Regulation of Transcription in the Fission Yeast *Schizosaccharomyces pombe*

A Dissertation Presented

by

Adam Phillip Rosebrock

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Molecular Genetics and Microbiology

Stony Brook University

August 2009

Stony Brook University

The Graduate School

Adam Phillip Rosebrock

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation.

A. Bruce Futcher, D. Phil – Dissertation Advisor
Professor, Department of Molecular Genetics and Microbiology

Aaron Neiman, PhD – Chairperson of Defense
Associate Professor, Department of Biochemistry and Cell Biology

J. Peter Gergen, PhD
Professor, Department of Biochemistry and Cell Biology

Janet K. Leatherwood, PhD
Associate Professor, Department of Molecular Genetics and Microbiology

Steven Skiena, PhD
Professor, Department of Computer Science

This dissertation is accepted by the Graduate School.

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Regulation of Transcription in the Fission Yeast *Schizosaccharomyces pombe*

by

Adam Phillip Rosebrock

Doctor of Philosophy

in

Molecular Genetics and Microbiology

Stony Brook University

2009

Although eukaryotic genomes contain a wide repertoire of sequence elements, only a subset are highly transcribed during any single stage of growth. Proper regulation of various transcriptional programs enables both the cyclic behavior of dividing cells and the capacity to respond to environmental or developmental cues. Conversely, dysregulated transcription is a hallmark of many diseases including cancer.

Previous high-throughput studies have successfully employed a reductionist approach of quantitating predicted functional transcripts. Using these methods, I have characterized the cell-cycle regulated genes of the fission yeast, *Schizosaccharomyces pombe*, assigned putative transcription factors to many of the oscillating genes, and identified transcription programs and regulatory schemes present from yeast to man.

In subsequent studies, I have shown that far more of the genome is transcribed than is accounted for by traditional expression microarrays; unbiased measurements using high-resolution strand-specific tiling arrays revealed that processed RNAs are generated from greater than 90% of the *S. pombe* genome, including previously unappreciated antisense species. I have identified hundreds of discrete non-coding RNAs that result from bidirectional activity of promoters upstream of known genes, and present evidence that non-coding transcription plays an evolutionarily conserved role in genome partitioning by de-coupling co-expression of adjacent genes.

Non-coding transcription also plays a direct role in regulation of gene expression. Many strong antisense transcripts are present in vegetative cells, and disproportionately represent genes exclusively expressed during mid-meiosis. I have demonstrated that antisense transcription represses expression of these genes, whose ectopic expression is often toxic to vegetative cells. Repression by antisense transcriptional interference supersedes previously proposed mechanisms for several genes, and appears to be a widespread method of regulating transcription.

For my family, without whom none of this would have been possible.

Contents

Nomenclature

Ago	Argonaute
AMV	Avian myeloblastoma virus
APE 1	human apurinic/apyrimidinic endonuclease
CAGE	Cap Analysis of Gene Expression
cdc	Cell division cycle
CDK	Cyclin dependent kinase
cDNAs	complementary DNA
ChIP	Chromatin immunoprecipitation
CUT	Cryptic unstable transcript
Cy	Cyanine dye
DHFR	Dihydrofolate reductase
DMSO	dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSR	determinant of selective removal
DTT	dithiothreitol
dUTP	deoxyuridine triphosphate
EMM	Edinburgh minimal medium
EtOH	ethanol
FHA	forkhead associated
FKH	Forkhead
HSE	Heat shock element
kb	Kilobase
KS	Kolmogorov-Smirnov
MMLV	Mahoney murine lukemia virus

NHS	N-hydroxysuccinimidyl
nt	Nucleotide
OD	Optical density
ORF	open reading frame
PABP	poly(A) binding protein
PCR	Polymerase chain reaction
RC	replicative complex
rDNA	ribosomal DNA
RNA	Ribonucleic acid
RNAi	RNA interference
RPE	R-phycoerytherin
rpm	revolutions per minute
RT	Reverse transcriptase
rtPCR	reverse transcription/polymerase chain reaction
SAGE	Serial Analysis of Gene Expression
SAPE	Streptavidin conjugated R-phycoerytherin
siRNAs	small interfering RNAs
TdT	terminal deoxynucleotidyl transferase
UDG	uracil DNA glycosidase
UTR	untranslated region
YEL	yeast extract liquid
YES	Yeast extract + supplements

List of Figures

List of Tables

1 Introduction

Genomes contain far more information than is used at any given time. Expression must be carefully regulated to ensure that cells can properly respond to external cues, progress through an ordered cell cycle, and to restrict the overlap of mutually incompatible programs. Regulation of transcript levels occurs through several mechanisms, such that steady state RNA concentrations represent the integrated effects of transcription and RNA turnover.

1.1 Yeast as model organisms

Yeast have served man since early civilization. *Saccharomyces cerevisiae*, known as budding, baker's, or brewer's yeast, has been domesticated for more than 9000 years (McGovern et al., 2004; Legras et al., 2007), and was one of the first microscopic organisms visualized by van Leeuwenhoek nearly a third of a millennium ago (Matthews and Lott, 1889). In the modern laboratory, yeast have become indispensable model organisms: they are cultured and manipulated with the ease of bacteria, yet house the same organelles and rely on many of the same biochemical and genetic functions of higher eukaryotes. Largely due to their diverse biology and clinical / industrial relevance, many different yeasts are used in the lab. My work has spanned several yeast genera, including *Saccharomyces* (Lo et al., 2008), *Candida* (Alvarez et al., 2008), and focused primarily on *Schizosaccharomyces* (Oliva et al., 2005; Mickle et al., 2007; Dutta et al., 2008; Patel et al., 2008; Viladevall et al., 2009). Although non-pathogenic and of little Western industrial use, the fission yeast *Schizosaccharomyces pombe* has been employed as a model organism for genetic and cell-cycle studies for more than fifty years . Budding and fission yeast represent vastly different branches of the mycological family tree, have diverged significantly across more than one billion

Organism	Genome Size	Gene #	gene density	Introns/gene
H. sapiens	3.2Gbp	~30,000	100,000nt	~8
S. pombe	14.1Mbp	4970	2,837nt	1.0
S. cerevisiae	12.1Mbp	6607	1,831nt	0.061
E. coli	4.6Mbp	4288	1,072nt	–

Table 1: Genome statistics for various model organisms

years (Heckman et al., 2001) – a distance greater than that between green algae and vascular plants – and provide orthogonal views of conserved eukaryotic biology.

Compared to metazoa, these model yeasts encode the entire repertoire of eukaryotic life into compact genomes of five to six thousand genes (see table 1). Prior to the widespread availability of transgenic mouse production or transient gene knockdown (reviewed in Paddison and Hannon (2002)), basic techniques used by yeast geneticists, such as targeted gene disruption, were inaccessible in higher eukaryotes. Even after the development of reverse genetic tools for complex organisms (Silva et al., 2004), parsimoniously-encoded yeast genomes have provided fertile ground for discovery of gene function; a recent screen for genes involved in mitochondrial biogenesis revealed more than 100 previously unassociated genes, half of which are conserved in mammals (Hess et al., 2009). As a testament to both the scientific importance of budding yeast and the small size of its genome, *S. cerevisiae* was the first eukaryote (and second free-living organism) whose genome was fully sequenced (Goffeau et al., 1996). Completion of the fission yeast genome more than half a decade later (Wood et al., 2002), underscored the differences in genomic organization between these two yeasts who, despite similar total genome size, have significantly different gene densities and transcript structure. The lower gene density of fission yeast is due to longer inter-gene spacing and far more introns than present in budding yeast, and we and other recent reports have shown that larger 5' and 3' untranslated regions (UTRs) of genes comprise roughly 10% of the genome (Wilhelm et al., 2008; Dutrow et al., 2008).

In addition to differences in genome organization, each yeast uniquely shares biochemical or organellar function with higher eukaryotes. Heterochromatic silencing of the telomeres and rDNA occurs in both yeasts, but is mediated by different mechanisms. RNA interference (RNAi), a mechanism conserved from fission yeast through man, guides heterochromatin formation in *S. pombe*, while budding yeast, which lacks all components of the RNAi machinery, uses a unique yeast-specific mechanism (Huang, 2002). Centromeres in budding yeast are extremely small, on the order of 200nt (Meluh and Koshland, 1997), while fission yeast centromeres are long, 35-110kb, heterochromatic sequences (reviewed in Ekwall (2004)) which more closely resemble those found in metazoa. Despite these similarities, *S. pombe* is not "closer" to mammals than *S. cerevisiae*, and several key biochemical pathways or even entire organelles are missing in *S. pombe*. For example, budding yeast have well-developed peroxisomes, mutations in which mimic dysfunction seen in human neurological disease (Lazarow, 1995). Long considered absent or vestigial, peroxisomes have only recently been identified conclusively in fission yeast (Jourdain et al., 2008). Despite these and numerous other idiosyncrasies, each yeast has served as an excellent model system and continues to provide a complementary set of tools for various aspects of molecular and cellular biology.

A compelling use for these "simple" model eukaryotes is in the reductive study of human disease. Deregulation of the cell cycle and bypass of checkpoint control is a key event in neoplasia (reviewed in Hanahan and Weinberg (2000)), and many of the key cell-cycle regulators and checkpoint mediators in man have orthologs in yeast. An elegant example of the direct applicability of cross-species analysis comes from early work by Nurse on Cdc2p, in which human cdc2 was identified by complementation of defective *S. pombe* cdc2 (Lee and Nurse, 1987). Later work demonstrated the inhibitory role of *S. pombe* Cdc2p Tyr15 phosphorylation on cell-cycle progression (Gould and Nurse, 1989), findings which were later mirrored (down to the exact phospho-residue!) in human cell lines (Jin et al., 1996). It is perhaps not surprising that such a fundamental process is conserved across species; it is the degree of conservation, 65% identity and nearly 80% similar matches, which is striking. Despite their compact genomes housing one fifth the genes in barely 1/2500th the total nucleotide space of mammals, yeast encode many proteins relevant to a wide range of clinical conditions. Studies in yeast have provided data relevant to a number of diseases and associated processes, including control of the cell cycle (Nurse, 2002), cancer and DNA damage (Workman et al., 2006), apoptosis (Madeo et al., 2004), lipid homeostasis (Hughes et al., 2005), regulation of drug metabolism (Hughes et al., 2007), diabetes (Hong et al., 2003), mitochondrial dysfunction (Perocchi et al., 2008; Barrientos, 2003), retroviral infection (Irwin et al., 2005), and prion diseases (Collinge and Clarke, 2007). Perhaps most striking, given the complexity of the human brain, is the number of neurodegenerative disorders with orthologous mutations in yeast, including Parkinson's, Alzheimer's, Huntington's (Outeiro and Muchowski, 2004), and Batten's disease (Vitiello et al., 2007), and the large number of neuronal proteins with strong yeast orthologs (Wood et al., 2002). Beyond elucidating basic cellular processes, yeast have provided insight into complex human conditions.

In addition to their utility as models of human disease, yeast continue to provide an excellent system in which to study basic genetic and biochemical processes. Yeast can be grown en masse using low cost media and equipment, and rapidly reach high culture densities. In addition, yeast cultures can be made highly synchronous such that large populations of cells corresponding to a single phase of the cell cycle can be analyzed. Many methods of synchronizing yeast cultures have been established (reviewed in Futcher (1999)), and include physical (eg. elutriation and gradient centrifugation), genetic (eg. temperature sensitive cell-cycle mutants such as cdc25-22 (*pombe*) and CDC28-13 (*cerevisiae*)), and chemical (eg. nocadazole or alpha factor arrest) techniques. Mitchison's measurement of macromolecular synthesis in *S. pombe* nearly 50 years ago demonstrated cell-cycle linked stepwise synthesis of RNA, DNA, and protein by metabolic labeling; technological advances over the past half century have permitted an "extension" of this elegant experiment is the analysis of individual transcript abundance throughout the cell cycle as presented in this thesis on page 41, below.

1.2 The Cell Cycle

Cell division is the only path to immortality –Schleiden and Schwann (1838).

Coordination of the growth, DNA replication, and division of individual cells is required for the perpetuation of free living organisms. Regulation of these events, together comprising the cell cycle, is central to life. The universal nature of this process is exemplified by the studies of sea urchins, fission, and budding yeast which shared the 2001 Nobel prize in medicine (Hartwell, 2004; Hunt, 2002; Nurse, 2002) due, in part, to their widespread applicability to human life and disease. While the exquisite mechanisms of proper cell cycle control went unnoticed for thousands of years of scientific study, the panoply of human morbidity caused by aberrant cell cycle progression has undoubtedly been witnessed since the dawn of man; dysregulation of the cell cycle is central to many illnesses including cancer.

Regulation of the cell cycle in higher eukaryotes results from the association of different CDKs with certain cyclins at various phases of the cell cycle. Fission yeast, in contrast, have a single cell-cycle driving CDK, Cdc2, which associates with different cyclins throughout the cell cycle; levels of Cdc2 remain relatively constant throughout the cell cycle, shifting regulation of Cdc2 activity to its associated cyclins. During G1, Cdc13-CDK activity is kept at bay by the inhibitory activity of Rum1, whose expression begins at mitotic anaphase. At the G1/S transition, Rum1 is itself phosphorylated and subsequently degraded, leaving the now uninhibited Cig2-CDK free to induce S-phase entry (Benito et al., 1998). At replication origins, the heterohexameric ORC recruits, among others, Cdc18 and Cdt1, leading to MCM2-7 association and formation of the pre-RC and origin licensing. Fewer than 50% of these licensed origins actually fire in a given S-phase (Mickle et al., 2007), and do so only after activation by the Hsk1-Dfp1 kinase(Patel et al., 2008). Re-licensing is precluded by the Cig2-CDK phosphorylation mediated proteolysis of Cdc18, whose transcription is downregulated in S-phase; Cdc18's MCM loading activity is restricted small window in mitosis and G1. Mitosis is initiated by a spike in Cdc13-CDK activity. Cdc13-Cdc2 is present throughout G2, but activity is kept low by inhibitory phosphorylation of Cdc2 on Tyr15 by Wee1 and Mik1. Inactivation of Wee1/Mik1 kinases and dephosphorylation of Cdc2 by Cdc25 permits a rapid increase in Cdc13-CDK activity; expression of Cdc25p is highly sensitive to protein biosynthetic capacity, and critical size may be sensed through increased eIF4A activity and the subsequent spike in Cdc25 levels (Daga and Jimenez, 1999). During mitosis, Rum1 begins to accumulate, inhibiting Cdc2-Cdc13 activity, and the cycle continues.

