From Department of Biosciences and Nutrition,
Karolinska Institute, Stockholm, Sweden
and
School of Life Sciences, Södertörn University, Huddinge, Sweden

DNA MICROARRAY APPROACHES TO UNDERSTANDING THE REGULATION AND EVOLUTION OF GENE EXPRESSION NETWORKS

Yongtao Xue-Franzén

Karolinska Institutet

Stockholm 2009

Published and printed by Karolinska University Press
Box 200, SE-171 77 Stockholm, Sweden

ISBN 978-91-7409-554-8

To my parents

献给我的父母

ABSTRACT

DNA microarray technology allows biological and medical research to shift from investigation of individual functions of a few related genes to the whole genome level. This creates opportunities for discovery of complex and coordinated transcriptional networks in biological systems. The aim of this thesis has been to study gene regulation and evolution using yeast responses to environmental cues as a model system.

We first developed and validated a fission yeast cDNA microarray for genome-wide expression analysis (Paper I). It is the first commercially available fission yeast microarray, which presents a useful resouce for yeast researchers and provides information required to contruct the array from scratch.

Next, we characterised the gene regulatory networks involved in the pheromone response (Paper II) and investigate the role of Gcn5 transcription co-regulator, a histone acetyltransferase (HAT), in re-programming gene expression during the salt stress response in fission yeast (Paper III).

We further investigated evolutionary conservation and divergence of Gcn5 in gene regulation by comparing its role in the evolutionarily distantly related yeast species. The parallel study of the fission yeast and budding yeast showed that Gcn5 has a conserved physiological role in salt stress responses, but it regulates diverged sets of stress response genes potentially via distinct mechanisms (paper IV).

Finally, we investigated interactions between different HATs and between HATs and HDACs (histone deacetylases). Phenotypic studies and gene expression profiling revealed that Gcn5 has overlapping functions with another HAT, Mst2, in the stress response and DNA damage repair (Paper V). We found that the HDAC Clr3 acts antagonistically to Gcn5 in transcriptional elongation and stress responses (Paper VI).

LIST OF PUBLICATIONS

This thesis is based on the following papers and manuscripts that will be referred to by their Roman numerals.

I **Xue Y**, Haas S, Brino L, Gusnanto A, Reimers M, Talibi D, Vingron M, Ekwall K and Wright APH (2004). A DNA microarray for fission yeast: minimal changes in global gene expression after a temperature shift. Yeast, 21:25-39.

II **Xue-Franzén Y**, Kjærulff S, Holmberg C, Wright APH, Nielsen O (2006). Genome-wide identification of pheromone-targeted transcription in fission yeast. BMC Genomics 7:303.

III Johnsson A, **Xue-Franzén Y**, Lundin M, Wright APH (2006). Stress-specific role of the fission yeast Gcn5 histone acetyltransferase in programming a subset of stress response genes. Eukaryotic Cell, 5:1337-1346.

IV **Xue-Franzén Y**, Johnsson A , Brodin D*, Henriksson J*, Bürglin TR and Wright APH (2009). Genome-wide characterisation of the Gcn5 histone acetyltransferase in budding yeast during stress adaptation reveals evolutionarily conserved and diverged roles. Submitted manuscript under revision.

V Nugent R*, Johnsson A*, Fleharty B, Gogol M, **Xue-Franzén Y**, Seidel C, Wright APH and Forsburg SL (2009). Expression profiling of S. pombe acetyltransferase mutants identifies redundant pathways of gene regulation. Submitted.

VI Johnsson A, Durand-Dubief M, **Xue-Franzén Y**, Rönnerblad M, Ekwall K and Wright APH (2009). HAT-HDAC interplay regulates global histone H3K14 acetylation in gene coding regions during stress. EMBO Reports, 10:1009-1014.

*These authors contributed equally to this work.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Gcn5	General control non-derepressible 5
HAT	Histone acetyltransferase
MPA	Mycophenolic acid
HDAC	Histone deacetylase
ChIP	Chromatin immunoprecipitation
PCR	Polymerase chain reaction
DNA	Deoxyribonucleic acid
KCl	Potassium chloride
ATP	Adenosine triphosphate
bp	Base pair
kb	Kilo base
SAGA	Apt-Ada-Gcn5-Acetyltrasferase
SLIK	SAGA-like
ORF	Open reading frame
IGR	Intergenic region
MAPK	Mitogen activated kinase
SNP	Single nucleotide polymorphism
MAQC	Microarray Quality Control
RNA	Ribonucleic acid
GNAT	Gcn5-N-acetyltransferase
MYST	MOZ-Ybf2/Sas3-Sas2-Tip60
NuA3	Nucleosome Acetylation on H3
NuA4	Nucleosome Acetylation on H4

1 INTRODUCTION

Technological innovation has revolutionized many fields, including biological and biomedical research. After years devoted to the isolation of individual genes involved in physiological or developmental process, DNA microarray technology has brought us to the world of whole genome processes, providing fundamental biological insights not possible with traditional approaches.

To date, the whole genome sequencing of more than 30 organisms has been completed or almost completed: from bacteria to fruit flies, from fish to human beings, they are highly variable in morphology and physiology. Amazingly, the variability of genome sequences does not reflect the variation manifested at the level of their ″phenotypes″. Additionally, the structure and function of many of the genes is often highly conserved across species. One explanation for how morphological diversity could have arisen in a relatively conserved genome background is that gene regulation patterns are remarkably complicated, diverse and flexible. This allows organisms to rapidly alter the network of genes expressed in response to new conditions, and to accommodate to evolutionary demands. Gene regulation is one of the central themes of molecular biology and evolution.

In the studies contributing to this thesis, we first established a fission yeast DNA microarray as a powerful tool for investigating gene regulation networks. We chose yeasts as model organisms, since these unicellular eukaryotes have made major contributions to current knowledge of eukaryotic transcriptional regulatory mechanisms. We focused our study on: how gene regulation networks respond to environmental stimuli; how the various transcription related factors coordinate with each other; and how the transcriptional networks differ in evolutionarily diverged yeast species.

2 DNA MICROARRAYS

Man is a tool-using animal… without tools he is nothing, with tools, he is all —Thomas Carlyle

The mechanic who wishes to do his work well, must first sharpen his tools —Confucius

Productive power, is contained in the tools and technology, developed by humans in their efforts to improve their material conditions, plus the human labour that goes into employing these tools —Karl Marx

2.1 INTRODUCTION OF DNA MICROARRAYS: HISTORY, PLATFORMS AND APPLICATIONS

DNA microarray technology has emerged as one of the most promising new tools in molecular biology.

A DNA microarray is a solid surface (glass slides or silicon chip) consisting of an arrayed series of thousands of microscopic spots of specific DNA sequences, each representing a gene or other DNA element. These spots are known as probes. The arrays hybridize to fluorescently labelled DNA samples (known as the target). Probe-target hybridization is detected and quantified to determine the relative abundances of specific nucleic acid sequences in the target. Since an array can contain a whole genome of an organism, it is a tool of major importance for genome-wide studies.

Before the era of DNA microarray technology, Differential Display technology (Liang and Pardee, 1992) was a major tool that opened the door to interpretation of genome-wide information. This technology allows detection of altered gene expression patterns by running a DNA sequencing gel after PCR amplification of total mRNA samples from control and experimental conditions. But the popularity of this technology did not last long due to its high false positive rate, and the impossibility of identifying more than a few genes. This “differential sequencing” was replaced by “differential hybridization” within a few years. This technique relied on a “Reverse Northern” hybridization approach: known cDNA probes are spotted on nylon membranes and hybridized by control and experimental mRNA samples. This eventually led to the development of the currently popular technology, DNA microarray, pioneered by Patrick Brown and David Botstein in 1995 (Schena et al., 1995) and Chee in 1996(Chee et al., 1996). This technique is in effect a repackaged differential

hybridization strategy, in which cDNA plaques are replaced with spotted cDNAs or photolithographic synthesized arrays, and radioactive labels are replaced with fluorescent ones. See figure 1 for a comparison of feature density of two modern platforms and the earlier nylon spotted arrays. In 1997, a complete eukaryotic genome (*Saccharomyces cerevisiae*) was established on a microarray (Lashkari et al., 1997).

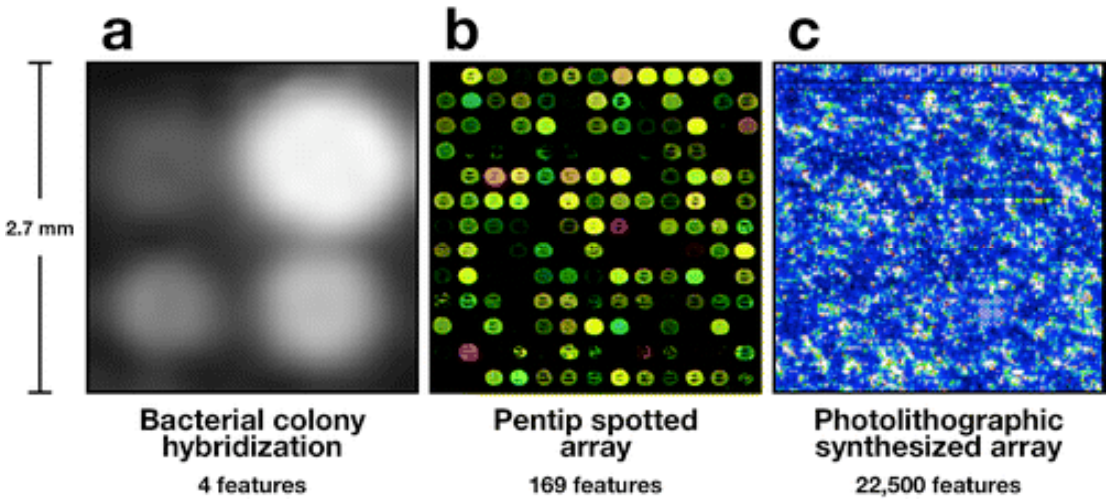

Figure 1 Feature density of representative microarrays. Each image shows a 2.7 mm square sub-region. (a) Bacterial colony spots on nylon from the 1980s. (b) Ink-jet in situ synthesized 60-mer oligo spots on glass. (c) Affymetrix human gene array with 18-µ features containing 25-mer oligos. Reprinted with permission from the Annual Review of Biochemistry, Volume 74 © 2005 (Stoughton) (by Annual Reviews www.annualreviews.org.) Updated microarray density information not shown.

In spotted microarrays, the probes are first synthesized and then "spotted" onto glass slides with fine-pointed pins (Figure1b). Paired samples are labelled with fluorescent cy3 or cy5 and hybridized to a single array slide (two-channel microarray). In Affymetrix oligonucleotide arrays (Figure1c), short nucleotide sequences are synthesized directly onto the silicon chips. Biotin labelled sample is hybridized to a single slide. Affymetrix oligonucleotide arrays offered high spotting density and excellent quality controls. However photolithographic synthesis limited the probe length (25mer in Affymetrix system) even though 60-mers allow a much better sensitivity-specificity trade-off than shorter oligos (Stoughton, 2005). Additionally, the cost of photolithographic mask is high. Spotted arrays, in contrast, allow a range of probes based on oligonucleotides, cDNA or small fragments of PCR products, with the additional advantage that the desired oligos or cDNA can be pre-synthesised. Spotting

is a relatively low-tech but robust and affordable method, although the probe density obtained is not comparable with Affymetrixtype microarrays.

For many people, DNA microarrays are synonymous with gene expression profiling, as the most common type of microarray in current use is the expression array. This type of array contains probes representing the protein coding regions, with one probe set for each gene, and has been the main focus of genome-wide investigations for many years. More recently, with the development of array synthesis techniques and the increasing availability the whole genome information, tiling arrays have entered the genomic biologist's toolkit In contrast to expression arrays, tiling arrays consist of probes designed to densely represent a genomic region of interest, or an entire genome including intergenic regions as well as ORF regions. Tiling arrays are often used for profiling DNA–protein interactions and epigenetic studies. Selected applications of DNA microarrays are listed in table 1.

Table 1 Selected applications of DNA microarrays.

Application	
Gene expression profiling	Measurement of gene expression (gene transcription or mRNA) level of thousands of genes or the entire genome simultaneously, to create a global picture of difference in gene expression between of control and experiment samples. (for example, treatment vs non-treatment, normal vs diseased) (Schena et al., 1995).
Protein-DNA interaction (ChIP-on-chip)	Also called genome-wide location analysis. DNA sequences bound to a particular protein can be isolated by immunoprecipitating that protein (ChIP), these fragments can then be hybridized to a microarray allowing the determination of protein binding site occupancy throughout the genome. Example applications include determining transcription binding sites, or mapping histone modification (Iyer et al., 2001; Lieb et al., 2001; Ren et al., 2000).
Comparative genomic hybridization (CGH)	Also called chromosomal microarray analysis. Measures regional copy number changes (gains/losses) in DNA content, for example finding abnormal regions in the genome by comparing tumour and normal DNA, or identifying novel mutations that responsible for dysmorphic features, developmental delays, mental retardation etc. (Moran et al., 2004; Pollack et al., 1999).
SNP detection	Identifying single nucleotide polymorphisms among alleles within or between populations. Used in genotyping, forensic analysis or genetic linkage analysis (Hacia et al., 1999).
Alternative splicing detection	To detect each individual exon or splicing isoforms for known or predicted genes (Clark et al., 2002; Yeakley et al., 2002)

2.2 IMPACT OF DNA MICROARRAYS IN BIOLOGICAL AND BIOMEDICAL RESEARCH

DNA microarray technology has been growing rapidly in popularity ever since it was developed. The number of publications based on the use of DNA microarray technology is increasing exponentially (Figure 2), while the total number of scientific publications increases only linearly (data not shown, total publication doubled in 2008 compared to 1995).

