tian C, Kasuga T, Sachs MS, Glass NL, 2007. Transcriptional profiling of cross pathway control in *Neurospora crassa*and comparative analysis of the Gcn4 and CPC1 regulons. *Eukaryotic Cell* **6**: 1018–1029.

van de Peppel J, Kettelarij N, van Bakel H, Kockelkorn TTJP, van Leenen D, Holstege FCP, 2005. Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets. *Molecular Cell* **19**: 511–522.

van Driessche N, Demsar J, Booth EO, Hill P, Juvan P, Zupan B, Kuspa A, Shaulsky G, 2005. Epistasis analysis with global transcriptional phenotypes. *Nature Genetics* **37**: 471–477.

Xie X, Wilkinson HH, Correa A, Lewis ZA, Bell-Pedersen D, Ebbole DJ, 2004. Transcriptional response to glucose starvation and functional analysis of a glucose transporter of Neurospora crassa. *Fungal Genetics and Biology* **41**: 1104–1119.

Zupan B, Demsar J, Bratko I, Juvan P, Halter JA, Kuspa A, Shaulsky G, 2003. GenePath: a system for automated construction of genetic networks from mutant data. *Bioinformatics* **19**: 383–389.