Periodic transcription and coherent gene loops

Periodic expression of cell-cycle relevant genes in a once-per-cycle fashion was first recognized for yeast histone mRNAs. Control of the cell cycle is afforded by multiple mechanisms. Regulated transcription, post-translational protein modification, and protein degradation are all involved in unidirectionally driving the cell cycle. In budding yeast, at least one transcription factor in each cell-cycle regulated complex oscillates with a once-per-cycle period (Pramila et al., 2002). This factor promotes expression of the next "wave" of transcription, generating forward motion through the cell cycle. In addition, a given transcriptional wave often generates factors which repress the preceding program, enforcing a ratchet-like forward progression . In budding yeast, the G_1 transcription factors SBF and MBF activate transcription of *NDD1*. Ndd1p, in turn, activates the Fkh2p-Mcm1p transcription factor complex, which leads to G2/M transcription of *ACE2* and *SWI5*. Mcm1p, Ace2p, and Swi5p direct transcription of *SWI4*, the DNA binding component of SBF, and activate transcription of *CLN3*, the result of both being increased SBF/MBF activity and completion of the cycle. Regulation of transcription factor activity relies on far more than activation of factor transcription (1). Ace2p, for example, is differentially localized to the daughter cell, where it acts on an overlapping, but not identical, set of genes as the mother-specific Swi5p. Both Ace2p and Swi5p activity is regulated by selective nuclear localization; phosphorylation by CDK masks the nuclear localization signal and results in cytoplasmic sequestration of the factors, precluding their DNA binding functions (Moll et al., 1991; O'Conallain et al., 1999).

Similar feed-forward transcriptional loops are present in fission yeast, and are detailed below (see manuscript on page 41).

Meiosis is an alternate cell "cycle"

In all sexual species, an alternate cell cycle exists in which one round of DNA synthesis is followed by two successive divisions of the cells' genetic complement. This process is fundamentally conserved across eukaryota, but has evolved to serve quite different functions in different organisms. In higher eukaryotes, meiosis is initiated within germ cells by extrinsic cues from the surrounding cell layers; the resulting gametes

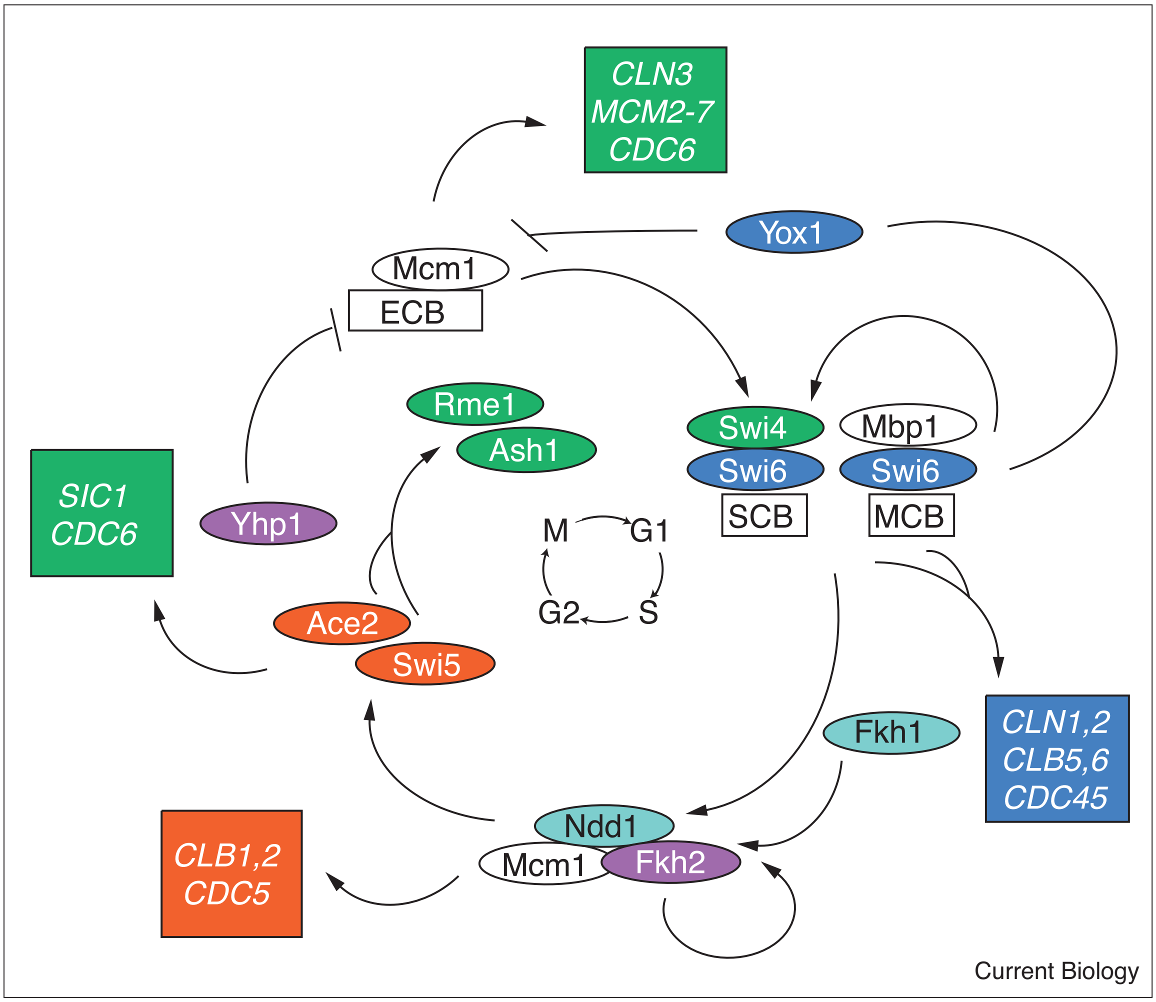


Figure 1: Serial feedback drives cell-cycle transcription in budding yeast. Tandem transcriptional waves generate a cycling behavior. Transcription factors driving a given step in the cell cycle are expressed by the previous wave, such as the SBF/MBF regulated expression of Fkh2. Expression of a given factor is often autoregulatory, both positively, such as the SBF-driven expression of Swi4, and negatively, as exists for the MBF-dependent repression of *MCM1* through Yox1p. Figure modified from Breeden (2003).

are often the sole means of reproduction for an organism. In both budding and fission yeast, meiosis occurs as a response to nutrient starvation (and quorum) signals, and generates environmentally resistant spores that are capable of weathering conditions unsurvivable by the vegetative cell. In both upper and lower eukaryotes, entry into the meiotic program is carefully controlled by multiple regulatory layers. Given the critical importance of proper meiosis on the agricultural and medical fields, the study of conserved mechanisms underlying meiotic regulation and fidelity are excellent uses for tractable model yeasts.

The fission yeast life-cycle is effectively triphasic. When grown in nutrient rich conditions capable of supporting proliferation, cells (generally haploid in the case of *S. pombe*) clonally divide through mitosis. In nutrient limiting conditions, mitotic cycling halts, and cells enter a stationary phase at either G_2 (for carbon starvation; in this state, post-S cells have a second copy of the genome from which to enact DNA repair as needed (Costello et al., 1986)) or accumulate at G_1 (for nitrogen starved cells). When nitrogen-starved cells of both mating types are present in mixed culture (including homothallic h^{90} strains), conjugation and mating occur followed immediately by meiosis which generates four haploid spores. Diploids will only cycle mitotically if moved from conjugation-promoting to rich media; when these cells experience nutrient depletion, G_1 arrest occurs, followed by entry into the meiotic program akin to freshly-conjugated haploid pairs.

Entry into and progression through the meiotic cell cycle requires several key mitotic regulators. Cdc2 is required for the transition from G_1 to pre-meiotic S, in cooperation with Cig2, and for the second meiotic division; and Cdc13 and Cdc25 are required for both the first and second meiotic divisions (Iino et al., 1995). During pre-meiotic S, genes involved in mitotic DNA synthesis including *cdc18*, *cdt1*, *cdc22* (detailed below in figure 13 on page 63) are expressed in addition to a number of meiosis-specific genes involved in meiotic recombination and meiotic cohesin loading, eg., *rec8* and *rec11* (detailed analysis of the fission yeast meiotic program is presented in Mata et al. (2002)). Regulation of these genes in both mitotic and meiotic cells is conferred by DSC1 (homologous to budding yeast MBF), consisting of Cdc10, Res1, Res2, and Rep2 in mitosis, and the combination of Cdc10, Res2, and uniquely Rep1 during pre-meiotic S (Sugiyama et al., 1994). Regulation of genes shared between mitotic and meiotic programs can be conferred by cycle-specific regulators, such as the induction of cdc25 by Mei4 in mid-meiosis.

Commitment to meiotic entry must be carefully regulated. Inappropriate entry into the sporulation program is energetically wasteful, and requires the expense of numerous cell-cycles worth of lost competitive time in mixed culture. Conversely, failure to heed the environmental cues which indicate inability to support further mitotic growth could result in the exposure of a relatively fragile vegetative cell to harsh external conditions. As a result, a "razor's-edge" regulatory switch has evolved, and includes overlapping layers of transcriptional regulation and regulation of meiotic message processing, stability, and turnover.

A number of meiotic genes are transcribed in mitotic cells, but contain a sequence element termed the "determinant of selective removal", or DSR which precludes their accumulation in non-meiotic cells. The DSR element is bound by Mmi1, an RNA binding protein containing a YT521-B homology domain, and directs their degradation in an Rrp6-dependent manner (Harigaya et al., 2006). Inactivation of Mmi1 in vegetative cells leads to aberrant expression of meiotic genes, including targets of Mei4, a meiosis-specific FHA domain containing transcription factor. Induction of meiotic genes following disruption of only Mmi1 strongly suggests that negative regulation conferred by the DSR/Mmi1/Rrp6 system is constantly required to keep expression of meiotic genes at bay, and does not serve as a dispensable "safety net" for ectopic expression of meiotic genes.

As Mmi1 acts in a pathway which degrades DSR-containing meiotic messages, downregulation of its function is required for entry into and progression through meiosis. In vegetative cells, Mmi1 is localized throughout the nucleus. During meiotic prophase, Mmi1 signal converges at a single point in the nucleus, colocalized with the Mei2 dot, in a *mei2* and meiRNA dependent manner. A model of Mmi1/Mei2 interaction from the Yamamoto lab proposes that during meiosis, a Mei2p/meiRNA complex sequesters Mmi1 to a discrete focus from which it is unable to direct degradation of DSR-containing transcripts. Mei2 expression in vegetative cells is regulated by the Pat1 Ser/Thr kinase, both by direct phosphorylation followed by ubiquitin mediated proteolysis and indirect repression of transcription by phosphorylation of the Ste11 transcription factor and its subsequent sequestration by 14-3-3 protein Rad24 (Kitamura et al., 2001).

Pat1 kinase activity negatively regulates meiotic entry, and can be perturbed to induce meiosis in otherwise unfavorable conditions. Presence of both *mat*-M and *mat*-P is required for meiotic entry, as is normally achieved by conjugation of h^- and h^+ cells. Heterothallic strains will arrest in G_1 when nitrogen starved, but can be pushed into meiosis by inactivation Pat1, as is accomplished by shifting cultures harboring homozygous copies of the temperature sensitive pat1-114 allele to restrictive temperature. The resulting meiosis is relatively synchronous, as it allows all cells to be arrested at G_1 prior to accumulation of Mei2 and Ste11 function. This synchronous meiotic induction has been widely used in the community, including my own work below.

In addition to targeted degradation of meiotic transcripts in vegetative cells, previous work in the lab of Janet Leatherwood demonstrated that regulated splicing provides an additional mechanism to preclude translation of meiotic gene products during mitosis. The splicing status of many early and mid-meiotic genes was examined, and it was demonstrated that many (extrapolated to 10%) of these meiosis-specific genes were transcribed but not spliced in vegetative cells. At various times during a synchronous meiosis, spliced transcripts were formed, often concomitant with upregulation of expression. Among the genes identified as differentially spliced were the meiosis-specific cyclins rem1 and crs1 (the latter named as "Cyclin Regulated by Splicing") (Averbeck et al., 2005). Of note, I believe that much of what was considered "regulated splicing" is a technical artifact of the RT-PCR analysis used; the majority of genes for whom regulated splicing was proposed are transcribed on the antisense strand in vegetative cells. These antisense transcripts inherently lack splicing signals, but are detected by PCR. Although regulated splicing does appear to be a valid regulatory mode for several genes (notably including crs1), my work (detailed below in section 4) strongly suggests that the interference of antisense transcription engenders regulatory function for these early genes. Regulated splicing appears to be used for only a minority of early meiotic genes.

1.3 Eukaryotic transcriptome complexity

In their seminal work characterizing the *lac* operon of E. coli, Jacob and Monod described a neatly organized paradigm for genetic regulation in which a discrete coding region is flanked by separable promoter and termination sequences (Jacob and Monod, 1961). This linear model of gene regulation has been supported by nearly a half century of genetic and biochemical studies, and conceptually extended to envision a genome consisting of tandem discrete transcription units. This model accounts for the output of a single transcript from any given locus, the regulation of transcription exclusively by upstream untranscribed control elements, and regulation

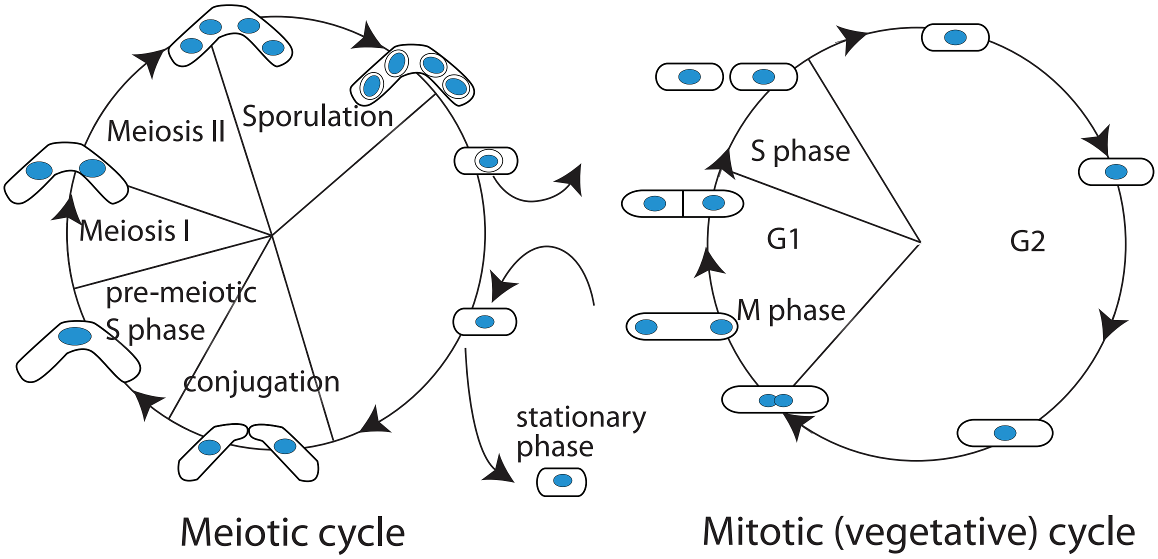


Figure 2: The mitotic and meiotic cell cycles of fission yeast.
Although considerable re-use of individual genes occurs, vegetative growth and meiosis are the result of separate transcriptional programs. Entry into the meiotic cycle normally occurs following the conjugation of nitrogen starved, G_1 arrested cells of opposite mating types. These diploid cells undergo a single round of pre-meiotic DNA synthesis coupled with extensive crossover formation and placement of meiotic cohesins, followed by two rounds of nuclear division without intervening S phases. Finally, the late meiotic program directs packaging of these haploid nuclei into environmentally resistant spores. Upon return to a favorable environment, germination occurs followed by re-enter into the (haploid) mitotic program.

of termination and transcript processing by downstream flanking sequences. With this model in mind, many useful tools have been developed and fruitful studies executed. Expression microarrays, detailed below on page 27, are designed provide a genome-wide snapshot of known transcripts, where a short (often strand-nonspecific) probe serves as a proxy for a single larger transcript. Studies of protein-DNA interactions, including chromatin immunoprecipitation (ChIP) on chip and bioinformatic sequence motif searches, have largely focused on the transcription start site proximal sequences upstream of interrogated transcripts. Probes contained on commercial (and many custom-spotted) arrays designed for ChIP/chip exclude transcribed sequences entirely, reflecting an presumed lack of factor-binding function by sequences outside of this narrowly defined window. Despite the successes of these reductionist approaches, it has been well established that transcription occurs from far more than these coding elements, and that regulation occurs by a wide variety of mechanisms not limited to protein-DNA interactions at upstream sequence elements.