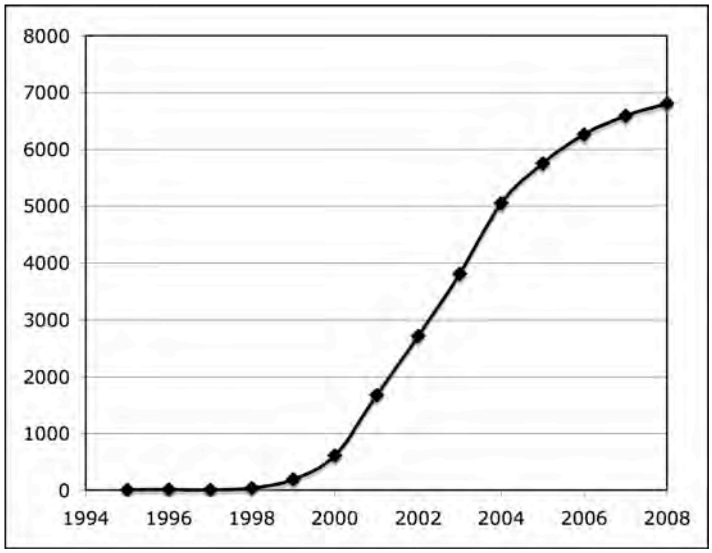

Figure 2. Number of articles in Pubmed containing keyword "DNA microarray". A search was conducted in http://www.ncbi.nlm.nih.gov/pubmed/advanced using "DNA microarray" as search term. Identical searches were conducted for each year from 1995 to 2008.

The popularity of microarray technology lies in the facts that, first, it has allowed a shift from studies of the individual biological functions of a few related genes to more global investigations, which provides an opportunity for discovery of complex and coordinated transcriptional control in biological systems. Second, it can be applied to research across a wide range of fields: cellular physiology, toxicology, disease characterization, evolutionary biology, stress responses and drug development.

A study using DNA microarray technology was carried out by Friend and colleagues in 2002 (van 't Veer et al., 2002), who provide a good example of the use of expression profiling by DNA microarray to improve breast cancer classification and therapy. Many breast-cancer patients receive unnecessary chemotherapy or hormonal treatment for possible tumour spread after the removal of a primary tumour. Classical

histopathological and clinical criteria are insufficient to determine which patients need this adjuvant treatment.

Friend and colleagues analysed primary breast tumours from a mix of patients who developed metastases or remained disease-free after 5 years, using DNA microarrays to study the gene expression profiles. They were able to identify 70 genes significantly associated with disease outcome, and developed a prognosis classifier based on this microarray study. This constituted a powerful tool for tailoring treatment to those patients at risk of recurring disease, and avoiding unnecessary treatment and the associated costs in patient quality of life and health-care expenditure.

DNA microarray allows us to survey, on a global scale, organisms and their patterns of gene and protein expression. Whole genome research has been used in multiple studies like this one, to pinpoint the links between the "genotype" and metastases tumour "phenotype".

This type of study is highly relevant in elucidating the networks and pathways that are fundamental for biological processes or disease development.

2.3 TECHNICAL CONSIDERATIONS OF DNA MICROARRAY TECHNOLOGY

In the early stage of my PhD studies, we developed a DNA microarray for fission yeast to study gene expression profiling (expression array) (paper I). Later, ChIP-on-chip high resolution tiling arrays were used in combination with expression array. In this section, I will address some perspectives on the technical considerations of DNA microarray technology based on these years' experience, mostly regarding two-channel spotting expression arrays.

DNA microarray analysis involves a number of steps. A complete microarray process including gene expression profiling and ChIP-on-chip for microarray experiments used in our study is illustrated (figure 3).

There are various sources of error, which can occur in microarray experiments. Most technical considerations in DNA microarray analysis boil down to concerns about the biological variation and systematic variation that could obscure the true difference

between control and experimental samples. Here we discuss some of the issues that are important for drawing statistically and biologically valid conclusions from microarray experiments at the stages of experimental design, handling of samples, hybridization, scanning, normalization and data analysis.

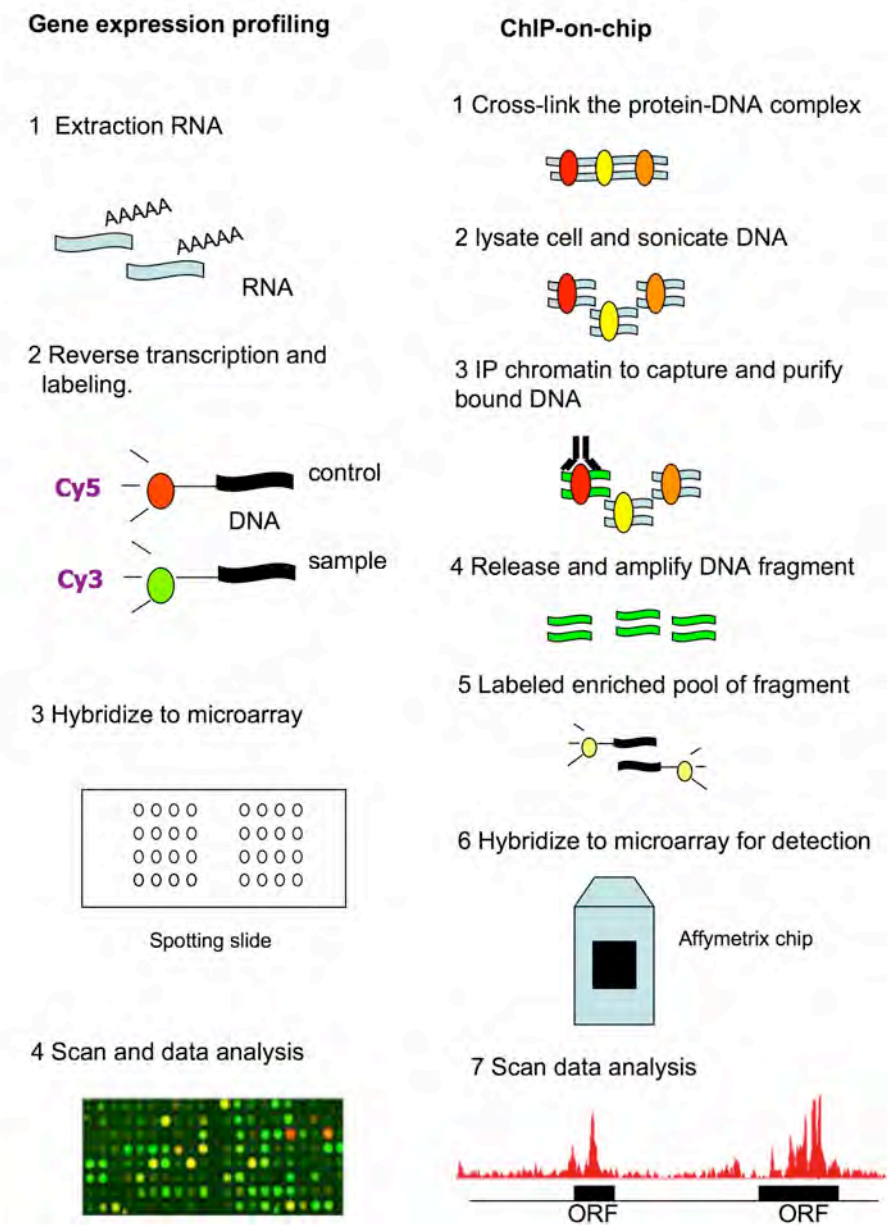

Figure 3 Experimental procedures for gene expression profiling and ChIP-on-chip.

2.3.1 Experimental design

Replicate samples are usually required for DNA microarray experiments because of the intrinsic biological variation between samples in addition to systematic variation during experimental procedures. It has reported that biological variation exceeds technical variation for DNA microarrays (Chudin et al., 2006; Shi et al., 2006; Zakharkin et al., 2005). This accords well with my own research experience: most variation comes from different biological samples; the results for repeating microarray experiments of the same set of biological samples are highly reproducible. Therefore it is more important to use biological replicates. How many replicates are needed is determined by several factors: goals of the study, sample resources, reliability of the technology and cost, etc. In the fission yeast cDNA microarray we designed, probes are double spotted at separate regions on a single slide to provide a measure of technical control for hybridization. This design provides technical replicates within a single hybridization.

One source of systematic variation in spotted microarrays is the bias caused between fluorescent dyes Cy5 and Cy3. These biases stem from a variety of factors, including physical properties of the dyes (heat and light sensitivity, relative half-life) and efficiency of dye incorporation (Yang, 2002). Gene specific dye bias also exists, in which certain genes tend to bind more efficiently to one dye than the other (Goryachev et al., 2001; Kerr et al., 2000; Tseng et al., 2001; Wang et al., 2001; Yang et al., 2002), and as a result, an observed difference between cy3 and cy5 channel intensities for a particular gene may be not due to differences in expression level, but differences in dye incorporation efficiency. Dye-swap design is therefore often used—samples are split in half and dye-label is swapped instead of running all control group or experimental group with the same dye. Even so, there are debates about whether dye-swap design should be used (Dobbin et al., 2005), and the choice to use dye-swap depends on several aspects of experiment design (Dobbin et al., 2003). For paired samples in our study, which compared samples "before treatment" and "after treatment", we were primarily interested in understanding the effect of treatment on gene expression. In this situation, dye-swap is an efficient design to correct for gene-specific dye binding inequality. In other situations, a reference can be used to compare to non-reference samples. For example, the reference sample may be a mixture of normal tissue and the non-reference sample RNA extracted from different tumours, so that the comparison would give some indication of genes expressed differently in tumours. This kind of

system is considered more effective without out dye-swap (Dobbin et al., 2003). The results tend to be more robust, relatively simple to analyse and produce better clusters (Dobbin and Simon, 2002). Normalization methods such as Lowess, which we will discuss below (Dabney and Storey, 2007), help to correct dye bias.

2.3.2 Handling of samples and hybridization

Even though systematic or technical variation in DNA microarrays is not avoidable, care in handing of samples could reduce such variations. Taking yeast experiments as an example, different batches of culture medium, treatment reagents, shaking speed when culturing cells, centrifuging when collecting samples, and so on are potential factors that cause systematic variation. Therefore, using the same batches of culture medium and chemical reagents, and keeping all the experimental conditions as similar as possible would be a good practice to reduce systematic variations. It is also important to perform control and experimental samples in parallel to avoid noise from batch differences. For gene expression profiling studies, the integrity of RNA or DNA is essential for a successful experiment.

DNA microarray is a hybridization-based method. A given microarray probe is designed as a perfect complementary strand to a given region of the target DNA. Based on the Waston-Crick paring, the probe will capture a certain number of target sequences. DNA microarray hybridization is designed to identify specific signals produced by the original labelled target. Ideally, the signal intensity should be proportional to the number of molecules of targets bound to probes. In reality, the signal produced by a given microarray probe can also be cross-hybridization signal produced by non-specific targets with significant sequence similarity with the probe (e. g., splice variants) or even a non-specific, background signal due to probe binding in the absence of any significant sequence similarity (Draghici et al., 2006). These non-specific signals are not avoidable but are negligible for a high specificity design of microarray platform. Optimizing fundamental parameters such as time, stringency and concentration of target sample can help to minimize the non-specific binding. We should also keep in mind that the property of the samples used for hybridization, such as fragment length, GC content and available binding sites. The first two secondary parameters are validated in our *S. pombe* arrays (Paper I).

2.3.3 Normalization

Normalization is to adjust the individual hybridization intensities to balance them appropriately so that meaningful biological comparisons can be made. There are numerous reasons why data must be normalized, including unequal quantities of starting material, differences in labelling or detection efficiencies between the fluorescent dyes used, and systematic biases in the measured expression levels. The choice of normalization method is determined by the underlying hypothesis. Most commonly, for expression arrays, the assumption is that gene expression levels of the majority of the transcripts on the arrays do not change between the samples. Total intensity normalization (global normalization) is based on this assumption. But, with a little experience, it becomes clear that a bias occurs in changes of expression level depending on spot intensities, i. e. changes in expression are not symmetric for all gene expression levels (Figure 5). This is not a biological phenomenon but results from dye bias at different gene expression levels, One explanation for this is quenching (Jeon, 2007) —a phenomenon where dye molecules in close proximity re-absorb light from each other, thus diminishing the signal. Quenching acts at different levels for each dye, leading to signal intensity bias at different gene expression levels. Lowess normalization (Intensity dependent normalization) is then used to compensate for this bias. The assumption for Lowess normalization is that changes in gene expression are roughly symmetric at all signal intensities. It is a good approach for spotted arrays to compensate for dye bias.