Eukaryotic genomes are extensively transcribed

Following the discretized model of genome organization presented by Jacob and Monod, transcribed material should correspond to functional elements, generally discrete protein coding and structural RNAs. Subsequent work has demonstrated that this overly simple model is incomplete. Studies of sea urchin gastrulae performed during the 1970s found that total nuclear RNA pools were greater than ten times more complex than the RNAs present on polysomes, and corresponded to nearly 30% of the nonrepetitive genome (Hough et al., 1975). Although the identity of these RNAs was unknown, far more material was transcribed than could be accounted for by translated sequences. Twenty years later, the Vogelstein lab identified robustly transcribed polyadenlyated RNAs lacking discernable open reading frames as part of the transcription complement of human cells (Velculescu et al., 1995). This finding, made during the development of serial analysis of gene expression (SAGE), was the harbinger of future methods for unbiased transcript detection and examination of unannotated RNAs.

Though Cot analysis and small-scale transcript sequencing had demonstrated that nuclear RNA pools contained far more than coding sequence, the extent of non-coding transcription was not fully appreciated until the advent of high resolution studies using large scale sequencing and tiling arrays. "Genome wide" studies of transcription in human cell lines were first performed along the relatively accessible 95 million base pairs comprising chromosomes 21 and 22. These pilot experiments revealed that more than 90% the bases present in stable transcripts were extra-exonic and non-coding. In addition, more than 26% of the genomic regions analyzed were transcribed (Kapranov et al., 2002), consistent with previous bulk hybridization studies. A plausible explanation for this (and previously estimated) widespread transcription was "sampling" of the genome by RNA Pol II, which would generate diverse and diffuse transcript species from permissive chromatin across the genome. Further examination along these exemplar human chromosomes revealed tens of thousands of binding events for several transcription factors, including Sp1, cMyc, and p53, and these ChIP on chip defined binding events were strongly correlated with the transcription start sites of the previously identified non-coding RNAs. These findings strongly suggested that instead of sampling, directed transcription was responsible for generation of many extra-exonic transcripts (Cawley et al., 2004).

Greater transcriptome complexity results from the presence of relatively unprocessed RNAs. In addition to the previously studied polyA+ fraction, much of the genome is transcribed into unadenylated or uncapped forms. Two thirds of all heterogeneous large nuclear RNAs are capped but unadenylated, and the majority of capped RNAs are absent from polysomes (Salditt-Georgieff et al., 1981). Additional complexity results from the different populations present in the cytoplasm and those confined to the nucleus. Subsequent tiling array studies addressed these complexities by profiling different "compartments" of transcripts, based on subcellular fraction (cytoplasmic vs nuclear localization) and polyadenylation status (Cheng et al., 2005). Technological advances permitted the analysis of 30% of the human genome, a far larger fraction than was previously possible, and enabled the comparison of transcripts derived from chromosomes with varying gene density. A striking feature of these studies was the relatively fixed transcript density of different chromosomes. On gene-rich chromosomes, such as Chr19, nearly half of the transcripts identified mapped to previous annotations, while only 13% were intergenic. The inverse was noted on gene bereft Chr13, where nearly 45% of transcripts were previously unidentified intergenic species and only 17% corresponded to previously identified or predicted RNAs. Unannotated "junk DNA" was found to be actively transcribed and processed largely without a clear functional role, demonstrating that at the very least, this genomic "dark-matter" contributes to cellular transcript pools.

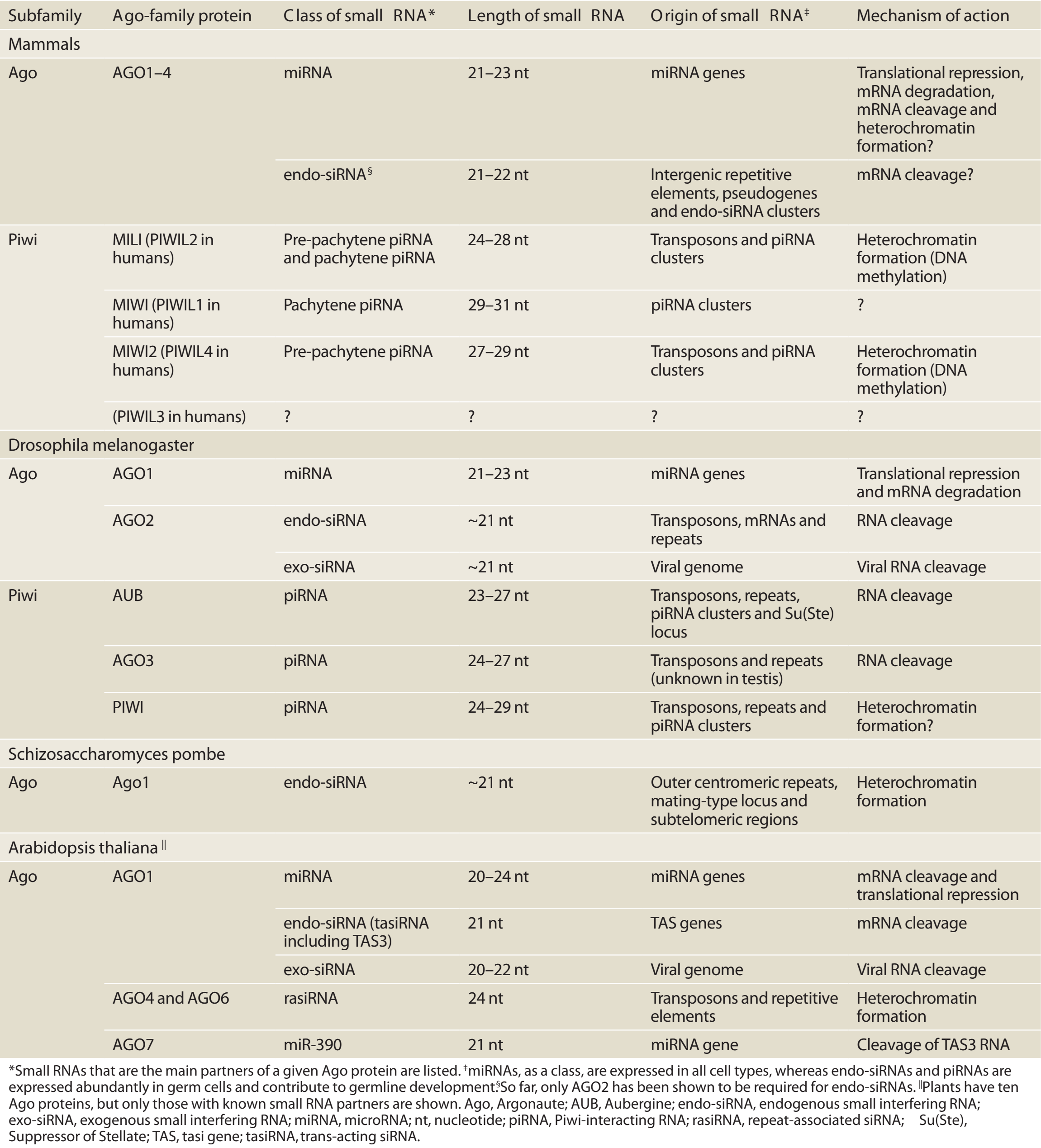
Hybridization based techniques are inherently limited to measuring the short stretch of RNA corresponding to each probe; the overall structure of associated transcripts remains unknown. A possible explanation for the observed widespread transcription was the presence of many overlapping relatively short transcripts, or transcripts with unconserved end positions. Large scale sequencing based approaches were subsequently used to provide both quantitation and detailed transcript structure analysis, exemplified by the massive collaborative effort between the RIKEN and FANTOM consortia to sequence the mouse transcriptome completed in 2005 (Carninci et al., 2005). Seven million 5' end tags (via Cap Analysis of Gene Expression, CAGE (Shiraki et al., 2003)), more than one million paired end tags (Ng et al., 2005), and roughly 100,000 full length mouse cDNAs were generated in an effort to characterize the breadth of mammalian transcription. From these data, nearly 200,000 unique transcripts were identified, reflecting a genome complexity at least an order of magnitude above the previously estimated 20,000 genes (Mouse Genome Sequencing Consortium, 2002). While many of the newly-identified transcripts were splice variants, transcripts with alternate 5' and 3' untranslated regions, or previously unannotated genes with protein coding potential, a vast number of non-coding transcripts were identified, accounting for more than 33% of all sequences. Only a fraction of these novel non-coding transcripts were identified as singletons, suggesting that many of these species are the output of directed transcription and processing and not the result of stochastic sampling.

Functional non-coding transcripts

Non-coding species represent the vast majority of transcriptional output from the human genome, recently estimated at 98% of all RNAs (Mattick, 2005). Over the last decade, a renaissance of RNA biology has occurred, resulting in the characterization of many new classes of functional RNAs. Studies of small RNAs have been especially fruitful, where a variety of mechanisms have been shown to regulate gene expression and chromatin structure through a wide range of processed species. Regulatory transcripts had been proposed by Jacob and Monod, who envisioned the *lac* repressor was, in fact, RNA; although LacI turned out to be protein, regulatory RNAs are indeed present in eubacteria. In addition, a richer palette of non-coding RNA regulatory functions has been described in eukaryotes; since the Ruvkin and Ambros labs discovery of the first functional small RNA in *C. elegans* in 1993 (Lee et al., 1993; Wightman et al., 1993), many classes of short functional species have been identified, and advances in high-throughput short read sequencing have made possible the exhaustive identification of variously sized populations from a number of organisms. The breadth of small RNA biology is beyond the scope of this document, and the brief overview that follows is intended to highlight their diversity of form, formation, and function.

The ability of non-coding RNAs to affect gene expression was demonstrated more than twenty years ago. Here at Stony Brook, Inouye and colleagues identified a short transcript in *E. coli* which had strong homology to the reverse complement of OmpC, a well-regulated component of the outer membrane (Mizuno et al., 1984). This “mRNA-interfering complementary RNA” was shown to decrease translation and overall stability of the sense transcript, presumably through formation of an RNA:RNA hybrid. Competing work performed in Nancy Kleckner’s lab demonstrated that activity was conferred in trans, and that translation of the Tn10 transposon’s active insert was suppressed by the presence of small RNA complementary to the 5’ of the transposase transcript (Simons and Kleckner, 1983). In both cases, binding of the regulatory RNA occurred near the 5’ end of the regulated transcript. While presence and binding of this complementary transcript reduces translation, an extant question is whether this occurs through direct occlusion of ribosome engagement, direction of transcript turnover, or a combination of the two (addressed in Gottesman (2005)).

Similar in spirit to eubacterial “interfering complementary” transcripts, microRNAs are short single stranded RNAs which function in direct modulation of translation and regulation of target turnover through direct target cleavage, deadenylation, and translational repression. These 21-23nt species result from the processing of highly structured pol II/ pol III transcripts by the RNAse III Drosha and its partner Pasha/DGCR8(Denli et al., 2004), are shuttled to the cytoplasm by exportin-5, and finally cleaved by Dicer into their mature from. Processed microRNAs are strand-specifically loaded into Argonaute proteins where they function independent of Ago’s



Subfamily	Ago-family protein	Class of small RNA*	Length of small RNA	Origin of small RNA‡	Mechanism of action
Mammals					
Ago	AGO1–4	miRNA	21–23 nt	miRNA genes	Translational repression, mRNA degradation, mRNA cleavage and heterochromatin formation?
		endo-siRNA§	21–22 nt	Intergenic repetitive elements, pseudogenes and endo-siRNA clusters	mRNA cleavage?
Piwi	MILI (PIWIL2 in humans)	Pre-pachytene piRNA and pachytene piRNA	24–28 nt	Transposons and piRNA clusters	Heterochromatin formation (DNA methylation)
	MIWI (PIWIL1 in humans)	Pachytene piRNA	29–31 nt	piRNA clusters	?
	MIWI2 (PIWIL4 in humans)	Pre-pachytene piRNA	27–29 nt	Transposons and piRNA clusters	Heterochromatin formation (DNA methylation)
	(PIWIL3 in humans)	?	?	?	?
Drosophila melanogaster					
Ago	AGO1	miRNA	21–23 nt	miRNA genes	Translational repression and mRNA degradation
	AGO2	endo-siRNA	~21 nt	Transposons, mRNAs and repeats	RNA cleavage
		exo-siRNA	~21 nt	Viral genome	Viral RNA cleavage
Piwi	AUB	piRNA	23–27 nt	Transposons, repeats, piRNA clusters and Su(Ste) locus	RNA cleavage
	AGO3	piRNA	24–27 nt	Transposons and repeats (unknown in testis)	RNA cleavage
	PIWI	piRNA	24–29 nt	Transposons, repeats and piRNA clusters	Heterochromatin formation?
Schizosaccharomyces pombe					
Ago	Ago1	endo-siRNA	~21 nt	Outer centromeric repeats, mating-type locus and subtelomeric regions	Heterochromatin formation
Arabidopsis thaliana‖					
Ago	AGO1	miRNA	20–24 nt	miRNA genes	mRNA cleavage and translational repression
		endo-siRNA (tasiRNA including TAS3)	21 nt	TAS genes	mRNA cleavage
		exo-siRNA	20–22 nt	Viral genome	Viral RNA cleavage
	AGO4 and AGO6	rasiRNA	24 nt	Transposons and repetitive elements	Heterochromatin formation
	AGO7	miR-390	21 nt	miRNA gene	Cleavage of TAS3 RNA

*Small RNAs that are the main partners of a given Ago protein are listed. ‡miRNAs, as a class, are expressed in all cell types, whereas endo-siRNAs and piRNAs are expressed abundantly in germ cells and contribute to germline development.§So far, only AGO2 has been shown to be required for endo-siRNAs. ‖Plants have ten Ago proteins, but only those with known small RNA partners are shown. Ago, Argonaute; AUB, Aubergine; endo-siRNA, endogenous small interfering RNA; exo-siRNA, exogenous small interfering RNA; miRNA, microRNA; nt, nucleotide; piRNA, Piwi-interacting RNA; rasiRNA, repeat-associated siRNA; Su(Ste), Suppressor of Stellate; TAS, tasi gene; tasiRNA, trans-acting siRNA.

Table 2: Classes of functional RNAs. Reproduced from Kim et al. (2009).

slicer and endonuclease activities (reviewed in Winter et al. (2009)) through mechanisms incompletely understood. This function of argonaute appears to be restricted to higher eukaryotes; to date, there is no evidence that microRNAs exist in either fission or budding yeast.

Small interfering RNAs (siRNAs) are found across multicellular plants and animals, and also in fission yeast. These short (usually 21nt) RNAs are generated from endogenous transcripts (outer centromeric and subtelomeric repeat regions in *S. pombe*, transposons and other repetitive sequences in mammals and flies) and function in either direct mRNA cleavage or formation of heterochromatin through Argonaute (Ago) family proteins. Ago loading requires a double-stranded RNA template. This can be accomplished by endonucleolytic cleavage of a double-stranded RNA template, as occurs during processing of pre-miRNAs or dsRNA viral genome intermediates, bidirectional transcription of complementary sequences, or following synthesis by an RNA dependent RNA polymerase (RdRP). The entire RNAi pathway is present in fission yeast, and one copy each of Dicer (Dcr1), Argonaute (Ago1), and the RdRP (Rdp1) is encoded in the *S. pombe* genome. They are required for proper establishment and maintenance of heterochromatin, but are apparently uninvolved in regulation of euchromatic genes transcribed from both strands (Hansen et al., 2005).

Transcription interference

In addition to the regulatory RNAs discussed above, many classes of proteins affect transcription, ranging from DNA-binding transcription factors and components of the basal transcription apparatus to silencing and RNA processing machinery. As has been previously described in context of the cell cycle, elegant regulatory mechanisms have evolved which increase or reduce transcription of a number of targets, often including the factor itself in regulatory feedback. In this paradigm, activity is conferred by the encoded protein, which in turn acts on the transcription machinery or chromatin to modulate transcription. Individual genes are not transcribed *in vacuo*, however, and both genomic and transcriptional context play important regulatory roles.

The very act of transcribing an adjacent or overlapping/abutting sequence has been shown to affect both transcript abundance, identity, and processing. Depending on the regulatory context, this influence can be either positive or negative. During the late 1970s to early 1980s, studies of phage λ transcription provided evidence of cis-regulatory interactions between co-localized RNA polymerases. The antagonistic interaction of colliding transcription complexes was described by the Murray lab (Ward and Murray, 1979), wherein mutual impairment of transcription was observed when two promoters were situated "nose-to-nose". The Gottesman lab identified cis-inhibitory activity of upstream transcription passing into or through a downstream promoter element (Adhya and Gottesman, 1982); in this promoter occlusion model, traversal of the "target" promoter by upstream-initiated polymerase precludes binding of polymerase at the target-gene's promoter. This process, since wrapped into the ill-defined term 'transcriptional interference', has been studied in multiple organisms including eubacteria, budding yeast, and *Drosophila*, where there now exist a number of well-characterized examples encompassing numerous mechanisms.

In budding yeast, a large number of non-coding RNAs have recently been mapped (David et al., 2006; Neil et al., 2009), several of which have already been shown to play regulatory roles by transcriptional interference. *SER3*, a 3-phosphoglycerate dehydrogenase which catalyzes an early step in serine and glycine biosynthesis (Albers et al., 2003), is situated downstream of the *SRG1* noncoding RNA. In nitrogen poor media, *SER3* expression and de novo serine biosynthesis occur; in rich media, *SRG1* is transcribed, and represses expression of the then-unnecessary *SER3* gene. At this locus, Winston and colleagues demonstrated a reduction of transcription factor binding in regions traversed by the upstream-initiated transcript, implying that "promoter occluding" transcription interference may operate at the level of reduced transcription factor recruitment (Martens et al., 2004).

Antisense transcription interference is also present in budding yeast, and helps regulate the developmental switch between mitotic and meiotic cell cycles. *IME4*, an *I*nitiator of *ME*iosis, is transcribed from the sense strand only in *MATa*/α diploids, where its activity is required for accumulation of *IME1* and entry into meiosis (Shah and Clancy, 1992). In cells lacking the a1-α2 repressive heterodimer, antisense transcripts covering the entire *IME4* gene are generated, antagonizing sense expression (Hongay et al., 2006). This repression occurs only in *cis*-, and depends on the relative strength of the promoters involved; overexpression of the sense transcript can surmount antisense interference.

These mechanisms are not limited to unicellular organisms. Studies in mammals have demonstrated direct antisense transcriptional interference of the sphingosine kinase gene, *sphk1*, where sense and antisense transcription appear to be mutually exclusive at the single-cell level (Imamura et al., 2004). Promoter occlusion also occurs in humans, where DHFR activity, and thereby DNA synthesis, is restricted in quiescent cells by transcription of an upstream noncoding RNA which reduces binding of basal transcription machinery at the major DHFR promoter (Martianov et al., 2007).

Several mechanisms of direct regulation *of* transcription *by* transcription have been recently demonstrated. Interfering transcription, either from an upstream or convergent promoter, can prevent or reduce recruitment of Pol II or other factors required for transcription initiation. In this case, firing and elongation rate of the regulatory RNA run counter to the size and binding affinity of the target gene promoter. In addition to reduced initiation of transcription, transcriptional interference can promote premature transcript termination. Transcription run-on assays performed in the Fink lab demonstrated that, during antisense-favoring conditions (non *MATa*/α), *IME4* sense transcript is initiated but only proceeds to generate a truncated form. Conversely, when sense transcription predominates, *IME4* antisense transcripts initiate but do not not extend to their erstwhile full length (supplemental data of Hongay et al. (2006)). Indirect regulation of target gene transcription can occur by modification of local chromatin structure or DNA topology. In fission yeast, studies of the fission yeast *fbp1+* locus have demonstrated transcriptional *activation* mediated by conversion of the *fbp1* promoter to an open chromatin state by transcription of an upstream-initiated noncoding RNA (Hirota et al., 2008). The passage of RNA pol II through the fbp1 promoter provides clearance for subsequent binding of transcriptional activators and formation of the pre-initiation complex (PIC) at an otherwise inaccessible locus. A similar scheme regulates budding yeast PHO5 expression, where transcription of an unstable non-coding RNA through the PHO5 promoter is necessary and sufficient to evict histones and permit promoter chromatin remodeling (Uhler et al., 2007b).

Despite these clear examples of biologically relevant action, potential mediators of transcriptional interference have only recently been cataloged en masse. Genome wide transcriptome studies, including those presented below in (David et al., 2006; Core et al., 2008; Preker et al., 2008; Seila et al., 2008; He et al., 2008; Wilhelm et al., 2008; Dutrow et al., 2008; Neil et al., 2009; Xu et al., 2009; Affymetrix ENCODE Transcriptome Project and Cold Spring Harbor Laboratory ENCODE Transcriptome Project, 2009), have begun to profile the non-coding RNA repertoire of several organisms. Although many of these transcripts are likely the side-effect of desired genic transcription, the regulatory potential of non-coding RNAs warrants examination of not only the transcripts formed, but also the genomic regions affected by their transcription.

RNA turnover and transcription quality control

Opposing transcription, RNA turnover is required to ensure that relevant, properly formed transcripts are present within the cell. Transcript quality is constantly assessed, beginning at the act of transcription itself and continuing alongside translation. RNA turnover occurs primarily through exonuclease activity. Processed mRNA transcripts harbor both a 5'-methylguanosine cap and protein-bound 3'-poly(A) tail. As cellular exonucleases usually act on accessible single-stranded terminii, these modifications confer resistance to exonuclease activity; it is only after decapping or deadenylation that degradation can occur. Conversely, transcripts which fail to acquire a 5' cap or are released as unadenylated polymers are rapidly degraded by 5'→3' and 3'→5' exonucleases, respectively. Although involving different ends, these pathways appear to have largely overlapping gross activities. Studies in budding yeast have demonstrated that disruption of either activity results in the upregulation of only a small number of transcripts, indicating that these pathways are functionally redundant (He et al., 2003).

Turnover of unprotected 5' terminii is generally constitutive, and regulation of exonucleolytic activity is conferred by selective decapping. In *S. pombe* and *S. cerevisiae*, decapping is performed by a dimeric enzyme consisting of Dcp1 and Dcp2, followed by cleavage through Exo2/Xrn1p, respectively. Decapping efficiency is enhanced by several additional factors, including several associated with 3' end binding such as Lsm1, the enhancer of decapping Edc3 (Kshirsagar and Parker, 2004), and in budding yeast, the PABP binding protein Pbp1. Components of the 5'→3' turnover pathway are found in P-bodies, cytoplasmic foci involved in various facets of RNA storage and metabolism, although 5'→3' exonuclease activity appears to be present throughout the cytoplasm (Sheth and Parker, 2003).

Transcription of much of the genome occurs, but only a fraction of these RNAs are stable and exported to the cytoplasm; the remaining material is hydrolyzed by one of several mechanisms. This turnover is highly regulated, and provides an additional layer controlling the cellular transcript pool. The exosome, a large multisubunit 3' exonuclease, is a primary factor in post-transcriptional regulation of RNAs, and guides both maturation and destruction of various species. In many respects, the exosome's activity on transcript pools parallels that of the proteasome on intracellular polypeptides. Both complexes are large macromolecular cages (although the eukaryotic exosome core appears to be catalytically inactive (Liu et al., 2006)), and associated "gatekeeper" subunits regulate access to the complex. Substrate recognition is engendered by yet additional subunits, and specificity often relies on tagging of polymers bound for degradation: poly(A) addition for RNAs, and polyubiquitination for proteins.

The exosome was identified in budding yeast for its essential role in maturation of the 5.8S rRNA, and was determined to be a multiprotein complex consisting of several predicted or previously confirmed exoribonucleases, including homologs of bacterial RNase II and RNase H. Strong cross-species conservation of most subunits was observed, including the ability of hRrp4 to complement mutation in *RRP4* (Mitchell et al., 1997). Subsequent studies have demonstrated diverse functions for the exosome, which is highly conserved from archaea to metazoa. The catalytically inactive eukaryotic exosome core appears to function as a scaffold and is associated with the exoribonucleases Rrp6 and Rrp44/Dis3, though the former is restricted to the nucleus in yeast.

In budding yeast, non-coding transcripts are terminated by Nrd1-Nab3, which recognize $GUA\frac{A}{G}$ and $UCUU$, respectively, weak motifs enriched in non-coding regions. Once terminated, these transcripts are subject to "maturation" by Rrp6 activity which is directed to these transcripts through polyadenylation by the Trf4/Trf5 non-canonical poly(A) polymerases present in the TRAMP complex(Thiebaut et al., 2006). Orthologs of all members of the TRAMP complex are found in fission yeast, including the poly(A) polymerase *cid14*, the RNA binding *air1*, and a pair of *mtr4* helicases; a *nab3* homolog is also present. Recruitment of Rrp6 to structured non-coding RNAs results in their exonucleolytic "trimming", as the exosome processes the unstructured tail of the transcript. Consequently, spurious transcripts lacking significant structure will be cleaved at specific 3' ends in a Nrd1-Nab3 dependent manner, polyadenylated, and degraded by Rrp6. An additional consequence of this process is that transcripts lacking a canonical cleavage/polyadenylation site will be terminated at a discrete site and gain a poly(A) tail. While the post-cleavage fragment is likely destroyed in an 5'→3' exonuclease dependent fashion, the processed fragment can be recovered by selective inactivation of Rrp6. This technique is used to identify the "hidden" unstable transcripts as described below on page 107.

1.4 Quantitation of expression

The ability to accurately measure RNA levels underpins many of the studies listed above, and nearly all work described herein. The fundamental necessity of transcription detection and quantitation is reflected in the breadth of techniques which have been developed and the subset that are currently in use. Advances have occurred along several axes, including specificity, sensitivity, and parallelism. Total RNA content and synthesis rate can be determined by bulk metabolic or pulse-chase labeling;

experiments in *S. pombe* demonstrated the stepwise synthesis of RNA throughout the cell cycle. Detection and quantitation of individual transcripts was made possible by two different methods developed independently in the late 1970s. S1 nuclease mapping (Berk and Sharp, 1977) and northern blotting (Alwine et al., 1977) provide sequence specific measurements of RNA level and remain in use today. While they provide focused quantitation and are well-suited for small numbers of genes, these approaches are laborious and require relatively large amounts of input material per measurement. Moderately high throughput, ranging up to hundreds or thousands of measurements, can be achieved using quantitative rtPCR, where reverse-transcribed cDNAs are amplified using gene specific primers and quantitated using one of a growing panel of methods. Measurement of thousands of genes in even a handful of samples is beyond the reasonable reach of quantitative PCR, which has become the de facto tool for validation of a few individual genes across many samples, and for validation of the high throughput methods, described below.

Spotted Arrays

Two color spotted PCR arrays have been used by our lab for more than a decade, and my involvement has spanned more than six years and multiple publications (Oliva et al., 2005; Mickle et al., 2007; Alvarez et al., 2008; Dutta et al., 2008; Lo et al., 2008; Patel et al., 2008; Viladevall et al., 2009). As such, expression arrays have become a tool which has not only been used to great success in the lab, it is one which has shaped the very design of experiments undertaken. At the heart of expression array experimental design is the comparison between "control" and "unknown" samples; relative ratios between samples are the only reliable output from arrays of this type[1].

[1] Due to the heterogeniety of spotted DNA concentration, amount of DNA deposited/crosslinked to the substrate, and size of the final spots, in-house arrays are inherently ill-suited for absolute quantitation. Advances in photolithographic and direct in-situ synthesis have expanded the role of

As such, two-color arrays are ideally suited for comparisons of linked samples, such as that of wild type to mutant strains, +/- chemical treatment of a given strain, or even time-courses compared to an asynchronous control. In these designs, ones' control is part and parcel of the data output; comparisons between various experiments lacking a common control cannot be reliably made.

The term "array" is suitably vague for the tool produced. Although DNA on glass arrays are by far the most common, and the sole format used in this dissertation, a wide variety of materials can be deposited on many different substrate types. Early macroarrays consisted of purified DNAs deposited onto nylon or nitrocellulose; ^{32}P or ^{33}P labeled cDNA probes were hybridized much like a massively parallel slot (or dot) blot. Development of relatively low cost, high-resolution stepper-motor driven robotic systems and suitable contact-printing pins permitted a large increase in spot density, and a concomitant decrease in substrate size. While hundreds to thousands of probes could be reliably arrayed onto a 15x15cm membrane, many thousand spots could be printed onto a standard format 25x75mm glass microscope slide. Parallel technologies of in-situ synthesis and direct deposition were developed simultaneously, but are beyond the scope of this document.