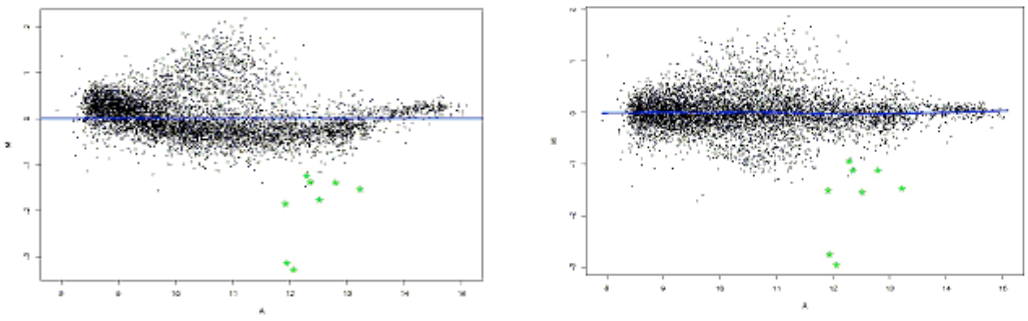

Figure 4 Ratio-Intensity plot showing bias depending on signal intensity (left panel) and corrected by lowess normalization (right panel). The picture is courtesy of Prof. Terry Speed.

Sometimes, changes are expected for the total gene expression levels between control and experimental groups, in other words, distribution of the up- and down-regulated genes is asymmetric, and in this situation, neither global nor Lowess normalization is

appropriate. Therefore, normalization relative to house-keeping genes is an alternative approach. In this normalization method, similar to that performed in quantitative PCR, we assume the expression level of house-keeping genes does not change between control and experimental samples. However this is not always true, as the expression level of house-keeping genes might be different in the groups under comparison. To avoid this problem, exogenous genes such as luciferase are used as a spiking control for normalization. Spiking control RNAs show no sequence similarity to the genome of the studied species and they are added in defined amounts to experimental RNA samples before labelling. The oligonucleotides specific for the spiking RNAs are spotted onto the slide. A possible technical limitation is the inclusion of spike probes in commercially available chips. In the fission yeast microarray we have designed, we have included the luciferase gene from *Renilla*, whose RNA is commercially available (Promega). Spiking control normalization for this microarray has been successfully applied by my colleagues (Djupedal et al., 2005).

2.3.4 Data analysis and biological interpretation.

The most challenging task for DNA microarray studies is not how to generate these vast bodies of genomic data, but rather in how best to analyse the data and make sound biological interpretations.

--Find significant genes: fold change or P-value?

The goal of an expression array experiment is to determine bona fide changes in gene expression profile between experimental paradigms. For biologists, the most direct and straightforward approach is to use fold change. High fold changes do indicate biological significance. Transcripts with high fold changes can relatively easily be validated by PCR, northern blot or other methods. However, using fold change has following limitations: First, it ignores small changes which might be statistical significant and biologically meaningful, and specifically, we miss important information when we are interested in investigating a global trend in expression pattern difference. Second, genes with low expression frequently have higher fold changes (Mariani et al., 2003), thus, simple static fold change thresholds are too stringent for highly expressed genes, and not stringent enough for low-expressed genes. Third, cross comparison of microarray data between different platforms, species or laboratories employing a single cut off would be difficult due to technical variability, if appropriate

normalization is not used. Most importantly, high fold change does not rule out the problem of false positives. Therefore testing significance by statistics is a necessity in microarray studies.

One widely used approach is to use a Student's t-test, which looks at the mean and variance of the two distributions and calculates the probability that they were sampled from the same distribution. When using t-test, it is important not to assume equal variance between test and control. e.g. the signal from test group could be 1389, 1089, 990 and signal from test 45, 60, 35. Here it is clear that the variance is numerically larger in the test than in the control. Welch's t-test takes this into account by assuming unequal variances between the two populations (Knudsen, 2002). If more than two conditions are studied, t-tests are not an ideal choice. The method known as analysis of variance (ANOVA) can be used instead. For all statistical tests that calculate P-values, it is important to consider the effect of multiple testing as we are looking at not just one gene but thousands of genes. If P-value is 0.01, for 6000 genes, we expect 60 false positives. Bonferroni correction is an approach which helps to correct for this (Bender and Lange, 2001; Duboit, 2000). (e.g. if we want a chance of 25% of having one false positive in the list of genes that change significantly in an experiment, a cut off of 0.25 divided by 6000 genes give a P-value of 4e-5.) Bonferroni correction gives very conservative interpretations. It often happens that no genes can pass this corrected p-value threshold, and in this case, genes with the smallest P-value can be chosen (Knudsen, 2002).

--Visualization, how important is it?

In my opinion, visualization is essential for microarray data analysis. Microarray data produces tens of thousands of numbers. Almost all of our data interpretations are based on visualization of those numbers. The MA plot in Figure 4 clearly shows the fold change bias depending on gene intensity (left) and dye bias correction after Lowess normalization (right). A wise choice of visualization help to reveal pattern changes and facilitate biological interpretation. Figure 5 provides some examples of visualization based on my own data produced using GeneSpring software. Figure 5A reveals the gene regulation pattern at a specific chromosome position; this type of visualization has allowed us to observe significant changes in gene regulation at telomeric regions. (Shown in paper V, figure 2). In figure 5B, three samples (A, B, C) are clustered based

on their expression pattern. It makes sense that the replicates (1 and 2) for each sample are clustered together.

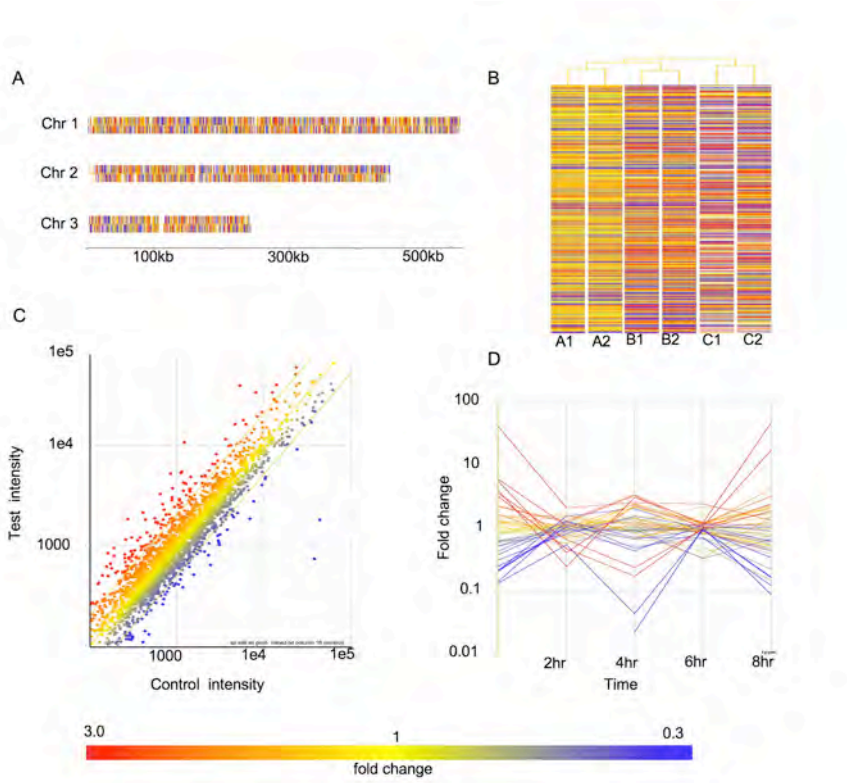

Figure 5 Visualization of microarray data. A. Distribution of regulated genes on chromosome. B. Condition tree clustering shows similarities and differences of gene regulation pattern between samples. C. Scatter plot of shows up or down regulated genes of test sample relative to control depending on their signal intensity. D. Graph shows gene regulation at different time points. Colour bar shows the scale of fold change.

--Microarray software: Commercial software or in house script?

The main advantage of using commercial software is its comprehensiveness. Most current analysis tools are included in a single package, with a user-friendly interface. For example, GeneSpring (Agilent Technologies) includes automated gene annotations for different species, and orthologous pairs between species can be translated, which greatly facilities interspecies comparisons. Commercial software usually provides various visualization functions, which are important for microarray data interpretation.

There are several potential limitations inherent in using commercial software. (Not to mention the yearly cost of commercial software is high for laboratories with tight budgets.) First, they might not be flexible enough, causing problems such as data comparability problems and conversion inconvenience. Another problem is their

capability of performing various data analyses with ease, which might lead the inexperienced user to overlook certain statistical limitations of the data analysis. Finally, a single comprehensive package is still not able to provide analysis tools for each individual customer's needs.

In house scripts can be created to perform one or a few specific analyses, e.g. clustering. They are usually straightforward to answer specific questions. Some open source software provide more comprehensive functions and allows users to analyse data using a host of existing techniques and to develop their own and integrate them within the system. Three of the most widely used open source statistical analysis tools are: Bioconductor written in R (http://www.bioconductor.org), the Java-based TM4 software system available from The Institute for Genomic Research (http://www.tigr.org/software), and BASE, the Web-based system developed at Lund University (http://base.thep.lu.se). (reviewed in (Dudoit et al., 2003)). The limitations of in house scripts or open source software include: lack of comprehensive visualization, in addition the requirement for knowledge of computer languages and a clear understanding of how to look for answers to confirm a hypothesis.

--Data confirmation, is it really necessary?

Each new breakthrough in technology immediately encourages investigators to make use of the new tool, and quite often gives rise to intensive activity to explore its place in the scientific repertoire and determine the significance of the technology. DNA microarray is no exception. Issues of sensitivity, accuracy, specificity and reproducibility in DNA microarray measurement have been discussed (Draghici et al., 2006; Shi et al., 2008). Many reviewers request verification of DNA microarray data by independent approaches such as quantitative PCR or northern blot. Indeed, every technology has its strengths and limitations. False positive signals can be generated by probes on the microarray hybridization, or other steps in the DNA microarray process. Nevertheless, the microarray quality control (MAQC) project demonstrated reliability and reproducibility of the microarray technology, and reported that the key factors influencing variations are the biological samples and human factors, rather than technical inconsistency {Shi, 2006 #126}. Based on my own experience, microarray data are reproducible even when using different platforms. Data from Eurogentec spotting arrays and Affymetrix correlated well when using identical sets of RNA samples, even though the Affymetrix result gave higher ratios, indicating higher

sensitivity of the platform. In Paper IV, we studied genome-wide localization of Gcn5 by the Affymetrix genechip system in *S. cerevisiae*. Gcn5 associated genes determined by this method show important overlaps with the previously reported results obtained by nylon membrane and spotting array (Robert et al., 2004; Rosaleny et al., 2007). In paper VI, we reported that, among the few genes regulated by Gcn5, the down regulation of Mei2 and ste11 Genes, and this down regulation is confirmed by other research work (Helmlinger et al., 2008).

This raises questions about the necessity for DNA data confirmation by verifying a number of individual genes by independent methodical approaches: first, how meaningful is it to demonstrate validity for a small number of genes (quite often highly regulated genes are selected instead of a random selection) when the conclusion drown from the microarray data is based on the global trend? Second, can we always confirm the result from a highly sensitive method (hybridization based DNA microarray) by a less sensitive method such as PCR? Of course, any data that cannot be confirmed are simply omitted from most publications. In my opinion, adequate replicates, good laboratory proficiency and appropriate data analysis provide reliable microarray results. Independent biological verification instead of technical verification is more convincing. For example, in paper VI, we draw a conclusion from the ChIP-on-chip data that Gcn5 might play a role in transcriptional elongation, because Gcn5 is localised in the coding regions of highly transcribed genes. Cell plating assays show that gcn5 deletion indeed causes deficiency in transcriptional elongation, this is more convincing than confirming the binding pattern by PCR for a group of highly transcribed genes associated with Gcn5.

3 YEASTS AS MODEL ORGANISMS

3.1 ADVANTAGES OF USING YEASTS AS MODEL ORGANISMS

Yeast has been used in biological research since the nineteenth century (Barnett, 2003) and ever since it has been a popular model organism in genetics and cell biology.

Several traits of yeast species make them popular model organisms. First, yeasts are simple: they are single celled organisms and easy to grow and replicate. They are non-pathogenic and few precautions are needed for handling them. Most importantly, basic biological process, such as transcriptional machinery, cell cycle, tRNA processing, stress associated functions and G-protein-based pheromone response pathway are conserved between yeasts and humans. Remarkably, Lee Hartwell, Sir Paul Nurse and Tim Hunt were awarded the 2001 Nobel Prize in physiology or medicine for pioneering the use of yeast genetics to define the cell cycle and to understand its control and role in carcinogenesis, which provides insight into human cancer. The similarity is not only in the general outline of the processes, but also the molecules themselves: More than half of human proteins have homologs in yeast (reviewed in (Wright, 2005), even human disease genes (Botstein et al., 1997; Wood et al., 2002). Finally, the techniques available in yeast genetics (such as allowing for either the addition of new genes or deletion through efficient homologous recombination) makes yeasts favourable.

The budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) and fission yeast *Schizosaccharomyces pombe* (*S. pombe)* are the best-studied model organisms. These two species diverged approximately 300 to 600 million years ago. Both species share features with more complex eukaryotes. In many aspects, *S. pombe* often resembles mammalian cells more closely than *S. cerevisiae*. For example, chromatin modifying enzymes in *S. pombe* mimic more closely the mammalian enzymes. Mechanisms of Histone-H3-K9 methylation and the RNAi/dicer system in gene repression are found in *S. pombe* but not in *S. cerevisiae* (Provost et al., 2002; Volpe et al., 2002). We have used both yeast models in our research. Additionally, we used *Saccharomyces kluyveri* (*S. kluyveri*), which is a budding yeast related to *Saccharomyces cerevisiae. S. kluyveri* appears to have become a species before the whole genome duplication that occurred in the *Saccharomyces* lineage (*S. kluyveri* genome sequencing project: http://genomeold.wustl.edu/). Comparison and parallel studies between these yeast

species may be give us significant evolutionary insights about the conservation of a biological process. Phylogenetic tree of yeast species is shown in Figure. 6.

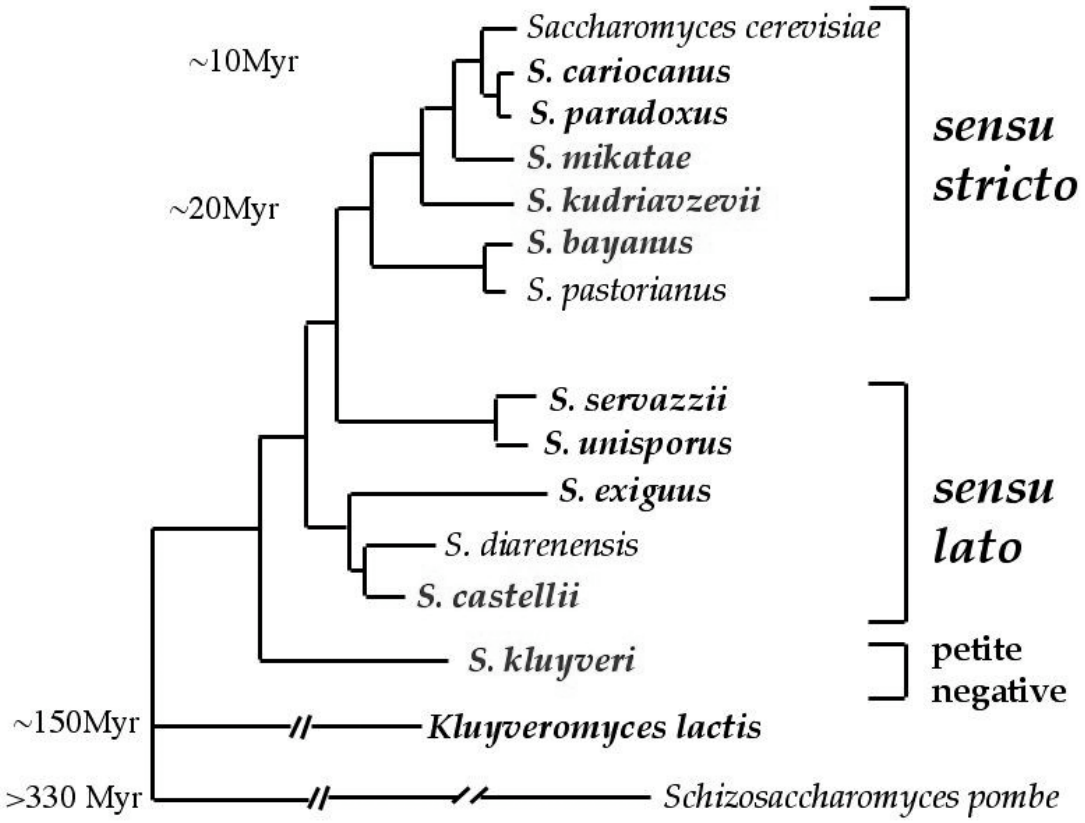

Figure 6. Phylogenetic tree of yeast species. *S. pombe* and *S. cerevisiae* are evolutionarily diverged. S. *kluyveri* is a distantly related yeast to *S. cerevisiae* in *Saccharomyces* linage. Picture is courtesy of Ashwin Dasikan and Washington University. Tools for genome sequence alignments: http://genomeold.wustl.edu/

3.2 COMPARATIVE GENOMICS USING YEAST MODELS

Comparative genomics is the study of the relationship between genome structure and function across different biological species or strains, helping us to gain new insights into evolutionary, biochemical, genetic, metabolic, and physiological pathways. In recent years, several key factors have contributed the development of comparative genomics. These are notably: first, the availability of increasing amounts of genome information from different organisms in public databases. Second, the development of microarray technology, especially the increased performances of array synthesis devices (allowing increased probes densities as well as array design flexibility and reduced production costs). Third, the increased computational capabilities that can handle the mass of data. Yeast models have been placed at the forefront of comparative genomics, particularly studying gene transcription networks and evolution. This can be

attributed not only to the complete sequencing of more than 10 yeast species (Dujon et al., 2004; Kellis et al., 2003), but also to the fact that different yeasts can be cultured and manipulated under comparable external conditions.

3.3 STUDYING YEAST RESPONSE TO ENVIRONMENTAL CUES

In general, all cells, whether free-living or part of multi-cellular organisms, must contend with a variety of environmental fluctuations that can be harmful or lethal to the cell. Cells exposed to different kinds of environmental cues rapidly alter their gene transcription, chromatin status and signalling pathways. Significant clues to the mechanisms involved in adaptation to new environments have come from studies of the genes that are expressed in response to specific stresses. Investigators have identified transcriptional activators and repressors that likely contribute to coordinate remodeling of genome expression (Kingston et al., 1987; Parker and Topol, 1984; Sorger and Pelham, 1987; Wu, 1985).

On another level, most proteins are completely dispensable for growth in favourable conditions (Varringer, 2003). Biological phenomena can be overlooked in favourable conditions, for example: 1) A specific protein can have a function which is important for the cell only in the face of particular environmental conditions. 2) Redundant functional relationship between proteins might not be easily observed in an ideal growth environment. 3) Oppositional effects between certain proteins can be observed under specific environmental stimuli, and this is particularly important for understanding regulatory mechanisms.

In the studies contributing to this thesis, we have analysed pheromone-induced genes involved in the sexual differentitation process (paper II). Subsequently, our main focus has been on gene regulation of Gcn5 under salt stress conditions, because under normal conditions (rich medium), relatively few genes are regulated by Gcn5, indicating that its function can be substituted by other proteins under these conditions. We found that Gcn5 becomes essential when cell are treated with KCl, $CaCl_2$ or $MnCl_2$. Our focus is primarily using KCl stress conditions as a model system for studying gene regulation by Gcn5, rather than characterising the physiology of the KCl response in yeasts. Indeed, we found that Gcn5 regulated genes under KCl stress overlap with common environmental stress response genes.

4 GENE REGULATION

Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product. These products are often proteins, or functional RNAs. Regulation of gene expression is fundamental for all living organisms, from viruses, prokaryotes, single celled yeast, to multi-cellular organisms with highly variable morphology and physiology. Regulation of gene expression increases the versatility and adaptability of an organism by allowing the cell to express gene products when needed. Furthermore, gene regulation drives the processes of cellular differentiation and morphogenesis, leading to the creation of different cell types in multi-cellular organisms where the different types of cells demonstrate different gene expression profiles though they all possess the same genome sequence. Gene regulation may also serve as a substrate for evolutionary change, since control of the timing, location, and amount of gene expression can have a profound effect on the functions (actions) of the gene in the organism.

Several steps in the gene expression process may be modulated, including the transcription, RNA splicing, translation, and post-translational modification of a protein. Gene transcription is the first step and essential for gene expression. In eukaryotes, the general transcriptional machinery and to some extend co-regulatory proteins are highly evolutionary conserved between organisms as diverse as yeast and mammals (Guarente and Bermingham-McDonogh, 1992; Schena, 1989).

4.1 BASAL POLII TRANSCRIPTION MACHINERY

Gene transcription is the process by which the genetic code in DNA is copied (transcribed) into a RNA molecule, The process of transcription may be divided into three parts; initiation, elongation and termination. A gene is transcribed from the 5'end to the 3' end. Genes can be grouped according to the mechanism by which they are transcribed and the eventual fate and role of the RNA product. RNA polymerase II transcribes messenger RNAs (mRNAs), which are the product encoded by the majority of gene sequences, as well as some small nuclear RNAs (Latchman, 2005). Other RNA types, mainly structural or enzymatic, are transcribed by RNA polymerases, RNA polymerase I and RNA polymerase III, but will not be discussed further in this thesis.

In eukaryotes, RNA polymerase II cannot bind to a promoter and initiate transcription by itself. Rather, it requires general transcription factors (GTFs: including TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH) and mediators added in a specific order for assembly of the transcription pre-initiation complex (PIC). These factors function collectively to recruit the polymerase and specify the transcription start site. PIC formation usually begins with TFIID binding to the TATA box, initiator, and/or downstream promoter element (DPE) found in most core promoters, TFIID, together with TFIIA, forms a stable protein-DNA complex on the promoter. The subsequent binding of TFIIB and TFIIE yields a complete initiation complex. At this stage, this promoter-bound complex is sufficient for a basal level of transcription in *vitro* and in *vivo*. Basal transcription factors are ubiquitous and interact with the core promoter region surrounding the transcription start site(s) of all classes of genes (Orphanides et al., 1996).

4.2 TRANSCRIPTIONAL REGULATION

Gene transcription is a highly regulated process. No individual factor is capable of playing a dominant role in generating the immense specificity required to make sure that genes are expressed in the right place, at the right time and in the right amount depending on the changing requirements of the organism. All transcription factors in eukaryotes interact not with naked DNA, but with the complete chromatin structure including a complex assortment of DNA-associated proteins. Transcription factors, co-regulators, modification of chromatin structure and non-coding RNA all play crucial roles in controlling the basal transcription machinery. These regulation mechanisms are not isolated but highly dependent to each other, a dependency that is reflected in structural alterations of the chromatin. These can give rise to i) increased or decreased access to the DNA for transcription factors, ii) stabilisation or prevention of binding of RNA polymerase to DNA regulatory sequences.

4.2.1 Regulation by chromatin remodeling and histone modifications

In all eukaryotic cells, chromosomal DNA is bundled together with proteins called histones and packed into a highly ordered structure: chromatin. At least one functional consequence of this packaging is that it prevents access of DNA-binding proteins that regulate transcription to the promoter. Nucleosomes, the basic structural unit of chromatin, consist of 146 base pairs of DNA wrapped twice around a cylindrical

protein core containing two copies of each histone: H2A, H2B, H3 and H4 (Kornberg, 1974; Luger et al., 1997). Nucleosomes impede all steps required for transcription. Thus, in the simplest case, repression can be achieved by creating a stable, inaccessible chromatin structure, and activation can be achieved by creating an accessible chromatin structure. Changes in chromatin structure may facilitate gene transcription by loosening the histone-DNA complex, allowing other proteins such as transcription factors access to the DNA (Strahl and Allis, 2000). Two highly conserved mechanisms for modifying chromatin structure are: chromatin remodelling and histone modification, carried out by ATP-dependent remodelling complexes and histone-modifying complexes respectively.

ATP-dependent remodelling complexes can expose nucleosome DNA by introducing torsion into the DNA strand and generating negative supercoils (Havas et al., 2000) . This results in a weakening of the DNA-histone interactions thus facilitating nucleosome sliding. ATP-dependent remodelling complexes may be divided into three groups, referred to as SWI/SNF, ISWI and Mi-2/CHD, based on the identity of their catalytic ATP-ase subunit, and their differing requirements for substrate.

Histone modifying complexes covalently modify nucleosomes by adding or removing chemical moieties to histone tails. Histone modifications are proposed to affect chromosome function through at least two distinct mechanisms. The first mechanism suggests modifications may alter the electrostatic charge of the histone, resulting in a structural change in histones or their binding to DNA. The second mechanism proposes that these modifications are binding sites for protein recognition modules, such as the bromodomains or chromodomains, that recognize acetylated lysines or methylated lysine, respectively. (Cosgrove et al., 2004; Strahl and Allis, 2000; Vaquero et al., 2003) Histone modifications include methylation, acetylation, phosporylation, ubiquitination and sumoylation etc. Cross talk between histone modifications adds a level of complexity to interpreting modification patterns. For a comprehensive review, see (Peterson and Laniel, 2004). In a later section, we will mainly discuss the functional consequences of acetylation in transcriptional regulation.