I joined the lab before their first successful printing of whole-genome arrays. An array spotting robot had been previously built using the open-source design of Jo DeRisi and Pat Brown at the Cold Spring Harbor Microarray course, but due to recent relocation of the lab and delays in probe production, no usable arrays had as of yet been produced. Probes for these arrays had been previously designed for both budding and fission yeast, and consisted of 3' biased PCR amplicons with a desired size range of ~150nt-1kb. Genomic DNA was used as template, which provided lit-

glass slide arrays to include absolute quantitation in single-color experiments. These platforms are being actively employed for separated tasks in the lab, although their use is beyond the scope of this section.

tle issue for budding yeast with a relative dearth of intronic features. Fission yeast amplicons avoided inclusion of intronic sequence where possible; in reality, many of the probes hybridize to both intronic and exonic sequence. Tandem PCR reactions were performed. A small volume reaction was carried out from genomic DNA template, followed by a larger volume reaction in which raw first-round product was used as template. Although identical primers were used for both reactions, inclusion of a second-round PCR reaction generated large (μg) amounts of DNA with low well-to-well variability; low template complexity and high input copy count combined resulted in reactions which were anectodally run to completion. Following amplification, amplified probe DNA was purified using 96 well ultrafiltration membranes (Millipore, Billerica MA), analyzed by agarose gel electropheresis, and aliquoted into high density plates suitable for array printing.

Many different substrates and attachment chemistries competed for the de facto standard in homemade microarray printing. Although covalent-binding chemistries, eg. acrydite/thiol, exist, most labs settled on surface modifications which provided transient electrostatic interactions with the phosphodiester backbone, followed by covalent crosslinking of DNA to substrate by UVC or vacuum baking. In our hands, aminopropylsilane coated substrates were determined to provide the best combination of stability, binding capacity, spot morphology, and post-hybridization background. Following printing and crosslinking, unoccupied primary amines were blocked by bulk action (fraction V BSA) or direct reduction of the reactive surface using sodium borohydride. When stored under vacuum and secondary dessication, these home-spotted arrays have proven stable over many years.

For relative expression measurements, labeling of reference and “treated” RNAs with two different fluors is required. Direct incorporation of dye-functionalized nucleotides during sample labeling is possible, but can lead to unpredictable biases due to inconsistent incorporation of sterically different dyes (Ideker et al., 2000). In addition, the dyes used are exquisitely sensitive to oxidative damage and photobleaching effects, leading to questionable longevity of fluro-modified nucleic acid samples. To circumvent these issues, we employed an indirect labeling approach in which functionalized cDNAs were generated by reverse transcription including an amino modified dUTP (amino-allyl dUTP), followed by coupling of an N-hydroxysuccinimidyl (NHS) ester functionalized dye. As the same modified nucleotide is used for reverse transcription of RNAs destined to be labeled with Cy3 or Cy5, no differential incorporation during reverse transcription occurs. Additionally, amino-functionalized cDNAs can be stably stored indefinitely, and coupled to unstable dyes only as needed.

Following hybridization, scanning, and determination of fluorescent signal corresponding to each dye, data from two-color arrays are distilled to a simple pair of values per "spot". Following background normalization and intensity-dependent detrending, expression ratios and sums of both channels' signals are calculated from the normalized red and green intensities. For home-spotted arrays, the sum of intensities of of little use (due to relatively inconsistent amounts of spotted material as mentioned in the footnote above), and analysis is focused solely on expression ratio. Prior to the enforcement of MIAME-compliant data deposition policies, results from many published array experiments were a simple genes × arrays matrix filled with relative expression values. In cases of well-designed experiments performed on quality arrays, this single measurement is sufficient to represent differences between two samples. Since a ratio, not absolute signal, is reported, probe-specific effects are minimized, and the reported value accurately reflects relative differences in material corresponding to the array spot examined[2].

[2] This does not imply that a probe's signal necessarily reflects (exclusively) the target for which it was originally designed. Many home-spotted array designs including our own made use of long double-stranded PCR amplicons. While the ratio reported for a given probe reflects relative abun-

Tiling Arrays

In contrast to the "one gene, one probe" design of expression microarrays[3], tiling arrays provide multiple non-redundant unbiased measurements across large swaths of genomic space, such as promoter regions, whole chromosomes, or entire genomes. Design of such arrays requires little *a priori* information about the features of the genome to be interrogated; once sequence is available, arrays can be designed to probe all (non-repetitive) regions without regard to their biological content. This seemingly simplistic design paradigm has several advantages when compared to expression microarrays. Most crucially, the information available from tiling array data does not rely on availability or accuracy of genome annotation. This approach permits discovery of novel genomic features, but dilutes the available number of probes across potentially uninformative sequences. As described below, this limitation is absent in practice when dealing with small genomes, such as that of fission yeast, on "modern" high-density arrays, where advances in manufacturing methods have permitted the inclusion of millions of probes on a single substrate.

Platform description High resolution tiling arrays covering the entire *S. pombe* genome are available commercially (Afffymetrix *S. pombe* tiling 1.0FR), and have been used extensively in the course of my work. These arrays consist of nearly 2.6 million twenty-five nucleotide *in situ* synthesized oligonucleotides. Of these, 1.1 million oligos, or probes, map uniquely to a recent *S. pombe* genome assembly (April 2007) as

dance of hybridizing material in the interrogated samples, the identity of the sequence which hybridizes to a given spot may not be unambiguously defined. This is especially true for gene families and pairs of genes resulting from an ancestral duplication. Further confound results from the inability of double-stranded arrays to distinguish orientation of hybridized material. This is the subject of some discussion, below in 3.3 on page 107.

[3] Many competing or conceptually similar microarray platforms are available. Except where explicitly noted, "two-color (micro)array" is synonymous with our home-spotted PCR-from-genome glass slide arrays.

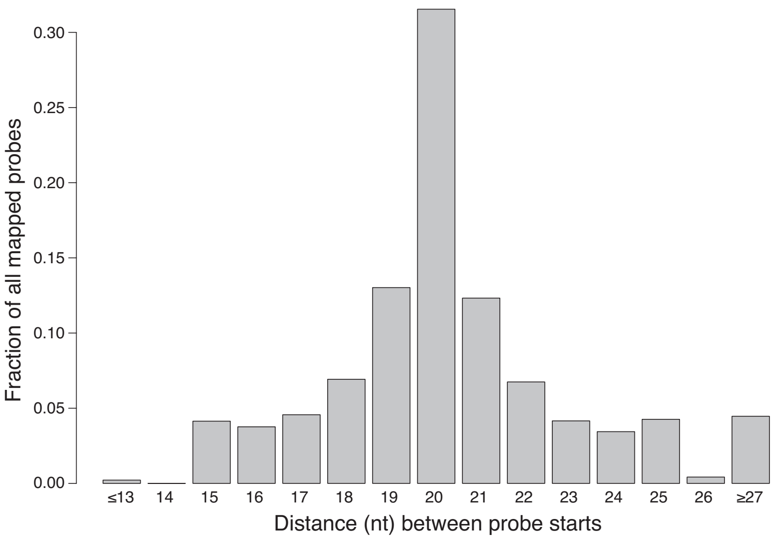


Figure 3: Spacing of mapped probes on Affymetrix *S. pombe* 1.0FR arrays. Probes which uniquely map to a modern reference genome maintain spacing similar to the original design goals. The majority of probes (>50%) have a 19-21 nucleotide start-to-start spacing. Coverage of genomic positions is relative quantized. Very few probes map with spacings of 15<x>25nt. 26nt represents the beginning of the right tale of the distribution; probes with wider spacing represent roughly 5% of total. Longer spacing between mapped probes is generally due to exclusion of repetitive regions.

described below (page 36). In addition to numerous spike-in and hybridization control probes, the vast majority of excluded probes were originally designed as mismatch controls (13th nucleotide is the complement of the perfect match probe). These probes were not used for further analysis, excepting those that aligned to the current genome assembly. Probe coverage was originally designed as one 25mer per 20nt of genomic sequence, generating a 5nt overlap; despite substantial changes to the sequence and assembly of the *S. pombe* reference genome, this figure is still generally (see figure 3).

The Affymetrix platform has proven incredibly flexible in practice. The platform was originally designed for detection of fragments following chromatin immunoprecipitation (ChIP on chip), but has been re-tasked for detection of labeled cDNA, chemically modified RNA, and labeled genomic DNA. Short fragments of nucleic acid (ideally in the range of 60-150nt) are biotinylated through direct chemical or enzymatic labeling, purified, and hybridized in a standard mixture including DMSO, formamide, and citrate salts. This loaded array is hybridized overnight at elevated temperature while constantly mixed to ensure thorough mixing with all potential hybridization partners, and to reduce non-specific adsorption. Following labeled probe recovery[4], arrays are washed under increasingly stringent salt and temperature conditions. Development of hybridized biotinylated material is achieved through a three step staining process. Streptavidin conjugated R-Phycoerytherin (SAPE) attaches an initial fluor to the biotinylated species. Following non-stringent washing, biotinylated goat anti-RPE antibody is used to decorate the previously labeled biotin/SAPE complex. This complex is further fluor-labeled by subsequent staining with SAPE resulting in a geometric amplification of the original biotinylation. Following a final wash, array cartridges are filled with buffer and scanned.

Differences from two-color expression microarrays Tiling arrays are designed to identify and quantitate material present in a single sample. A single nucleic acid preparation is labeled and hybridized, and the intensity of each spot is used to infer the abundance of the corresponding material present in the original sample. In so doing, the burden of a consistent and relevant reference sample is lessened, and direct comparison between any two samples can be made. Direct quantitation on arrays is not without unique computational challenges, however, as significant experiment- and probe-wise normalization must be performed to generate intelligible output. Normal-

[4] Due to a left-skewed equilibrium, only a negligible fraction of labeled material is “lost” from this liquor during hybridization to the array. Serial hybridization using the same material is possible, and is in fact integral to multiple-chip array sets used by designs for which a single chip provides insufficient probe counts. In context of my experiments, re-hybridization in the context of technical replicate generation was considered but soon dismissed – systematic noise between serial hybridizations of the same sample was virtually nonexistent. All hybridization cocktails were preserved as “backup”, where they occasionally found use following failure of downstream array washing or scanning.

ized probe intensities across millions of individual probes are analyzed further.

Unlike two-color arrays, the output of tiling arrays is not gene- or feature- centric. Our two-color array probes were designed to specifically target the ~6000 annotated *S. pombe* genes; probe names were, in fact, simply the gene to which the amplicon (and probe) corresponded. The expression ratio for any (covered) gene is provided by the single probe covering the 3' end of the gene in question. Tiling array probes, as mentioned previously, were generated without specific regard to the genes or other features which they cover. Multiple probes cover each *S. pombe* gene, and many probes correspond to portions of the genome with no annotated content. Quantitation of gene expression on tiling arrays therefore required mapping of probes to genomic coordinates, determining which probes overlap each genomic feature, and calculation of a value which summarizes the information content of all relevant probes for each gene.

The two array platforms are functionally quite different. Aside from the benefits of commercial availability and quality control over homemade arrays and processing equipment, marked differences in sample requirements exist. Targets for our two color arrays were long cDNAs which incorporated an primary-amine modified dUTP during reverse transcription. These ostensibly full-length cDNAs had to be dye-coupled with an hydroxysuccinimide reactive dye moiety prior to hybridization. The dyes used are inherently unstable, severely limiting the long-term storage of coupled material. Although reverse transcription of a common modified nucleotide theoretically reduced dye-dependent effects over direct incorporation of dye-functionalized nucleotide analogs, poorly characterized differences in fluorescence between Cy3 and Cy5 required the deconvolution of dye effects.

Data management issues Microarrays are inherently prolific data sources. Unlike more focused biological tools, thousands to millions of data points are generated from each array experiment, and the resulting information can often be re-mined to address additional specific questions. As such, it is critical to maintain a complete record of all data generated by each hybridization. For all experiments, images of the hybridized substrate are archived and the resulting numeric matrix of spot intensities is carried through further analysis. In the case of yeast expression arrays covering a few thousand genes, these lists are manageable by most off-the-shelf spreadsheet packages. Indeed, early work in the lab relied exclusively on Excel® for analysis, though the limitations of this package quickly became apparent[5]. As mentioned above, expression experiments normally encompass a large number of hybridizations each consisting of a (relatively) small number of probes. Ones major data management concern becomes organization of individual hybridizations and array designs across a large number of experiments, and comparison of heterogeneous array designs; the sheer volume of values being analyzed is quite low. In stark contrast, tiling arrays generate deep data, generally across a smaller number of experiments or conditions. 200 array hybridizations each profiling 6,000 genes generate 1.2 million values. Each of my tiling hybridizations generates more (numeric) information than that contained in the sum total of all published *S. pombe* cell-cycle experiments (Rustici et al., 2004; Oliva et al., 2005; Peng et al., 2005). The computational burden becomes simultaneous analysis of millions of measurements across a relatively small number of hybridizations.

[5] The use of three character "common names" in *S. pombe* nomenclature has a few unfortunate serendipitous matches in other English conventions, including calendar months. Genes such as "sep1" are quickly mangled into formatted dates (10-01). While we noted this behavior early in our analyses (which only helped reinforce my exclusion of Excel from any analytical pipeline), such assisted renaming issues have propagated throughout the community and are visible in numerous high-profile public datasets.

Computational methods are becoming part of the prototypical biologist's toolbox. Many ad-hoc packages for the analysis of different data types exist (often provided by the manufacturers of each respective instrument/assay), but these tools often lack flexibility, transparency, and expandability. The Bioconductor Project (Gentleman et al., 2004), and the underlying R project for statistical computing have become the de facto standard platform for microarray analysis. As a community supported, open-source, peer-reviewed set of tools, Bioconductor/R addresses many of the shortcomings of individual analysis tools, and provides a common hub for integrating and comparing disparate data types. Of particular relevance to my transcript mapping studies are the affy (Irizarry et al., 2003), tilingArray (Huber et al., 2006), and vsn (Huber et al., 2002) packages , which provided the framework for import and normalization of raw scan data.

Data warehousing has become an integral part of high-throughput biology. For many publications, supporting work product has shifted from a handful of strains or antibodies to thousands -> billions of individual measurements. For example, including the output of one tiling array experiment as a supplement to this document would add nearly 40,000 filled pages[6]. The scientific community has responded to this need for centralized, standardized storage following numerous pleas of the very labs at the forefront of data generation (Ball et al., 2002a,b, 2004). All microarray data included in this manuscript have been deposited in ArrayExpress, under accession numbers listed in the respective reprints.

Mapping tiling probes to genomic space Since its formal publication in 2002 (Wood et al., 2002), the *S. pombe* reference genome has undergone many non-trivial

[6] Fitting 2598544 values at 66 lines per page would permit only the most brief of annotations in a total of 39372 pages. Unfortunately, I doubt that the standard repositories can boast similar longevity as the acid-free paper on which this document is printed.

structural revisions and sequence changes. The Affymetrix tiling arrays used in these studies were designed to cover a depreciated revision of the *S. pombe* genome (September 2004), and were designed without consideration for probes which could potentially cross hybridize to multiple genomic coordinates. An early challenge in analysis of data generated on these arrays was the assignment of probes to current coordinates and the classification of probes into different levels of target uniqueness (ie, zero, one, two, or >2 genomic matches).

Briefly, probe sequences for all 2,598,544 (1612^2) positions on the array were determined and the resulting 25mers aligned to a current reference genome (April 2007) using MUMmer (Delcher et al., 2002). Only probes which matched perfectly (25/25 nt) were assigned genomic coordinates and considered mapped, and any probe for which ≥23/25nt matched multiple genomic positions was flagged as having cross-hybridization potential (placing the onus of excluding these potentially confounding probes on downstream tools). In addition, a separate flag was added to probes that overlapped an annotated sequence feature, providing a very rough categorization into "likely transcribed" and "background" probes[7].