4.2.2 Regulation by transcription factors

Transcription factors generally simultaneously bind DNA together with an RNA polymerase, serving as a platform for recruitment of the transcriptional machinery. A defining feature of specific transcription factors is that they contain one or more DNA binding domains (DBDs), which attach to specific sequences of DNA adjacent to the genes that they regulate. Transcription factors act as either activators or repressors to increase or decrease the rate of transcription. Transcription factors also recruit co-activators or corepressors to regulated gene transcription.

4.2.3 Regulation by coactivators or corepressors.

In contrast to transcription factors, coactivators and corepressors do not bind directly to DNA. They are recruited to the transcription factor and DNA complex activating or repressing transcription. The coactivator can enhance transcription initiation by stabilizing the formation of the RNA polymerase complex. It may also control various other downstream steps of transcription, including elongation, RNA splicing, and termination and degradation of the coactivator-activator complex. Some coactivators possess histone acetyltransferase (HAT) activity, which modulate chromatin structure and causes chromatin to relax in a limited region allowing increased access to the DNA. Gcn5, CBP and p300 are examples of coactivators with HAT activity. The corepressor can, in the opposite way, repress transcriptional initiation, by deacetylating histone and making DNA less accessible to transcription (Goodson et al., 2005; Lazar, 2003). Coactivators and corepressors can interact antagonistically and modulate transcription dynamically. See Figure 7.

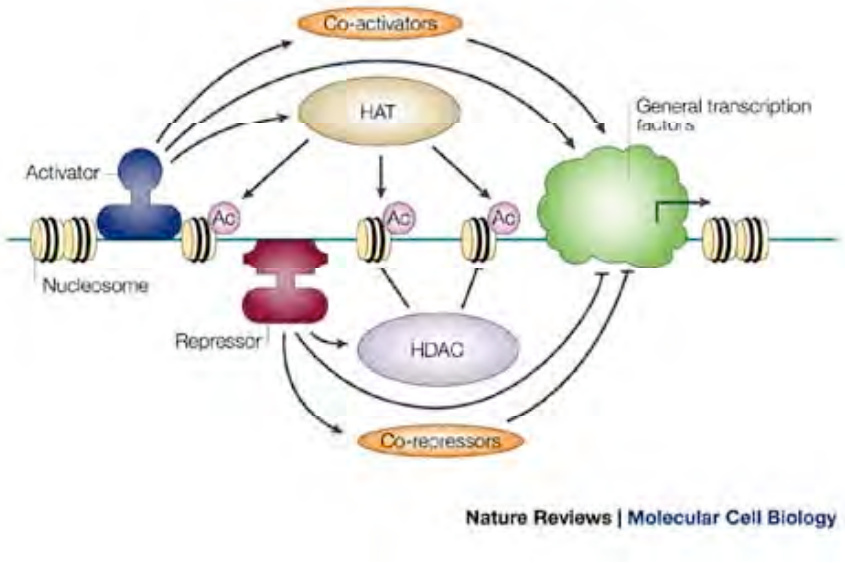

Figure 7 Positive and negative regulation of gene transcription: transcription factors (activators) can activate gene transcription via recruitment of enzymes with histone-modifying and -remodelling activities, such as HAT (histone acetyltransferase), or direct contact with components of the general transcription machinery, or interaction with transcriptional co-activators. Transcriptional repressors antagonize many of the same steps; the deacetylating of histones by histone deacetylase (HDAC), blocking the recruitment of the general transcription machinery and interacting with transcriptional co-repressors. Regulation by non-coding RNA is not included in this figure. Reprint from Nature Reviews Molecular Cell Biology 4, 192-201 (March 2003) with permission from Nature Publishing Group.

4.2.4 Regulation by non-coding RNA

The term non-coding RNA (ncRNA) refers to functional RNA molecules that are not translated in to protein. ncRNA can be as short as 20bp, or longer than 4kb. Recent studies indicate that many of the non-coding RNAs, previously thought to be functionally inert, are actually transcriptionally active in various features. Several classes of non-coding RNAs that regulate transcription have been identified, including microRNAs (miRNA) and short interfering RNAs (siRNA). miRNA can directly regulate gene transcription through binding enhancers and promoters. miRNA is also involved in post-translational repression through mRNA targeting and degradation (Wagner and Lynch, 2008). Several lines of evidence implicate non-coding RNA in transcriptional regulation in several species: a recent study showed that RNA polymerase II transcription of non-coding RNA is required for chromatin remodelling at the fission yeast *S. pombe* fbp1+ locus during transcriptional activation by converting chromatin to an open configuration (Hirota et al., 2008). In human, a model has been suggested in which ncRNAs serve as a molecular "ligand" for a specific RNA-binding protein, TLS, which in turn modulates co-regulators and regulates transcription (Wang et al., 2008). In *S. cerevisiae*, a 2.4 kb non-coding RNA accelerates chromatin

remodelling and recruitment of PNAPII to activate PHO5 gene transcription (Uhler et al., 2007) In future, high-density tiling arrays may be used to survey a well-annotated genome and identify the boundaries, structure and level of coding and non-coding transcripts, this will help to understand the significance and complexity of non-coding RNA in gene regulation.

4.2.5 Transcription networks and evolution

In summary, gene transcription is complicated and regulated by multiple factors and complexes in a highly cooperative manner. From modulation of higher order chromatin structure, to binding to promoters or enhancers, from modifying nucleosomal structure to communicating between activators or repressors at the site of transcription initiation, and eventually generating transcripts, all these steps are coordinated to ensure proper regulation. In this transcriptional system, the individual components interact with each other; these interactions include both protein-protein interactions among transcription factors and regulatory interactions between transcription factors and their sites in DNA. These interactions can be collectively represented as a network. The transcription network is not static but dynamic. One scenario is that the same transcription factors can utilise different transcription co-regulators under different conditions. One interesting question is to characterise how different combinations of co-regulator proteins participate in re-programming expression of the genome in response to particular conditions.

From an evolutionary point of view, general transriptional machinery, and to some extent co-regulatory proteins and non-coding miRNA are highly conserved in eukaryotes (Bentwich et al., 2005; Berezikov et al., 2006; Guarente and Bermingham-McDonogh, 1992; Mattick and Makunin, 2006; Schena, 1989). However, transcription factors are generally highly diverged in evolution except within their DNA-binding domains. In spite of this divergence, transcription factors can often act across species boundaries when they are expressed in heterologous organisms. The flexible mechanism by which transcription factor activation domains interact with co-regulator proteins (Hermann et al., 2001). One explanation is that the flexible mechanism by which transcription factor activation domains interacts with co-regulator proteins. This could provide an opportunity for evolving new functions without the immediate loss of the old ones. Indeed in many cases, transcription factors are adapted to be able to

interact with a wide range of structurally distinct co-regulators proteins via their activation/ repression domains (Flinn et al., 2002; McKinsey et al., 2002). This same flexible mechanism of transcription factor activation domain interaction could give transcription factors a key role in adaptation of gene expression pathways during evolution. This helps to answer one key question in biology: how several distinct organisms evolve from a small range of ancestors to create biodiversity?

Our interest in this question led us to focus this thesis study on evolutionarily conserved transcriptional co-regulator histone acetyltransferase (HATs). We used yeast as a model system to characterise how HAT proteins, in particular Gcn5, participate in re-programming expression of the genome in response to environmental changes. Further, we performed a parallel study of gene regulation using evolutionarily diverged *S. pombe* and *S. cerevisiae* yeasts cultured under similar environmental conditions. This provides a system to examine conserved regulatory components and their cellular applications in different organisms.

4.2.6 Histone acetylation and gene transcription

The phenomenon of histone acetylation in the eukaryotic cell has been known for decades. And since the early 1970s, various HAT activities haven been isolated and partially characterized. Histone acetyltransferases (HATs) apply the acetyl moiety from acetyl coenzyme A to the e-NH3+ group of internal lysine residues of histone proteins. Introduction of the acetyl group to lysine neutralizes the positive charge and increases the hydrophobicity. In the opposing deacetylation reaction, histone deacetylases (HDAC) remove the acetyl group, re-establishing the positive charge in the histone. Hyperacetylated histones have been associated with transcriptionally active genes while hypoacetylated histones are found in repressed or silenced regions. HATs can also acetylate lysine residues within transcription-related non-histone proteins (Gu and Roeder, 1997; Sterner and Berger, 2000).

One mechanism by which histone acetylation facilitates transcription is that it changes the state of chromatin structure: lysine acetylation of histone tails results in neutralizing the positively charged histone tail region, which is postulated to weaken histone-DNA binding (Hong et al., 1993; Steger and Workman, 1996) or nucleosome-nucleosome interaction (Fletcher and Hansen, 1996; Steger and Workman, 1996). This enhances the

accessibility of the DNA to transcription complexes (Anderson et al., 2001; Lee et al., 1993; Sewack et al., 2001). Another possibility is that acetylated residues in different combinations can be involved in recruiting additional transcription factors. When other tail modifications are taken into account, the combinatorial possibilities expand rapidly. This has led to the hypothesis that certain combinations of modifications in one or more tails act sequentially or concomitantly to form a histone code (Strahl and Allis, 2000). Recognition of this code via binding of specific regulatory proteins is proposed to lead to additional downstream events.

In vivo, the HATs often interact with various regulatory proteins and/or the transcription apparatus to form HAT complexes that participate in transcription. HATs can be separated into two families based on primary sequence, namely the GNAT and MYST families. Table 1 shows a summary of HATs present in *S. cerevisiae* and *S. pombe*.

Table 2 Summary of HATs in yeast

HATs	Substrate	HAT complex	Transcriptional related function	Organism
GNAT family				
Gcn5	H2B, H3	SAGA, ADA, SLIK	Transcription co-regulator	*S. cerevisiae*, *S. pombe*, *human*
Elp3	H3, H4	Elongator and Polymerase II holoenyzme	Transcriptional elongation	*S. cerevisiae*, *S. pombe*
Hat1	H4	HAT B	Telomeric silencing and DNA double-strand break repair	*S. cerevisiae*
Hpa2	H3, H4		unknown	*S. cerevisiae*
MYST family				
Esa1	H4	NuA4	Required for cell cycle progression transcriptional silencing.	*S. cerevisiae*
Sas2	Free histones	SAS complex	Regulate transcriptional silencing	*S. cerevisiae*
Sas3	H3,	NuA3	Involved in transcriptional silencing	*S. cerevisiae*
Mst1	H2A,H2B, H3, H4	NuA4	Damage response and chromosome segregation	*S. pombe*
Mst2	unknown		Negatively regulate telomeric silencing	*S. pombe*

Gcn5 (general control non-derepressible) is one of the best-characterized HATs, and is the focus of our study because it is a phylogenetically conserved transcriptional co-regulator that is found throughout the eukaryotes. Its catalytic HAT domain is the most conserved domain and has been shown to be inter-changeable between human and yeast (Wang et al., 1997). This domain can act as the catalytic subunit of several

complexes (see table 1), notably SAGA. The SAGA complex is the direct target for recruitment by transcription factors *in vitro* (Rosaleny et al., 2007; Utley et al., 1998).

Considering the importance of histone acetylation in transcriptional activation, it is somewhat surprising to discover that some HATs, such as Gcn5, are not essential in yeast. Single deletions of Gcn5 or Mst2 do not cause growth defects in *S. cerevisiae* or *S. pombe*. This implies two things: first, we need a functional model system to study HATs in gene regulation, and different environmental stress conditions are a possibility. Indeed, the Gcn5-dependent SAGA complex appears to play a key role in regulation of stress-induced transcription, via mediation of TBP binding at stress-responsive promoters (Uffenbeck and Krebs, 2006). Second, since a double mutation of HATs (gcn5Δ and mst2Δ, gcn5Δ and elp3Δ) is significantly sicker than either single mutation (paper V and (Wittschieben et al., 2000)) one can postulate overlapping functions of different HATs. This is consistent with the notions that co-regulators are functionally redundant *in vivo* (Wallberg et al., 1999). With the extensive list of histone-acetylating enzymes at hand, studying how these enzymes function co-ordinately *in vivo* will give a deeper understanding of HATs in gene regulation.

Not only the network of HAT-HAT interactions, but also the dynamic relationships between HAT and HDAC activities result in an active mode of regulation that is responsive to the ever changing demands of the cell. Shahbazian and Grunstein (Shahbazian and Grunstein, 2007) reviewed two models of HAT and HDAC interactions: one is a concerted action model where HATs are first recruited through transcriptional activators, leading to a high level of histone acetylation. HDACs are also either recruited to the active gene or act at the gene as part of their global function. When HAT activities outcompete HDAC activities on active genes, this results in hyperacetylation and transcriptional activation, and conversely when HADCs activities dominate HATs, this results in hypoacetylation and transcriptional repression. Alternatively, it could be that the overlapping specificity of HATs and HDACs is such that most lysines are acetylated by HATs, but certain lysines are stronger targets of deacetylation by HDACs. Several studies *in vitro* and *in vivo* have shown that different HATs and HDACs exhibit different substrate preference, but interplay between HAT and HDACs has been less extensively studied. In *S. pombe*, there are six HDACs: Clr6 and Hos2 (classI), Clr3 (classII) and the three class III HDACs: Sir2, Hst2 and Hst4 (Ekwall, 2005). We found that the HAT Gcn5 and the HADC Clr3 play opposite roles

in a range of stress responses (paper VI), and we further investigated the antagonistic role of Gcn5-Clr3 by means of a genome-wide approach.

At the whole genome level, HATs and HDACs have unique site preferences and many enzymes act only on certain chromosomal regions. Therefore, determination of substrate specificities of histone-modifying enzymes and their preferred chromosomal sites of action is key to understanding the distribution of histones that carry different modification patterns (Li et al., 2007; Millar and Grunstein, 2006). Genome-wide distributions of histone modification on an average gene level, in which the histones and their modifications are, mapped on an arbitrary gene relative to promoter, 5'IGR, ORF or 3'IGR. They surmised that the general acetylation pattern of H3 and H4 specifically increases at the 5'ends of genes. Based on our research comparing both normal conditions and stress conditions in *S. cerevisiae*, we found a dynamic pattern of activity for the HAT Gcn5 in *S. cerevisiae*, leading to questions of how the regulatory machinery works at the chromatin level and how Gcn5 participates in the stress response. We also sought to understand the role of Gcn5 in *S. pombe* in transcriptional elongation, based on its occupancy of the coding regions of highly transcribed genes.

5 AIMS OF THE STUDY

The aim of this project was to study gene regulation networks using microarray technology. Specifically, we are interested in studying an evolutionarily conserved co-regulator and its flexible role in adaptation of gene expression pathways during evolution. At the time when I started my PhD education, there was no commercial *S. pombe* DNA microarray available. The first goal was therefore to develop a *S. pombe* DNA microarray. The specific aims concerning technical and biological questions are listed below.

* To establish and validate a *S. pombe* cDNA expression microarray (Paper I)

* To study gene regulation networks in response to environmental cues (Paper II and Paper III)

* To study the conservation and divergence in gene regulation networks involving transcription co-regulator Gcn5 using evolutionarily divergent yeasts. (Paper IV)

*To study interactions between transcription co-regulators (interaction between HATs, and interaction between HATs and HDACs) (Paper V and VI)

6 RESULTS AND DISCUSSIONS

6.1 ESTABLISHMENT AND VALIDATION OF A *S. POMBE* DNA MICROARRAY (PAPER I)

The aim of this study was to construct and validate a fission yeast cDNA microarray, which permits genome-wide expression analysis in *Schizosaccharomyces pombe*. Here we report microarray design, construction, exprerimental procedure data qualification and data analysis.

The microarray contains DNA fragments, PCR-amplified from a genomic DNA template that represent 99% of the 5000 or so annotated *S. pombe* genes, as well as a number of control sequences.

To validate our design principles, we investigated the influence on hybridization of fragment length, relative fragment position, GC content and intron content. Further we tested the reproducibility of our data by comparing independent labelling microarray experiments.

In order to test microarray performance, we performed yeast temperature shift experiments, in which the low magnitude of fold changes makes the identification of reproducibly altered genes difficult but, on the other hand, it provides a good test of microarray performance. When temperature for yeast culture shift from 25 ℃ to 36 ℃. We observed that the vast majority of genes do not change more than two-fold, supporting the widely held view that temperature-shift experiments specifically reveal phenotypes associated with temperature-sensitive mutants. However, we did identify a small group of genes that showed a reproducible large magnitude change in expression. Importantly, most of these genes corresponded to previously characterized heat-shock genes, whose expression has been reported to change after more extreme temperature shifts than those used here.

This study provides information required to contruct the array from scratch as well as to use the complete array. It presents a useful resouce for fission yeast researchers as well as the broader yeast community, since it will facilate comparison with the distantly related budding yeast, *Saccharomyces cerevisiae.*

6.2 GENOME-WIDE IDENTIFICATION OF PHEROMONE-TARGETED TRANSCRIPTION IN FISSION YEAST (PAPER II)

In *S. pombe,* nutritional starvation is the major signal that activates sexual differentiation. As long as the nutritional conditions are favourable, haploid cells will propagate vegetatively, but under nitrogen starvation, the cells exit from the mitotic cycle and undergo a differentiation process, which requires sexual agglutination, conjugation, nuclear fusion, meiosis and spore formation. The process of conjugation involves the action of diffusible pheromones secreted by P and M cell types in order to attract cells of the opposite mating type. When exposed to the opposite pheromone, the cells form projections toward each other, and fuse upon cell-cell contact.

We use a *cyr1* gene genetic background which allowed us to study pheromone signalling indepdendently of nitrogen starvation. When h+ and h- *cyr1*- disruptants are mixed, they undergo mating even in rich medium without nitrogen starvation. In these mutant strains, induction of pheromone-induced genes can thus be decoupled from other aspects of the nitrogen starvation response.

We identified classes of genes that are consistently induced by pheromone, and found that genes induced by M-factor and P-factor stimulation were highly overlapping; this is consistent with the fact that pheromone response pathway is common to both mating types. This group of phermone controlled genes have an elevalated occurence of ste11 binding sites.

We also discovered new pheromone-induced genes involved in the differentiation process, which we validated and functionally characterised using PCR and yeast genetic methods.

We further identified cell type specific genes by direct comparsion of the M- and P-factor induced expression pattern. In this manner we found that pheromone control suprisingly extended to genes fulfiling their function well beyond the point of entry into meiosis, including numerous genes required for meiotic recombination. The results also suggest that the ste11 transcription factor is responsible for the majority of pheromone-induced transcription. Our observation of P-specfic transcripts extending into the

neighbouring genes demonstrates the importatnce of genome mapping and this information will be useful for future microarray design.

6.3 THE FISSION YEAST GCN5 PROTEIN REPROGRAMS GENE EXPRESSION IN RESPONSE TO ENVIRONMENTAL CHANGES (PAPER III)

In this paper, we studied how the transcriptonal co-regulator, Gcn5 histone acetyltransferase, contributes to reprogramming genome expression in response to the external signal KCl, in fission yeast.

First we found that Gcn5 is not requried for cell growth and proliferation but specifically required for adaptation to KCl and $CaCl_2$ mediated stress. We have characterized the genome-wide gene expression response to KCl stress and show that Gcn5 in involved in the regulation of a subset of common stress response genes. Gcn5 is most clearly associated with KCl-induced genes, but there is no correlation between Gcn5 dependence and the extent of their induction.

We report that Gcn5-dependent KCl-induced genes are specifically enriched in four different DNA motifs. The Gcn5-dependent KCl-induced genes are also associated with biological process gene ontology terms such as carbohydrate metabolism, glycolysis, and nicotinamide metabolism that together constitute a subset of the ontology parameters associated with KCl-induced genes.

This study shows Gcn5 plays important role in stress responses and the stress conditions used here provide a good model for investigating gene regulation controlled by Gcn5.

6.4 CONSERVED AND DIVERGED ROLES OF GCN5 HISTONE ACEETYLTRANSFERASE (PAPER IV)

In this paper, we characterized the role of Gcn5 histone acetyltransferase during stress adaptation in *S. cerevisiae* and compared its role in the evolutionarily diverged yeast, *S. pombe.*

We showed that Gcn5 is important for a specific set of stress responses in evolutionarily diverged yeast species including *S. cerevisiae, S. pombe* and *S. kluyveri.*

We proved that the conserved histone acetyltransferase domain is required for stress adaptation in *S. cerevisiae.*

We defined a group of KCl stress response genes in *S. cerevisiae* that are specifically dependent on Gcn5. We found that Gcn5 represses the *FLO8* gene, which provides support for a role of Gcn5 as a corepressor as well as a coactivator. Gcn5 localization studies showed a shift in the distribution of Gcn5 within genes to the coding region during KCl stress adaptation. Gcn5 is preferentially localized within highly expressed genes and long genes during stress adaptation.

Comparative analysis of gene regulation by Gcn5 during KCl stress adaptation between *S. cerevisiae* and *S. pombe* showed that Gcn5 regulates divergent sets of KCl responsive genes even though it regulates the same stress response, adaptation to KCl in both organisms. The shift in Gcn5 localization during normal conditions and stress conditions in *S. cerevisiae* is not observed in *S. pombe.* The results are significant for understanding the molecular mechanisms by which Gcn5 regulates transcription as well as for an appreciation of their conservation or divergence of gene regulation networks in different eukaryotic organisms.

6.5 REDUNDANT GENE REGULATION OF HISTONE ACETYLTRANSFERASES (PAPER V)

In paper V we studied the gene expression profiles of different HAT mutants (elp3Δ, gcn5Δ, mst2Δ and mst1) in *S. pombe* and identified redundant functions in gene regulation.

First, we isolated fission yeast elp3+, the gene encoding the histone acetyltransferase (HAT) subunit of the transcription elongator complex. We characterized *elp3Δ* and its interaction with other HAT mutants: *gcn5*Δ and *mst2*Δ. Next, using whole-genome microarrays, we compared the effects of Δ*elp3* on gene expression to the effects of two other HAT deletion strains, Δ*mst2* and Δ*gcn5*, Further we studied global acetylation levels in HAT deficient strains.

Comparison of phenotypes and gene expression profiles in single, double and triple mutants suggests that Gcn5 and Mst2 have overlapping functions. Consistently, overlapping specificity in histone H3 acetylation is observed.

6.6 HISTONE ACETYLTRANSFERASE-DEACETYLASE INTERPLAY IN STRESS RESPONSE (PAPER VI)

In this paper, we used ChIP-on-chip high resolution tiling arrays and expression microarrays to study HAT-HDAC interaction in fission yeast.

First we mapped genome-wide Gcn5 occupancy and found that Gcn5 is most strongly associated in the coding region, suggesting a role for Gcn5 in transcriptional elongation in addtion to its known role in transcriptional initiation. We confimed experimentally this role of Gcn5 in transcriptional elongation by MPA assay.

Next we found that Clr3, and not the other HATs, can antagonise Gcn5 to modulate H3K14ac levels, elongation and stress responses.

Finally, gene expression profiling showed that clr3Δ mediated changes reversed gene expression changes resulting from deltion of gcn5.

The results show an important role for H3K14 acetylation in stress responses. Interplay between Gcn5, Clr3 and H3K14 affects the transcriptional elongation efficiency of important stress genes.

7 FUTURE PERSPECTIVES

DNA microarray-based assays have moved from a technology-driven early stage to application-oriented high-output era. The technology allows investigators to make use of arrays of probes ranging from a few thousand to a few million sequences. At the same time, bioinformatics and statistical approaches essential to analyse, store and communicate the huge amount of information generated by this technology are maturing in parallel. DNA microarray technology has conquered highly diversified fields to even prove to be mature enough for clinical applications (Draghici et al., 2006; Shi et al., 2008; Shi et al., 2006).

At the stage where DNA microarrays have succeeded in silencing most of the early critics (on matters such as reproducibility, shortcomings of hybridization based methods and so on), it has to face new challengers: the new sequencing instruments (also called next generation or massively parallel) which sequence millions of DNA fragments on a single chip in parallel. These sequence read-outs can be first mapped onto a reference genome, and then counted, their distribution throughout the genome analysed. This approach presents several advantages over the hybridization-based microarray methodologies, making it successful in several applications, notably ncRNA profiling, DNA-protein interaction, SNP and mutation discovery. Here I will address advantages of next generation sequencing compared to DNA microarray in ncRNA profiling and DNA-protein interaction assays.

One key advantage of next generation sequencing is that unknown or incompletely known genomes can be analysed since no sequence specific probes are needed for detections. It allows the discovery of novel sequences, especially ncRNA. Genome location of ncRNA is complicated; certain classes of ncRNAs are poorly conserved over evolution. DNA microarray approaches can target only the known miRNAs previously identified by sequencing or homology searches (Yin et al., 2008). In contrast, next generation sequencing can readily discover novel ncRNAs. Additionally, the length of short ncRNA is typically around 21-35 bp. Low molecular weight and low abundance detection is challenging for DNA microarray technology, while sequencing can distinguish single nucleotide differences.

Replacing ChIP-on-chip, ChIP-seq approaches can be used to detect DNA-protein interactions. Due to the short size of the reads, ChIP-seq approaches provide a better resolution in the definition and DNA-protein interaction sites, and comparison studies also show fewer replicates are needed; therefore less DNA material is required (Euskirchen et al., 2007).

With its enormous potential in life science application, will next generation sequencing take over from DNA microarray in the future? Aside from the fact that it is still young and issues regarding cost still need to be worked out, the new methods can not solve everything: although a large fraction of genome is accessible by Seq-based methods, 15-20% of the reads in human genome cannot be unambiguously mapped to a single location because they occur more than once in the genome. Furthermore, the field faces bioinformatic challenges such as analysis algorithm improvement. Over time more applications will migrate from microarray to sequencing platforms, while at the moment these two approaches appear to coexist. Indeed, combination or comparison studies of sequencing approaches with ChIP-on-chip, or gene expression profiling give more convincing results and more complete information about DNA-protein interactions or transcriptomes (Euskirchen et al., 2007; Liu et al., 2007; Oudes et al., 2005).