Annotations resulting from this mapping process relate genomic position to a coordinate on the original array surface. An unanticipated benefit of this convention was the ability to quickly import raw scan files from outside labs (using the same platform, of course) into a homogeneous coordinate system[8]. This re-mapping has

[7] This was known to be a gross assumption even at the time of genome alignment, as (unknown) UTRs and unannotated genes were included in the "background" category. The extent of this fallacy was not clear until after the analysis of our tiling data, at which time it became clear that stable transcripts corresponding to more than 90% of the genome exist in vegetative cells (see 3.3 on page 107). Downstream use of these "background" probes was always done with the assumption that the population would be a mixture of "transcribed" and "untranscribed" intervals.

[8] An unfortunate inconsistency of Affymetrix scan control software is the autorotation of array data to place the "human readable" imprint in conventional left-to-right orientation, instead of the bottom-to-top orientation as it is physically present on the chip (90° clockwise rotation). This function was removed for new (non-upgraded) installations of the software, resulting in data heterogeniety between labs and individual scanner installations. I have not discovered the header byte

permitted the side-by-side analysis of my data with transcription and chromatin IP experiments from other labs using the Affy tiling platform (Wilhelm et al., 2008). Furthermore, the tools created for mapping oligonucleotide probes to a given genome are not limited to Affymetrix designs; I have used the methods described above to map heterogeneous array designs to a common genome build, such as those generated on Agilent tiling arrays (Dutrow et al., 2008). This method of mapping has proven quite effective for other platforms and enabled a more direct comparison of disparate data sets, as detailed in the manuscripts below.

2 The Cell-Cycle Regulated Genes of *Schizosaccharomyces pombe*

2.1 Rationale

Previous landmark work in the lab systematically described the cell-cycle oscillating genes of budding yeast (Spellman et al., 1998). Briefly, microarrays containing probes for nearly all known *S. cerevisiae* genes (created in similar fashion to DeRisi et al. (1997)) were used to measure transcript abundance as cells progressed through the vegetative cell cycle as synchronized by various means (see Futcher (1999) for an excellent review of methods). Additional experiments were included to directly determine the targets of known cell-cycle regulators, for example induced expression of SBF and MBF responsive genes by induction of GAL-CLN3. Using this approach, 800 genes were identified as "cell cycle regulated", nearly eight fold more genes than had previously been shown to oscillate in a cell-cycle dependent manner, and more

which denotes rotation of the included data (although some distinguishing feature is clearly present, as Affy binaries are able to deal seamlessly with both types of scans), so great care must be taken to account for this rotation, when present.

than three times the previously projected number of regulated genes (Price et al., 1991).

A major conceptual advance made in Spellman et al. was the analysis of upstream sequence elements shared between co-regulated groups of genes. Gibbs sampling algorithms designed to identify overrepresented sequences, or motifs, had previously been applied to find shared domains in hand-curated groups of protein primary sequence (Lawrence et al., 1993; Neuwald et al., 1995), and were modified to identify DNA sequence elements which were statistically overrepresented in the 5' non-coding regions of co-regulated genes. The definition of "co-regulated genes" was fundamentally extended in this work, under the assumption that co-expressed genes are implicitly co-regulated. Unsupervised agglomerative hierarchical clustering (described in Eisen et al. (1998)) was used to group genes with quantitatively similar expression patterns, vastly extending the repertoire of genes whose upstream sequences could be searched for regulatory motifs. This definition of 'clusters as coregulated genes' provided a fundamental shift in the analysis of gene regulation. Although any expression data set can be clustered, in practice, data from multiple experiments targeting varied regulatory pathways are necessary to separate genes whose expression is incidentally correlated. To a point, the wider variety of expression states that can be measured, the better.

Later work from the lab (Zhu et al., 2000b) used expression profiles of strains deleted for one or more related transcription factors (in this case, all Forkhead associated domain (Weigel et al., 1989) containing proteins) to provide indirect evidence of regulation by the disrupted factor(s). In addition to assigning a pair of transcription factors to an M/G1 cluster which previously lacked a regulator, this work demonstrated the feasibility and value of examining expression in transcription factor knockout backgrounds as a method of determining target gene identity; the combination of wild-type and mutant cell cycle profiles as input for cluster analysis permitted the discrimination of primary and secondary targets of the factors tested. These two manuscripts provided the conceptual groundwork for numerous studies of periodic expression in other organisms ranging from the developmentally dimorphic Caulobacter crescentus (Laub et al., 2000) to human cell lines (Cho et al., 2001; Whitfield et al., 2002b).

Studies of periodic expression in higher eukaryotes were performed before the availability of transient gene knockdown (Paddison et al., 2004) limiting early genome-wide expression studies in these organisms to available immortalized cell lines and tumor samples. As such, strategies used to tease apart temporally co-expressed clusters in budding yeast were largely unavailable in higher model organisms. Despite this limitation, biologically relevant co-regulated clusters were generated, and appreciable overlap in the cell-cycle regulated genes shared between budding yeast and mammals was observed. The immense phylogenetic distance between *Saccharomyces* and mammals presented fundamental difficulties in predicting the evolutionary fate of transcriptional regulators and their target genes; relatively recent changes unique to the lineage of either system would obscure erstwhile common patterns.

An orthogonal organism system would provide information crucial to determining the common ancestral regulatory scheme. The fission yeast *Schizosaccharomyces pombe*, roughly evolutionarily equidistant from yeast and mammal, is perfectly situated to provide such information. Fission yeast has been used as a model organism for genetic and cell-cycle research across more than fifty years , and has a robust repertoire of genetic tools available (reviewed in Forsburg (2003)). The fission yeast genome sequence, formally published in 2002 (Wood et al., 2002), made possible the creation of microarray probes and high-throughput analysis of expression in *S. pombe*. Generation of probes and manufacture of arrays was undertaken in the lab as part of a National Center for Research Resources (NCRR) core facility grant beginning in 2001; my tenure in the lab began shortly thereafter.

2.2 Methods

Our experimental approach was similar to that of the lab's previous work in budding yeast (Spellman et al., 1998), and is fully described in the included reprint. The definition of an "oscillating" gene is fundamentally different between the two works, however. For budding yeast, genes were ranked based on the magnitude of their oscillations' fit to a sinusoid. The resultant "CDC score" was strongly biased for genes with a high peak-to-trough ratio, with less regard to the robustness of fit. For example, an otherwise invariant gene with a few strong spurious "in-phase" measurements would score higher than a gene that oscillates, albeit weakly, with a perfect once-per-cycle periodicity. To overcome this limitation, I chose to quantitate a standard score (also called a Z-score) for each gene, providing a homogeneous comparison between genes regardless of their peak expression levels. Systematic noise still provides a hard floor precluding the detection of extremely weakly expressed transcripts in all systems, but the standard scores generated in this study more accurately reflect cell cycle regulation when compared to a simple magnitude.

2.3 Paper Reprint

The following was previously published in PLoS Biol. 2005 Jul;3(7):e225. Epub 2005 Jun 28.

The cell cycle-regulated genes of Schizosaccharomyces pombe. Oliva A, Rosebrock A, Ferrezuelo F, Pyne S, Chen H, Skiena S, Futcher B, Leatherwood J.

Department of Molecular Genetics and Microbiology, Stony Brook University,

Stony Brook, New York, USA.

Abstract

Many genes are regulated as an innate part of the eukaryotic cell cycle, and a complex transcriptional network helps enable the cyclic behavior of dividing cells. This transcriptional network has been studied in *Saccharomyces cerevisiae* (budding yeast) and elsewhere. To provide more perspective on these regulatory mechanisms, we have used microarrays to measure gene expression through the cell cycle of *Schizosaccharomyces pombe* (fission yeast). The 750 genes with the most significant oscillations were identified and analyzed. There were two broad waves of cell cycle transcription, one in early/mid G2 phase, and the other near the G2/M transition. The early/mid G2 wave included many genes involved in ribosome biogenesis, possibly explaining the cell cycle oscillation in protein synthesis in *S. pombe*. The G2/M wave included at least three distinctly regulated clusters of genes: one large cluster including mitosis, mitotic exit, and cell separation functions, one small cluster dedicated to DNA replication, and another small cluster dedicated to cytokinesis and division. *S. pombe* cell cycle genes have relatively long, complex promoters containing groups of multiple DNA sequence motifs, often of two, three, or more different kinds. Many of the genes, transcription factors, and regulatory mechanisms are conserved between *S. pombe* and *S. cerevisiae*. Finally, we found preliminary evidence for a nearly genome-wide oscillation in gene expression: 2,000 or more genes undergo slight oscillations in expression as a function of the cell cycle, although whether this is adaptive, or incidental to other events in the cell, such as chromatin condensation, we do not know.

Introduction

The yeasts *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* are excellent organisms for the study of the cell division cycle. Both yeasts have many well-characterized cell division cycle (cdc) mutants (Nurse et al., 1976; Culotti and Hartwell, 1971; Hartwell et al., 1970; Hartwell, 1971a,b), and both have a long history of genetic and molecular cell cycle studies. However, they diverged more than 1 billion years ago, and have many lifestyle differences.

In particular, the two yeasts have different cell cycles. *S. pombe* divides by fission, a symmetrical process in which a septum grows across the center of a long cylindrical cell, dividing the old cell into two equal new cells. Moreover, the main control point in the *S. pombe* cell cycle is a size control in G2, not in G1 as in *S. cerevisiae* and many other organisms. In *S. pombe*, when cells reach a critical size, the Cdc2 protein kinase is activated both by cyclin binding and also by Cdc25 phosphatase removal of the inhibitory phosphate from Tyr15 of Cdc2, and this leads to mitosis. Once nuclear division has occurred, the cell moves quickly into S phase without an appreciable G1. Therefore S phase is largely completed by the time cytokinesis/cell separation occurs. Thus, when the cells are growing in good conditions, cells have a long G2, and most cell cycle–specific events are completed in a relatively small portion of the cell cycle encompassing M, G1, and S, with S occurring coincident with cytokinesis. When conditions are poor, a cryptic size control appears in G1 phase; that is, a G1 phase appears and becomes longer as growth rate becomes slower.

In contrast, *S. cerevisiae* divides by “budding,” an inherently asymmetrical process whereby a large mother cell generates a small daughter bud. Once born as a separate cell, the small daughter grows in volume through a long G1, and commits to division at a G1 event called “START.” START involves the activation of a pair of closely related transcription factors, MBF and SBF, and the induction of 100 or more genes. After START, DNA synthesis is initiated, and a bud forms. There is a short G2 phase, followed by mitosis and cytokinesis, and then cells enter the next G1. When cells are growing rapidly in good conditions, G1, S, G2, and M phases are of similar lengths, and so various cell cycle–specific events are distributed somewhat equally around the cycle. However, when cells are growing slowly in poor conditions, almost all the increased length of the cell cycle is accounted for by an increased G1, and most cell cycle–specific events occur over a relatively small percentage of the cell cycle, encompassing "START," S phase, and mitosis.

Microarrays have been used to analyze gene expression in synchronized *S. cerevisiae*. There are at least 800 genes whose transcripts oscillate as a function of the cell cycle (Spellman et al., 1998). The cataloging of these transcripts has helped describe what happens in a cell cycle. In addition, because many of the oscillating genes are regulatory, the microarray analysis has helped us understand how the *S. cerevisiae* cycle is regulated. In view of the fact that *S. pombe* also has a well-studied cell cycle and because these two yeasts have both differences and similarities in the way they carry out a cell cycle, it is of interest to characterize oscillating transcripts in *S. pombe* also, to understand at a deeper level what is preserved and what changes across the cell cycles of these two model eukaryotes.

Recently, Rustici et al. (2004) and Peng et al. (2005) have published microarray analyses of *S. pombe* cell cycle genes. Our results are broadly similar to theirs, but as described below, each group finds a somewhat different set of genes. There is excellent agreement between the groups with respect to the most strongly regulated genes, but naturally there is less agreement for more weakly regulated genes. Here, we concentrate on the 750 genes that are most strongly regulated, but we believe that there may be a total of 2,000 or more genes that have at least weak cell cycle regulation. A large number of weakly to moderately oscillating genes peak in G2 phase, and these are highly enriched for functions in ribosome biogenesis. Our analysis of the cell cycle–regulated promoters shows them to be surprisingly complex, and shows clusters of multiple regulatory motifs similar to clusters of motifs found in the developmental genes of *Drosophila*. Although Rustici et al. have pointed out several differences between the cell cycles of *S. pombe* and *S. cerevisiae*, we find that there are also striking similarities, suggesting deeply conserved mechanisms.

Results

Synchronous Cultures and Identification of Cell Cycle–Regulated Transcripts Three synchronous cultures were studied, one generated by cdc25 block release, and two generated by elutriation. Each culture was sampled through three cell cycles, giving nine cell cycles of data. Synchrony and cell cycle position were assayed by scoring initiation of anaphase and septation microscopically (figure 4). RNA was extracted, converted to cDNA, labeled, and hybridized to arrays, and then fluorescence was analyzed. Gene expression was assayed as a ratio of experimental cDNA to asynchronous control cDNA. In total, approximately 1.2 million data points were generated from cell synchrony experiments. Fourier analysis identified cyclic expression patterns. Monte Carlo simulations were used on shuffled expression-ratio data, and compared to the actual, cyclic data, to generate a p-value for the cyclicity of each gene. These p-values ranged from less than 10^{-16} for the most cyclic genes (i.e., the probability that the observed oscillation occurs by chance is less than 10^{-16}), to 0.997 for the least cyclic gene of the 5,000 studied. We ranked all 5,000 genes by p-value, with the most significantly oscillating genes at the top. The amplitude of the oscillation is a major contributor to the p-value, so genes with higher amplitude oscillations tend to rank higher than genes with lower amplitudes.

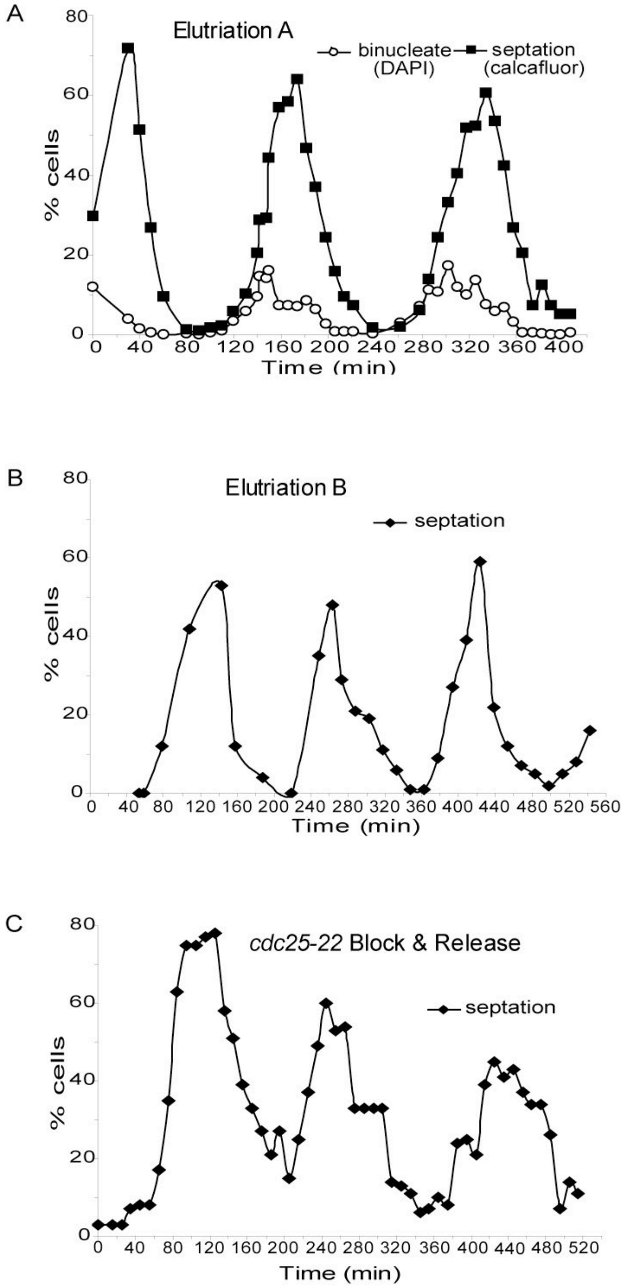


Figure 4: Synchrony of time courses analyzed.

A: Samples from elutriation A were double stained with calcofluor (for septa) and DAPI (for nuclei). Cells were assayed for initiation of anaphase by scoring cells with two nuclei but no septum (binucleates, open circles). The cells were also scored for septation (filled squares). B: Cells from elutriation B were assayed for septation by phase contrast microscopy. C: Cells from the cdc25-22 block release were assayed for septation by phase contrast microscopy.

The list of all 5,000 genes ranked by p-value and other associated information such as time of peak expression is given in Table S1. The raw data have been deposited at ArrayExpress[9]. The raw data, including all figures and all tables, are available online[10].

The distribution of genes versus p-values is shown in figure 5 on the following page . There is no clear distinction between "cyclic" and "non-cyclic" genes. Rather, after the best 203 genes, there are simply more and more genes as one goes to poorer and poorer p-values.

Because the distribution of genes versus p-values continuously increases after gene 203, one must choose a somewhat arbitrary threshold for discussion of cell cycle–regulated genes. We have chosen to discuss the best 750 genes in our p-value list. This number is similar to the number of genes chosen by Peng et al. and Rustici et al. as being cell cycle regulated (747 and 407, respectively), and similar to the number of genes chosen for the yeast *S. cerevisiae* (800) (Spellman et al., 1998), thus facilitating comparison of these gene sets. In the vicinity of the 750th gene (and even below), most genes display an oscillatory behavior to the eye, at least in one or two of the three experiments. Finally, the number 750 is obviously somewhat arbitrary, and indeed we have no basis for anything other than an arbitrary cutoff. Because the list of genes is ranked, other investigators may choose their own sets of oscillatory genes from our *p*-value list (Table S1) by choosing any desired cutoff. For the top 750 genes, the false discovery rate is 0.00022, so on a statistical basis, less than one false positive is expected in the list of 750.

Although we will discuss primarily these 750 best genes, there are many more genes that appear to oscillate slightly. A total of 2,262 genes (nearly half the genes

[9] http://www.ebi.ac.uk/arrayexpress/

[10] http://publications.redgreengene.com/oliva_plos_2005/

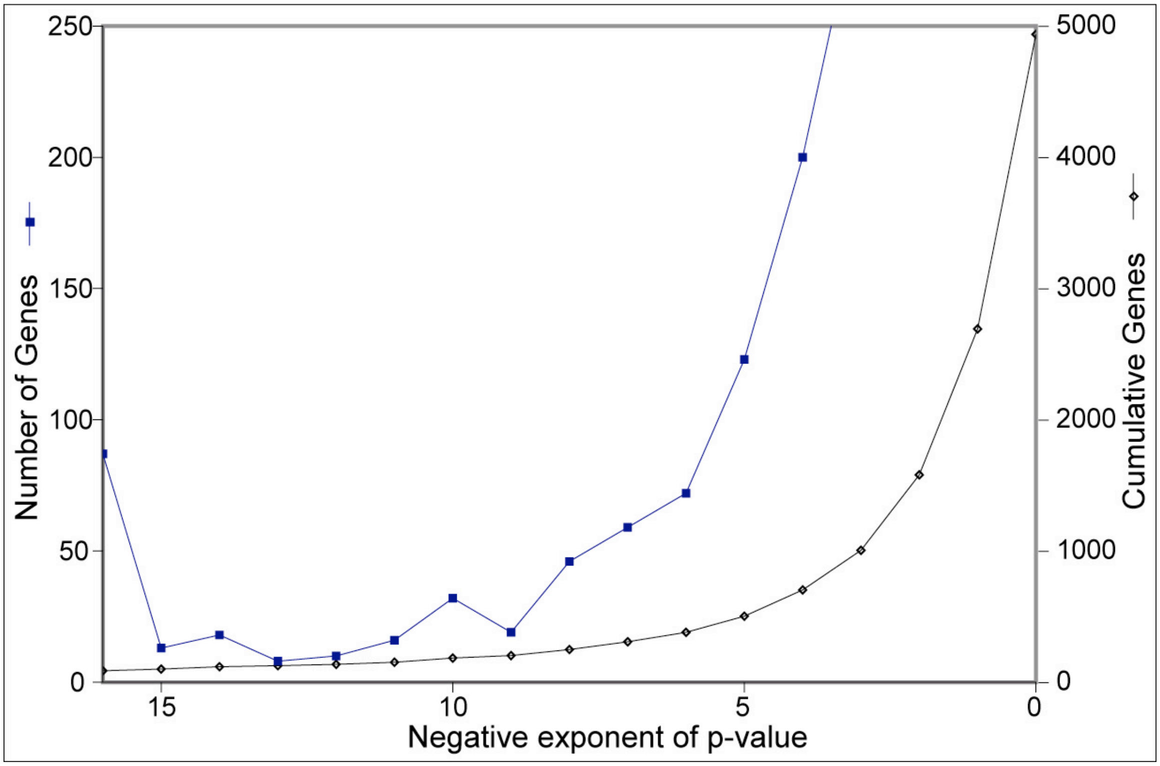


Figure 5: Distribution of p-values for cell cycle regulated genes.
The x-axis shows bins of p-values of the significance of cell cycle regulation. From the left, the bins are: 1. Genes with p-values less than 10^{-16} (87 genes); 2. Genes with p-values between 10^{-15} and 10^{-16} (13 genes); 3. Genes with p-values between 10^{-14} and 10^{-15} (13 genes); 4. Genes with p-values between 10^{-13} and 10^{-14} (8 genes); etc. The number of genes in each bin is shown on the left y-axis (dark blue squares). Also shown (right y-axis, magenta diamonds) are the cumulative number of genes at each p-value or lower. Thus there are about 1000 genes with a p-value of 10^{-3} or less.

in the genome!) have a p-value less than 0.05, the usual statistical cutoff. Based on the false discovery rate, we would expect about 53 of these to be false positives, but even so, this leaves well over 2,000 genes with a slight but statistical oscillation.

Previously, 37 cell cycle–regulated genes have been reported in *S. pombe*; 29 of these (78%) are in our top 750. Of the eight that are not in our top 750, two are in the top 1,000. The remaining six (cdc19/mcm2, cmk1, dmf1/mid1, ppb1, uvi22/rrg1, and suc22) are also not in list of 407 of Rustici et al., and three of these (cdc19/mcm2, cmk1, and ppb1) are also not in the list of Peng et al.. Thus, these genes are probably quite weakly regulated (except for suc22, for which there are two transcripts, one regulated and one not (Harris et al., 1996)). The top 750 genes are shown in figure 6 on the next page in order of time of expression (i.e., ordered by cell cycle phase). About halfway down this phasogram is an apparent discontinuity; this corresponds to the mid to late G2 trough, when there are relatively few cell cycle–regulated genes (see below).

Rustici et al. have recently compiled a list of 407 periodically-expressed *S. pombe* genes, and while our manuscript was in review, Peng et al. identified 747 similar genes. A comparison of the three studies is shown in figure 7 . The total number of genes found to oscillate in at least one study is 1,373. Of these, 1,013 were unique to just one of the studies, whereas 360 were found in two or three studies, and 171 were found in all three.

Despite the fact that 1,013 genes were found in only one of the three studies, we believe that most of these 1,013 do indeed oscillate to some extent. There are two lines of evidence. First, most of the genes do display a clear oscillatory pattern to the eye, at least in one of the studies. For instance, figure 8 on page 52 shows the 516 genes found by us but not by Rustici et al. (63 of these were also found by Peng et al., but the remainder were unique to us). At least in part, we found these

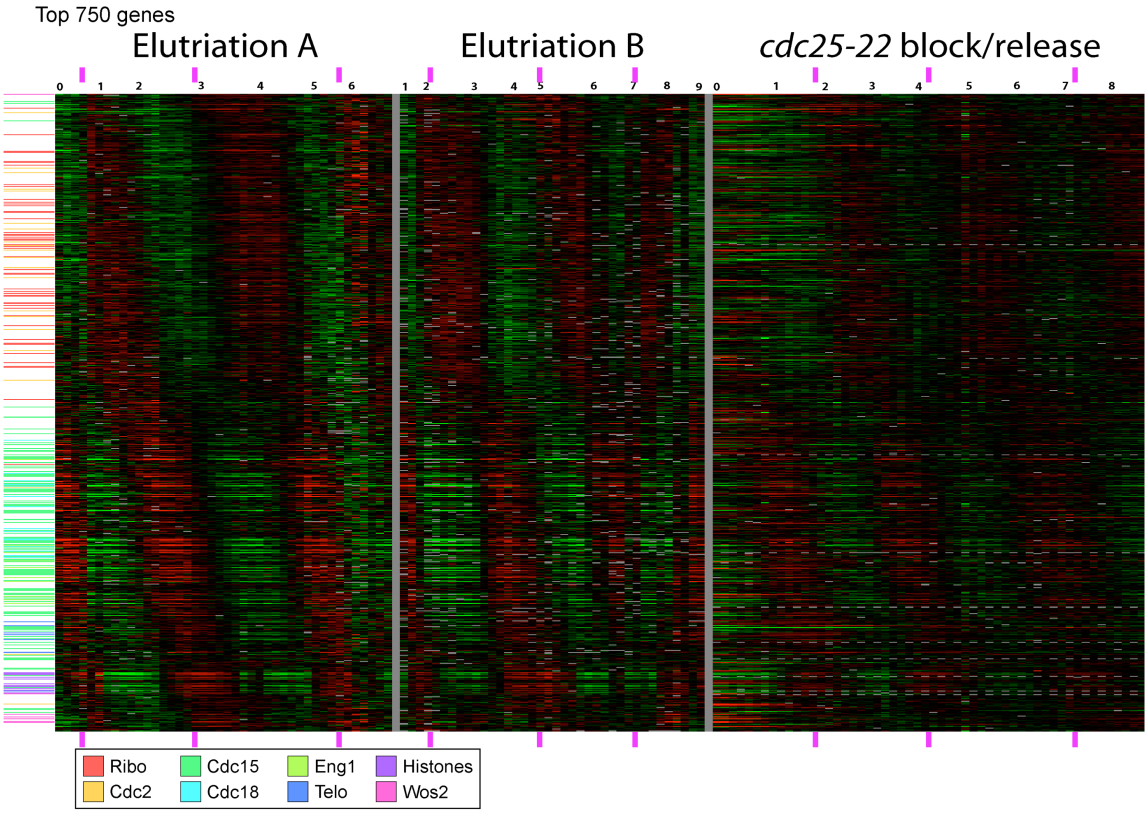


Figure 6: Top cycling genes ordered by time of peak expression

Expression data for the top 750 genes is shown, with genes ordered by time of peak expression. Every row represents a gene; every column represents an array from a time-course experiment. Red signifies up-regulation (i.e., an experiment/control ratio greater than one); green signifies down-regulation (i.e., an experiment/control ratio less than one). Black is a ratio close to one, and grey is missing data. Dynamic range is 16-fold from reddest red to greenest green. The time in hours since the beginning of the time course is shown in black numerals at the top of Figure 3. The peaks in septation index are marked with purple rectangles at the top and bottom of the figure. Genes from defined clusters are marked on the left by colored lines, according to the cluster color code shown at the bottom of the figure.

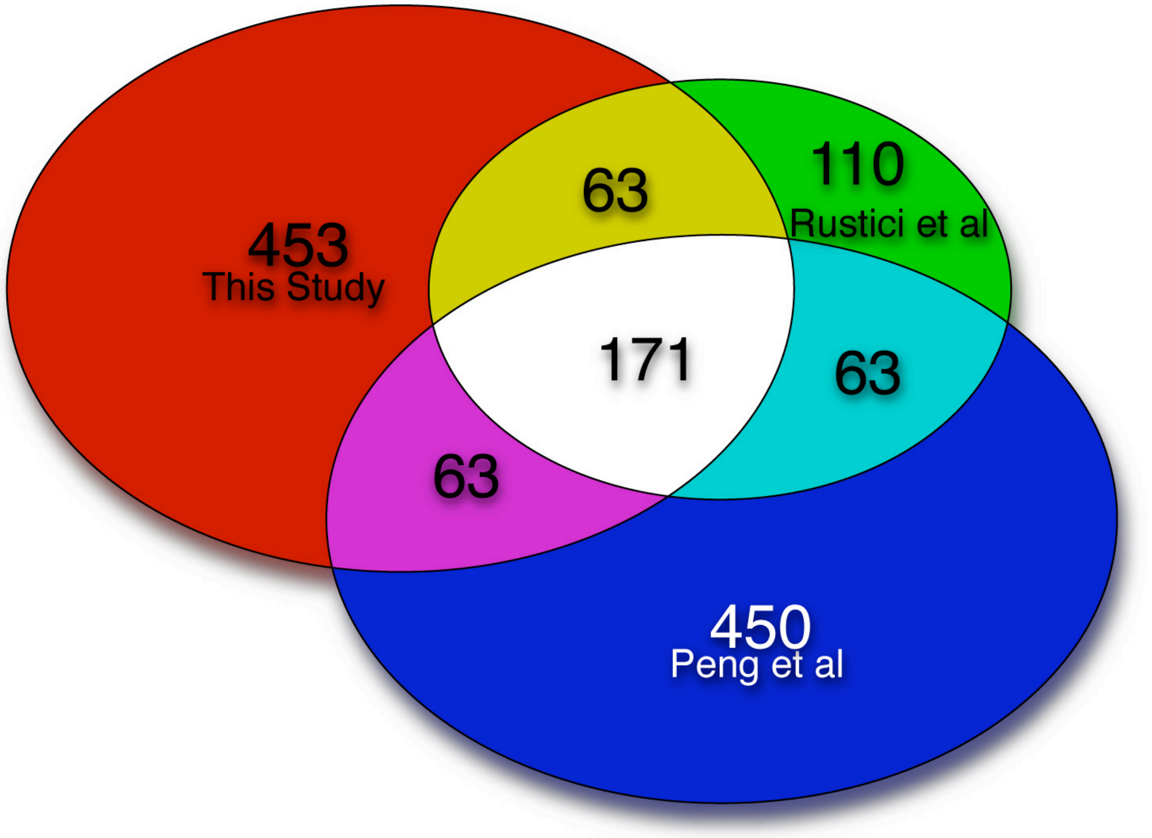


Figure 7: Overlap between cell cycle microarray studies.
A Venn diagram of the overlap between the three lists of cell cycle regulated genes from this study, Rustici et al, and Peng et al listing genes.