In this PhD work, we have mostly presented the gene expression profiles and DNA-protein interaction patterns in different yeast species by expression array and ChIP-on-chip approaches. We have not obtained clear information about mechanistic details of gene regulation. For example, in paper V, how do different HATs collaborate with each other to achieve the redundant roles? ChIP-on-chip or ChIP-seq studies would help us to read the "histone code" of HATs: whether they have overlapped acetylation sites, how they collaborate under normal or stress conditions? In paper IV, we observed a switch in Gcn5 localization within genes to the coding region in the *S. cerevisiae*. This altered pattern between normal conditions and stress conditions is not observed in *S. pombe*; what is the mechanistic difference in stress response between these two yeast species? Would the Gcn5 localization pattern switch back when *S. cerevisiae* is returned to normal conditions? We found Gcn5 is also localized to non-coding RNA genes; how is ncRNA involved in stress responses? Why does gene transcription not always occur at genes where Gcn5 is localised? (We only find 30% percent of Gcn5 localized genes are transcriptionally regulated by this factor.) How does a protein (such

as Gcn5) control different sets of target genes while maintaining the same physiological role in evolutionarily diverged species? Understanding the constraints of its gene regulation might be a key to transferring knowledge obtained from model organisms to human applications.

8 ACKNOWLEDGEMENTS

This work is financially supported by Södertörn University, EU grant and the Swedish Research Council.

As a student from a foreign country, there is always a lot to say when comes to this important moment in life.

Thank you, **Tony**, my supervisor, for taking me as a student when I am at the deepest valley in my life. Thank you for the weekly discussions about science, which lead me step by step to maturity in research. Thank you for providing a friendly working environment, and being supportive and considerate all the time. Also, thanks for your sense of humour and positive attitude to life.

Professor **Gong Ju**, Thanks for being so inspiring both in sciences and life! For opening the door to the world for me, for being a forever understanding friend!

Karin Dahlman-Wright, my former supervisor, for your encouragement and being such a cheerful person.

All **my collaborators**, for your contributions to make this thesis possible!

Karl Ekwall, **Claes Gustafsson,** and all the present and previous yeast group members: **Helmi**, for being so helpful and responsible, and for daily "fika" time! **Anna**, thanks for the communication in sciences and life. **Rachel**, everyone feels lucky to have you in the group! Thanks for your help with English language in manuscript and thesis writing. You are so patient in listening to everyone's trouble. **Chiounan,** you are always the one who helps to take care of my samples. **Johan N,** for dragging me to the gym, I need it! All the other lab members: **Monica, Leila, Fredrik, Azadeh, Ingela**, **Carolina**, **Mariana**, **Julian**, **Pernilla**, **Indranil**,, **Mickaël**, **Michelle**, **Andreas**, **Ulrika**, **Annelie**, **Jenna**, **Vera**, **Olga** and the rest.

People in Södertörns and Novum**: Johan H** and **David Brodin**, it is such enjoyment to collaborate with you! **Yann,** for proof reading my manuscript and thesis, plus the fun time at Södertörn. **Gayathri**, **Satishi**, **Qi Dai, Agata, Ida, johan Deffs, Ivo, YongGuang, Louise** and **Doctor Stefan,** for your friendship.

Administration people at Novum and SH**, Lena**, **Monica** and **Marie,** for all these years of help and the warm feelings you give. **Sam** and **Asim**, for helping with computer problems with amazing speed. **Marie Granroth**, **Lotta, Elisabeth, Helena Fors** and the rest.

Previous colleagues at Novum, **Anette**, **Maria** and **Louvisa, Jurgen**, you are so nice to me. **Elisabeth** at KI Solna core facility for the early time of microarray.

All my friends, for your support and help on every occasion! And the enjoyable gathering time! **Ming** (for every lunch meeting!) and **David**, **Bjorn** and **Camilla**, **Tao**

and **Stefan**, **YongHong** and **XueFeng**, **Rebecca** and **Sam**, **Thomas** and **Elinor** (Thank you so much for your hosting in Gotland), **QingYang** and **Xiao Niu**, **Ann** and **David, WuHao** and **MoHao**, **Zhang Pu** and **XiaoLei**, **Dawei** and **Tian Fei**, **Jonny** and **Maria**, **Elinor** and **Jim**, **Lisa**, **Bjorn** and **Jenny**, **Wang Xiao** and **JianGuang**, **ChunYan, Hui Gao** and **Wen**, **XiaoFeng Zhang**, and my Korean friend, **Jin**.

Young and beautiful people in Stallbacken! I really love you! You made the after-work time so lovely, **HuiXin** and **JinJing** (for all the communications), **Karin** and **Victor** (I enjoy so much hanging out with you!), **Mikael and Malja** (for being a Kevin-lover ☺), **Marie and Ulf (**for tolerating William's endless disturbances), **Daniel and Ann-Marie** (for every late night party!), **Unn and John, Anders and Mona, Kalle** and **Thomas** (for your entertaining garden that charmed my children!), **Anna (**for the passionate Russian hugs**), Paula** and **Magnus**, **Sanna** and **Peter, Richard (**for your music!) and **Malin.** Also, **Lotten** and **Ingvar** at Vårsta for all the good memories!

Special thanks for **Gunilla** for teaching me guitar and Swedish music culture, **Mona Sahlin** and **Mia Bielkheden** for teaching me Swedish Language!

Family in Sweden, **Bertil** and **Marita**, Thank you for showing up every time when we need help! Thank for your love to grandchildren! Tack för att ni visa er upp varje gång när vi behöver hjälp! Also thanks for you love to grandchildren! **Ulla** and **Rune,** for giving me such a warm welcome when I just arrived to Sweden (I was already one of the family from then !), thanks for all the dinner and cakes! **Marie**, as I always say, how lucky I am to have you as a sister! You are so cheerful, reliable, considerate, supportive, competent, and beautiful! And **Anders**, **Ann-sofie** and **Jocke**, for being so energetic and full of fun (or craziness?)! **Mattias** and **Matilda**, we are so proud of you!

Mom and **Dad**, thank you by all my heart, for your endless giving! I am proud of you being so open-minded and wise! 亲爱的爸妈, 感谢你们对儿女无止境地付出! 我对你们的开明和智慧深感自豪。感谢舅舅，俐俐姨，小姨和姨父多年的关心和爱护! My sister **Grace**, I am always proud of you! For your love, and for sharing everything! **Robert**, for showing your love by heart! 薛峰和桑华，感谢你们照顾父母， 感谢07年的美好记忆！

William and **Kevin**, my beautiful boys! You fill up my life with joy in all dimensions! Christoffer, my bonus son, sensitive, considerate, full of love!

Mats, my dearest, thank you, for being a caring husband, a loving father, and a friend everyone likes. You love makes me feel that I am the richest in the world, and the strongest, or the weakest… I cannot imagine a life without you.

9 REFERENCES

Anderson, J.D., Lowary, P.T. and Widom, J. (2001) Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Biol*, **307**, 977-985.

Barnett, J.A. (2003) Beginnings of microbiology and biochemistry: the contribution of yeast research. *Microbiology*, **149**, 557-567.

Bender, R. and Lange, S. (2001) Adjusting for multiple testing--when and how? *J Clin Epidemiol*, **54**, 343-349.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., Sharon, E., Spector, Y. and Bentwich, Z. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet*, **37**, 766-770.

Berezikov, E., van Tetering, G., Verheul, M., van de Belt, J., van Laake, L., Vos, J., Verloop, R., van de Wetering, M., Guryev, V., Takada, S., van Zonneveld, A.J., Mano, H., Plasterk, R. and Cuppen, E. (2006) Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. *Genome Res*, **16**, 1289-1298.

Botstein, D., Chervitz, S.A. and Cherry, J.M. (1997) Yeast as a model organism. *Science*, **277**, 1259-1260.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S. and Fodor, S.P. (1996) Accessing genetic information with high-density DNA arrays. *Science*, **274**, 610-614.

Chudin, E., Kruglyak, S., Baker, S.C., Oeser, S., Barker, D. and McDaniel, T.K. (2006) A model of technical variation of microarray signals. *J Comput Biol*, **13**, 996-1003.

Clark, T.A., Sugnet, C.W. and Ares, M., Jr. (2002) Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science*, **296**, 907-910.

Cosgrove, M.S., Boeke, J.D. and Wolberger, C. (2004) Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol*, **11**, 1037-1043.

Dabney, A.R. and Storey, J.D. (2007) Normalization of two-channel microarrays accounting for experimental design and intensity-dependent relationships. *Genome Biol*, **8**, R44.

Djupedal, I., Portoso, M., Spahr, H., Bonilla, C., Gustafsson, C.M., Allshire, R.C. and Ekwall, K. (2005) RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev*, **19**, 2301-2306.

Dobbin, K., Shih, J.H. and Simon, R. (2003) Statistical design of reverse dye microarrays. *Bioinformatics*, **19**, 803-810.

Dobbin, K. and Simon, R. (2002) Comparison of microarray designs for class comparison and class discovery. *Bioinformatics*, **18**, 1438-1445.

Dobbin, K.K., Shih, J.H. and Simon, R.M. (2005) Comment on 'Evaluation of the gene-specific dye bias in cDNA microarray experiments'. *Bioinformatics*, **21**, 2803-2804.

Draghici, S., Khatri, P., Eklund, A.C. and Szallasi, Z. (2006) Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet*, **22**, 101-109.

Duboit, S., Yang, Y. Callow, M. J. and Speed T.P. (2000) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Technical report #578.

Dudoit, S., Gentleman, R.C. and Quackenbush, J. (2003) Open source software for the analysis of microarray data. *Biotechniques*, **Suppl**, 45-51.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.M., Beyne, E., Bleykasten, C., Boisrame, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L.,

Muller, H., Nicaud, J.M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.F., Straub, M.L., Suleau, A., Swennen, D., Tekaia, F., Wesolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P. and Souciet, J.L. (2004) Genome evolution in yeasts. *Nature*, **430**, 35-44.

Ekwall, K. (2005) Genome-wide analysis of HDAC function. *Trends Genet*, **21**, 608-615.

Euskirchen, G.M., Rozowsky, J.S., Wei, C.L., Lee, W.H., Zhang, Z.D., Hartman, S., Emanuelsson, O., Stolc, V., Weissman, S., Gerstein, M.B., Ruan, Y. and Snyder, M. (2007) Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. *Genome Res*, **17**, 898-909.

Fletcher, T.M. and Hansen, J.C. (1996) The nucleosomal array: structure/function relationships. *Crit Rev Eukaryot Gene Expr*, **6**, 149-188.

Flinn, E.M., Wallberg, A.E., Hermann, S., Grant, P.A., Workman, J.L. and Wright, A.P. (2002) Recruitment of Gcn5-containing complexes during c-Myc-dependent gene activation. Structure and function aspects. *J Biol Chem*, **277**, 23399-23406.

Goodson, M., Jonas, B.A. and Privalsky, M.A. (2005) Corepressors: custom tailoring and alterations while you wait. *Nucl Recept Signal*, **3**, e003.

Goryachev, A.B., Macgregor, P.F. and Edwards, A.M. (2001) Unfolding of microarray data. *J Comput Biol*, **8**, 443-461.

Gu, W. and Roeder, R.G. (1997) Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell*, **90**, 595-606.

Guarente, L. and Bermingham-McDonogh, O. (1992) Conservation and evolution of transcriptional mechanisms in eukaryotes. *Trends Genet*, **8**, 27-32.

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. and Collins, F.S. (1999) Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nat Genet*, **22**, 164-167.

Havas, K., Flaus, A., Phelan, M., Kingston, R., Wade, P.A., Lilley, D.M. and Owen-Hughes, T. (2000) Generation of superhelical torsion by ATP-dependent chromatin remodeling activities. *Cell*, **103**, 1133-1142.

Helmlinger, D., Marguerat, S., Villen, J., Gygi, S.P., Bahler, J. and Winston, F. (2008) The S. pombe SAGA complex controls the switch from proliferation to sexual differentiation through the opposing roles of its subunits Gcn5 and Spt8. *Genes Dev*, **22**, 3184-3195.

Hermann, S., Berndt, K.D. and Wright, A.P. (2001) How transcriptional activators bind target proteins. *J Biol Chem*, **276**, 40127-40132.

Hirota, K., Miyoshi, T., Kugou, K., Hoffman, C.S., Shibata, T. and Ohta, K. (2008) Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature*, **456**, 130-134.

Hong, L., Schroth, G.P., Matthews, H.R., Yau, P. and Bradbury, E.M. (1993) Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. *J Biol Chem*, **268**, 305-314.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M. and Brown, P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, **409**, 533-538.

Jeon, H.C., S. (2007) Fluorescence quenching causes systematic dye bias in microarray experiments using cyanine dye. *Genomics & Informatics*, **5**, 113-117.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B. and Lander, E.S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, **423**, 241-254.

Kerr, M.K., Martin, M. and Churchill, G.A. (2000) Analysis of variance for gene expression microarray data. *J Comput Biol*, **7**, 819-837.

Kingston, R.E., Schuetz, T.J. and Larin, Z. (1987) Heat-inducible human factor that binds to a human hsp70 promoter. *Mol Cell Biol*, **7**, 1530-1534.

Knudsen, S. (2002) *A biologist's guide to analysis of DNA microarray data*. John Wiley &sons, Inc., New York.
Kornberg, R.D. (1974) Chromatin structure: a repeating unit of histones and DNA. *Science*, **184**, 868-871.
Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O. and Davis, R.W. (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc Natl Acad Sci U S A*, **94**, 13057-13062.
Latchman, D.S. (2005) *Gene regulation*. Taylor & Francis Group.
Lazar, M.A. (2003) Nuclear receptor corepressors. *Nucl Recept Signal*, **1**, e001.
Lee, D.Y., Hayes, J.J., Pruss, D. and Wolffe, A.P. (1993) A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell*, **72**, 73-84.
Li, B., Carey, M. and Workman, J.L. (2007) The role of chromatin during transcription. *Cell*, **128**, 707-719.
Liang, P. and Pardee, A.B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, **257**, 967-971.
Lieb, J.D., Liu, X., Botstein, D. and Brown, P.O. (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat Genet*, **28**, 327-334.
Liu, F., Jenssen, T.K., Trimarchi, J., Punzo, C., Cepko, C.L., Ohno-Machado, L., Hovig, E. and Kuo, W.P. (2007) Comparison of hybridization-based and sequencing-based gene expression technologies on biological replicates. *BMC Genomics*, **8**, 153.
Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F. and Richmond, T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, **389**, 251-260.
Mariani, T.J., Budhraja, V., Mecham, B.H., Gu, C.C., Watson, M.A. and Sadovsky, Y. (2003) A variable fold change threshold determines significance for expression microarrays. *Faseb J*, **17**, 321-323.
Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. *Hum Mol Genet*, **15 Spec No 1**, R17-29.
McKinsey, T.A., Zhang, C.L. and Olson, E.N. (2002) MEF2: a calcium-dependent regulator of cell division, differentiation and death. *Trends Biochem Sci*, **27**, 40-47.
Millar, C.B. and Grunstein, M. (2006) Genome-wide patterns of histone modifications in yeast. *Nat Rev Mol Cell Biol*, **7**, 657-666.
Moran, G., Stokes, C., Thewes, S., Hube, B., Coleman, D.C. and Sullivan, D. (2004) Comparative genomics using Candida albicans DNA microarrays reveals absence and divergence of virulence-associated genes in Candida dubliniensis. *Microbiology*, **150**, 3363-3382.
Orphanides, G., Lagrange, T. and Reinberg, D. (1996) The general transcription factors of RNA polymerase II. *Genes Dev*, **10**, 2657-2683.
Oudes, A.J., Roach, J.C., Walashek, L.S., Eichner, L.J., True, L.D., Vessella, R.L. and Liu, A.Y. (2005) Application of Affymetrix array and Massively Parallel Signature Sequencing for identification of genes involved in prostate cancer progression. *BMC Cancer*, **5**, 86.
Parker, C.S. and Topol, J. (1984) A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. *Cell*, **37**, 273-283.
Peterson, C.L. and Laniel, M.A. (2004) Histones and histone modifications. *Curr Biol*, **14**, R546-551.
Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D. and Brown, P.O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet*, **23**, 41-46.
Provost, P., Silverstein, R.A., Dishart, D., Walfridsson, J., Djupedal, I., Kniola, B., Wright, A., Samuelsson, B., Radmark, O. and Ekwall, K. (2002) Dicer is required for chromosome segregation and gene silencing in fission yeast cells. *Proc Natl Acad Sci U S A*, **99**, 16648-16653.
Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T.L., Wilson, C.J., Bell, S.P. and

Young, R.A. (2000) Genome-wide location and function of DNA binding proteins. *Science*, **290**, 2306-2309.
Robert, F., Pokholok, D.K., Hannett, N.M., Rinaldi, N.J., Chandy, M., Rolfe, A., Workman, J.L., Gifford, D.K. and Young, R.A. (2004) Global position and recruitment of HATs and HDACs in the yeast genome. *Mol Cell*, **16**, 199-209.
Rosaleny, L.E., Ruiz-Garcia, A.B., Garcia-Martinez, J., Perez-Ortin, J.E. and Tordera, V. (2007) The Sas3p and Gcn5p histone acetyltransferases are recruited to similar genes. *Genome Biol*, **8**, R119.
Schena, M. (1989) The evolutionary conservation of eukaryotic gene transcription. *Experientia*, **45**, 972-983.
Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467-470.
Sewack, G.F., Ellis, T.W. and Hansen, U. (2001) Binding of TATA binding protein to a naturally positioned nucleosome is facilitated by histone acetylation. *Mol Cell Biol*, **21**, 1404-1415.
Shahbazian, M.D. and Grunstein, M. (2007) Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem*, **76**, 75-100.
Shi, L., Perkins, R.G., Fang, H. and Tong, W. (2008) Reproducible and reliable microarray results through quality control: good laboratory proficiency and appropriate data analysis practices are essential. *Curr Opin Biotechnol*, **19**, 10-18.
Shi, L., Reid, L.H., Jones, W.D., Shippy, R., Warrington, J.A., Baker, S.C., Collins, P.J., de Longueville, F., Kawasaki, E.S., Lee, K.Y., Luo, Y., Sun, Y.A., Willey, J.C., Setterquist, R.A., Fischer, G.M., Tong, W., Dragan, Y.P., Dix, D.J., Frueh, F.W., Goodsaid, F.M., Herman, D., Jensen, R.V., Johnson, C.D., Lobenhofer, E.K., Puri, R.K., Schrf, U., Thierry-Mieg, J., Wang, C., Wilson, M., Wolber, P.K., Zhang, L., Amur, S., Bao, W., Barbacioru, C.C., Lucas, A.B., Bertholet, V., Boysen, C., Bromley, B., Brown, D., Brunner, A., Canales, R., Cao, X.M., Cebula, T.A., Chen, J.J., Cheng, J., Chu, T.M., Chudin, E., Corson, J., Corton, J.C., Croner, L.J., Davies, C., Davison, T.S., Delenstarr, G., Deng, X., Dorris, D., Eklund, A.C., Fan, X.H., Fang, H., Fulmer-Smentek, S., Fuscoe, J.C., Gallagher, K., Ge, W., Guo, L., Guo, X., Hager, J., Haje, P.K., Han, J., Han, T., Harbottle, H.C., Harris, S.C., Hatchwell, E., Hauser, C.A., Hester, S., Hong, H., Hurban, P., Jackson, S.A., Ji, H., Knight, C.R., Kuo, W.P., LeClerc, J.E., Levy, S., Li, Q.Z., Liu, C., Liu, Y., Lombardi, M.J., Ma, Y., Magnuson, S.R., Maqsodi, B., McDaniel, T., Mei, N., Myklebost, O., Ning, B., Novoradovskaya, N., Orr, M.S., Osborn, T.W., Papallo, A., Patterson, T.A., Perkins, R.G., Peters, E.H., Peterson, R., Philips, K.L., Pine, P.S., Pusztai, L., Qian, F., Ren, H., Rosen, M., Rosenzweig, B.A., Samaha, R.R., Schena, M., Schroth, G.P., Shchegrova, S., Smith, D.D., Staedtler, F., Su, Z., Sun, H., Szallasi, Z., Tezak, Z., Thierry-Mieg, D., Thompson, K.L., Tikhonova, I., Turpaz, Y., Vallanat, B., Van, C., Walker, S.J., Wang, S.J., Wang, Y., Wolfinger, R., Wong, A., Wu, J., Xiao, C., Xie, Q., Xu, J., Yang, W., Zhang, L., Zhong, S., Zong, Y. and Slikker, W., Jr. (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol*, **24**, 1151-1161.
Sorger, P.K. and Pelham, H.R. (1987) Purification and characterization of a heat-shock element binding protein from yeast. *Embo J*, **6**, 3035-3041.
Steger, D.J. and Workman, J.L. (1996) Remodeling chromatin structures for transcription: what happens to the histones? *Bioessays*, **18**, 875-884.
Sterner, D.E. and Berger, S.L. (2000) Acetylation of histones and transcription-related factors. *Microbiol Mol Biol Rev*, **64**, 435-459.
Stoughton, R.B. (2005) Applications of DNA microarrays in biology. *Annu Rev Biochem*, **74**, 53-82.
Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41-45.
Tseng, G.C., Oh, M.K., Rohlin, L., Liao, J.C. and Wong, W.H. (2001) Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res*, **29**, 2549-2557.

Uffenbeck, S.R. and Krebs, J.E. (2006) The role of chromatin structure in regulating stress-induced transcription in Saccharomyces cerevisiae. *Biochem Cell Biol*, **84**, 477-489.
Uhler, J.P., Hertel, C. and Svejstrup, J.Q. (2007) A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc Natl Acad Sci U S A*, **104**, 8011-8016.
Utley, R.T., Ikeda, K., Grant, P.A., Cote, J., Steger, D.J., Eberharter, A., John, S. and Workman, J.L. (1998) Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature*, **394**, 498-502.
van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R. and Friend, S.H. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, **415**, 530-536.
Vaquero, A., Loyola, A. and Reinberg, D. (2003) The constantly changing face of chromatin. *Sci Aging Knowledge Environ*, **2003**, RE4.
Varringer, J. (2003) Genetic robustness during environmental stress. *Department of Cell and Molecular Biology*. Gothenborg Univeristy, Gothenborg, Vol. Ph.D.
Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I. and Martienssen, R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, **297**, 1833-1837.
Wagner, G.P. and Lynch, V.J. (2008) The gene regulatory logic of transcription factor evolution. *Trends Ecol Evol*, **23**, 377-385.
Wallberg, A.E., Neely, K.E., Gustafsson, J.A., Workman, J.L., Wright, A.P. and Grant, P.A. (1999) Histone acetyltransferase complexes can mediate transcriptional activation by the major glucocorticoid receptor activation domain. *Mol Cell Biol*, **19**, 5952-5959.
Wang, L., Mizzen, C., Ying, C., Candau, R., Barlev, N., Brownell, J., Allis, C.D. and Berger, S.L. (1997) Histone acetyltransferase activity is conserved between yeast and human GCN5 and is required for complementation of growth and transcriptional activation. *Mol Cell Biol*, **17**, 519-527.
Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K. and Kurokawa, R. (2008) Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. *Nature*, **454**, 126-130.
Wang, X., Ghosh, S. and Guo, S.W. (2001) Quantitative quality control in microarray image processing and data acquisition. *Nucleic Acids Res*, **29**, E75-75.
Wittschieben, B.O., Fellows, J., Du, W., Stillman, D.J. and Svejstrup, J.Q. (2000) Overlapping roles for the histone acetyltransferase activities of SAGA and elongator in vivo. *Embo J*, **19**, 3060-3068.
Wood, V., Gwilliam, R., Rajandream, M.A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor, R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E.J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S., Pearson, D., Quail, M.A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R.G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schafer, M., Muller-Auer, S., Gabel, C., Fuchs, M., Dusterhoft, A., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R., Pohl, T.M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dreano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S.J., Xiang, Z., Hunt, C., Moore, K., Hurst, S.M., Lucas, M., Rochet, M., Gaillardin, C., Tallada, V.A., Garzon, A., Thode, G., Daga, R.R., Cruzado, L., Jimenez, J., Sanchez, M., del Rey, F., Benito, J., Dominguez, A., Revuelta, J.L., Moreno, S., Armstrong, J., Forsburg, S.L., Cerutti, L., Lowe, T., McCombie, W.R., Paulsen, I., Potashkin, J.,

Shpakovski, G.V., Ussery, D., Barrell, B.G. and Nurse, P. (2002) The genome sequence of Schizosaccharomyces pombe. *Nature*, **415**, 871-880.
Wright, A.P. (2005) Raising the standard - yeast as a model eukaryote: the first 70 years and beyond. *The biochemist*, 9.
Wu, C. (1985) An exonuclease protection assay reveals heat-shock element and TATA box DNA-binding proteins in crude nuclear extracts. *Nature*, **317**, 84-87.
Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J. and Speed, T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*, **30**, e15.
Yang, Y.H.D., S. Luu, P. Lin D. M. Peng, V, Ngai, J. and Speed, T. P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*, **30**, e15.
Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S. and Fu, X.D. (2002) Profiling alternative splicing on fiber-optic arrays. *Nat Biotechnol*, **20**, 353-358.
Yin, J.Q., Zhao, R.C. and Morris, K.V. (2008) Profiling microRNA expression with microarrays. *Trends Biotechnol*, **26**, 70-76.
Zakharkin, S.O., Kim, K., Mehta, T., Chen, L., Barnes, S., Scheirer, K.E., Parrish, R.S., Allison, D.B. and Page, G.P. (2005) Sources of variation in Affymetrix microarray experiments. *BMC Bioinformatics*, **6**, 214.