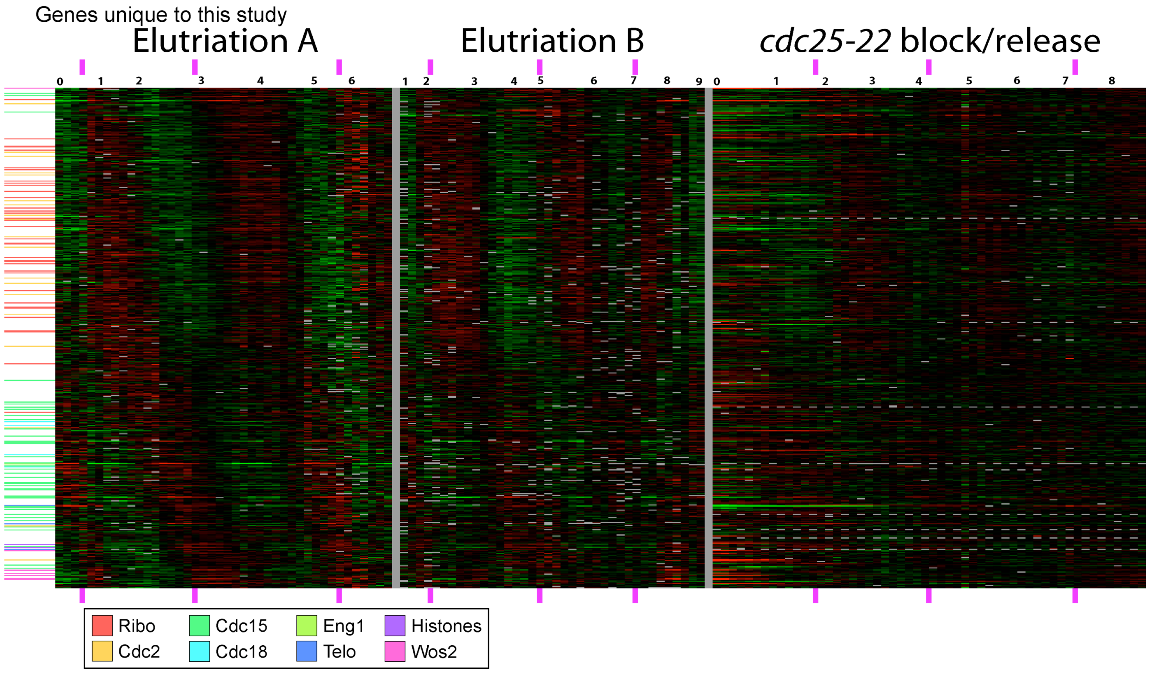


Figure 8: Cycling genes uniquely identified in this study
As 6 on page 50, but only the 514 genes found in our study but not found by Rustici et al. are shown.

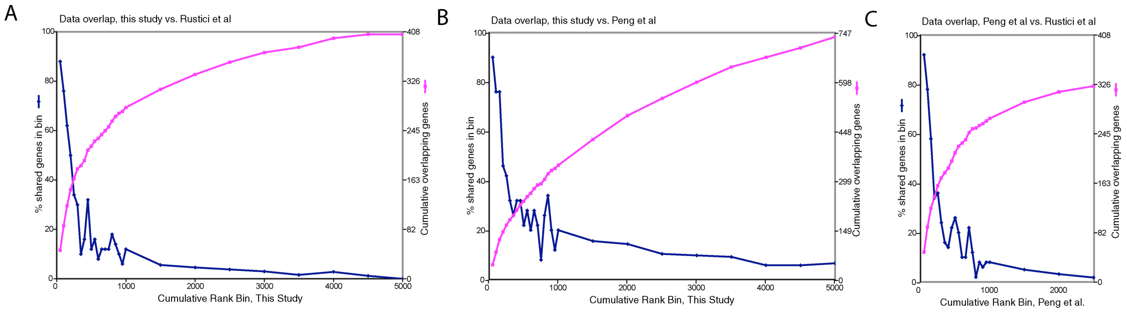


Figure 9: Overlap between Different Cell Cycle Microarray Studies, by Rank
(A) Our ranked list of cell cycle–regulated genes is divided into consecutive sets, or bins, of 50 genes. For each set of 50 genes, the number of genes in that set also found in the list of 407 cell cycle genes of Rustici et al. is plotted on the left y-axis. For instance, of our best 50 genes, 44 (88%) are found in the list of of 407 genes of Rustici et al., and of our next-best 50 genes, 38 (76%) are also in their list. For the top 15 bins (750 genes), every bin of 50 genes is represented. Afterward, the number plotted represents an average over several bins. The cumulative number of genes in the list of 407 is plotted on the right y-axis.
(B) As (A), but the bins in our study are compared to the list of 747 genes of Peng et al..
(C) As (A), but the ranked list of Peng et al. is divided into bins, and compared to the list of 407 of Rustici et al. Because Peng et al. ranked only their top 2,700 genes, the graph is truncated after gene 2,700, and the cumulative number of genes rises to only 325.

genes because our elutriated cells were more synchronous than those of Rustici et al. (compare our figure 4 on page 46to Figure 1B in the supplemental data of Rustici et al.), thus allowing detection of genes with moderate amplitudes.

The second line of evidence is that most of the 1,013 genes unique to one study also display some statistical oscillatory behavior in one or both of the other studies, even though this behavior is not strong enough to surpass the threshold for inclusion on the cell cycle list in those studies. This effect is shown in figure 9 As might be expected, the top genes in our ranked list and the ranked list of Peng et al. show excellent (approximately 85%) agreement with Rustici et al.. The degree of agreement then drops as one proceeds down the ranked lists. But genes below rank 750 but above rank 2,500 in either list are much more likely to be in the list of Rustici et al. than are genes below rank 2,500. In other words, a gene unique to the study of Rustici et al. is likely to show some oscillatory behavior in the other two studies (i.e., be in the top half of the lists). Analogous comments apply to the genes unique to us, and genes unique to Peng et al..

Before the publication of Peng et al., we had compared our study to that of Rustici et al. to look for discrepancies. We identified a total of 21 genes (11 from Rustici et al., ten from us) that appeared very strongly regulated in one study, but not at all regulated in the other. We have now checked these 21 genes against the results of Peng et al., and find that 17 of the 21 appear regulated in Peng et al., whereas four (three from us and one from Rustici et al.) do not appear regulated. Thus it seems that both we and Rustici et al. have been conservative in our identification of cell cycle–regulated genes and tend to get false negatives rather than false positives.

In summary, the three cell cycle lists together implicate about 1,300 genes, and our ranked p-value list does not become worse than a p-value of 0.05 until gene number 2,262. We believe that a very large number of *S. pombe* genes, 2,000 or more, have at least a weak cell cycle oscillation.

Two Genome-Wide Waves of Transcription To examine the distribution of gene expression around the cycle, Fourier analysis was used to determine the time at which each gene's expression peaked (the "phase angle" of peak expression). For genes in the bottom half of the 5,000 gene rank list (i.e., genes that did not cycle appreciably), phase angles were largely determined by noise, but nevertheless would tend toward the peak of any weak cyclic behavior that may have existed. The number of genes peaking at each time in the cycle was plotted (figure 10) for four groups of genes: the most-regulated 750 genes (panel A of figure 10), all genes (panel B of

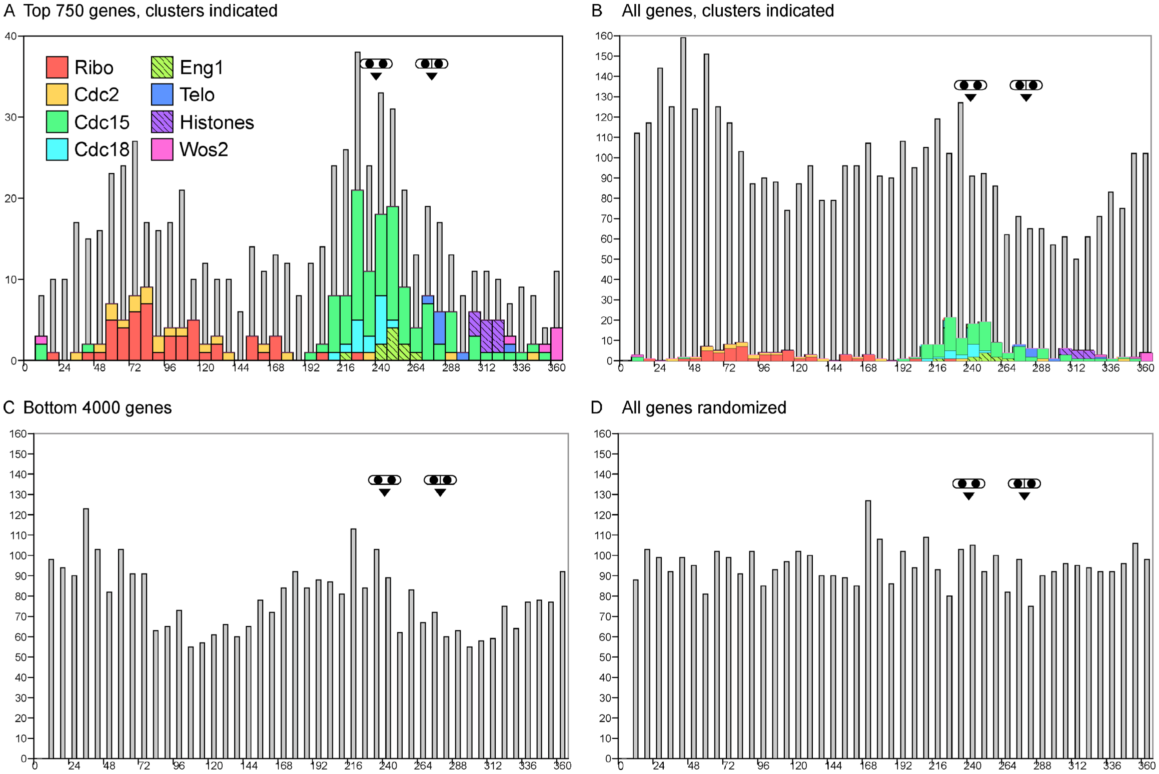


Figure 10: The Number of Genes Peaking during Each Portion of the Cell Cycle The cell cycle was divided into 45 consecutive portions, or bins. If the cell cycle is considered as a circle of 360°, then each bin occupies 8°. Every gene was analyzed using a Fourier transform to determine the time of peak expression in elutriation A, from 0° to 360°. The number of genes peaking in each bin was summed and plotted (grey bars, background), with the number of genes in each bin shown on the y-axis. Genes in specific clusters are shown by colored bars (foreground). The Fourier transform calculation was similarly used to derive the time of peak septation and peak binucleate cells (from Figure 1) and these cell cycle landmarks are indicated.

(A) The top 750 cell cycle–regulated genes were analyzed for time of peak expression. Genes from different clusters are "stacked" when they occur in the same bin.

(B) All genes (~5,000) were analyzed for time of peak expression, exactly as in (A).

(C) The bottom 4,000 genes were analyzed for time of peak expression. Data was extracted from arrays before the red/green normalization step, so that the bottom 4,000 genes would not be affected by the normalization and the cyclic expression of the top 750 genes.

(D) All genes (~5,000) were analyzed for time of peak expression after genewise random shuffling of microarray observations. This randomization serves as a negative control for the Fourier calculation in parts (A), (B), and (C).

figure 10), the least regulated 4,000 genes (panel C of figure 10), and all genes after random shuffling of ratio data (panel D of figure 10). The peaks of septation and binucleates were also determined by Fourier analysis. (See Materials and Methods for information on red/green normalization.)

There were two striking findings. First, it appears that there are two broad waves of gene expression, one peaking in early to mid G2, and the second peaking in late G2/M, whereas there are troughs in mid to late G2, and in S. The early/mid G2 peak contains the Ribosome biogenesis cluster (see below) and associated genes, whereas the late G2/M peak contains the genes of the Cdc15, Cdc18, and Eng1 clusters (see below), which are important for M and S. Second, the two waves of gene expression were seen even in the 4,000 least-cyclic genes. As noted above, there is statistical evidence from p-values that 2,000 or more genes may oscillate slightly. The two waves of expression seen for the bottom 4,000 genes confirm that many of these genes do indeed oscillate. If the fluctuations in these 4,000 genes had simply been due to noise, then the peak phase angles would have been uniformly distributed from 0° to 360° (as confirmed by repeating the analysis on shuffled data; panel D of figure 10 Combined with the evidence of the p-values, this analysis suggests that many, or most (or possibly all!) of the least-cyclic 4,000 genes do in fact oscillate slightly, and that there are two nearly genome-wide peaks in gene expression. These peaks might represent periods when the cell is preparing for a high level of cell cycle–specific activity, or when transcription (of any kind) is activated on a genome-wide basis. Alternatively, one might focus on the troughs, which might represent periods with little cell cycle–specific activity, or periods when transcription is repressed on a genome-wide basis (see Discussion).

Cluster Analysis To study the regulation of the cell cycle, we wished to find clusters of co-regulated genes potentially responding to the same transcription factor. However for this purpose it is not sufficient to find genes expressed at the same time, because such genes might be responding to different mechanisms of regulation. This is an acute problem in *S. pombe*, because mitosis, DNA synthesis, and cytokinesis all occur in a small window of the cell cycle under standard growth conditions.

Therefore our analysis included not only our three time courses of synchronous cells, but also eleven other array experiments that more directly addressed regulatory mechanisms. These experiments (see Materials and Methods) included small cells grown in poor nitrogen to induce a G1 phase; a cdc10-M17 block-release experiment, to separate S phase events from cytokinesis and septation events; an arrest at G1 (using cdc10-M17, encoding MBF transcription factor subunit); an arrest at S (using cdc22-M45, encoding ribonucleotide reductase); an arrest at late G2 (using cdc25–22, encoding the phosphatase that activates Cdc2); an arrest at M (using nuc2–663, encoding a subunit of the anaphase promoting complex); and finally, from the data of Rustici et al., experiments using a constitutively active allele of cdc10 (cdc10-c4), null and over-expressor alleles of the forkhead transcription factor sep1, and null and overexpresser alleles of the transcription factor ace2.

Hierarchical clustering was used (Eisen et al., 1998) because the underlying structure of a gene regulatory network is somewhat hierarchical. Thus, a hierarchy found by the clustering algorithm is often interpretable in terms of a hierarchical transcriptional network existing in the cell (see S. J. Gould's essay, "Linnaeus's Luck?"[11], for an illuminating discussion of this issue in a different context).

The clustergram of 750 genes is shown in figure 11 on the following page . We chose eight clusters for analysis and discussion, on the basis that the genes in these

[11] http://www.findarticles.com/p/articles/mi_m1134/is_7_109/ai_65132190

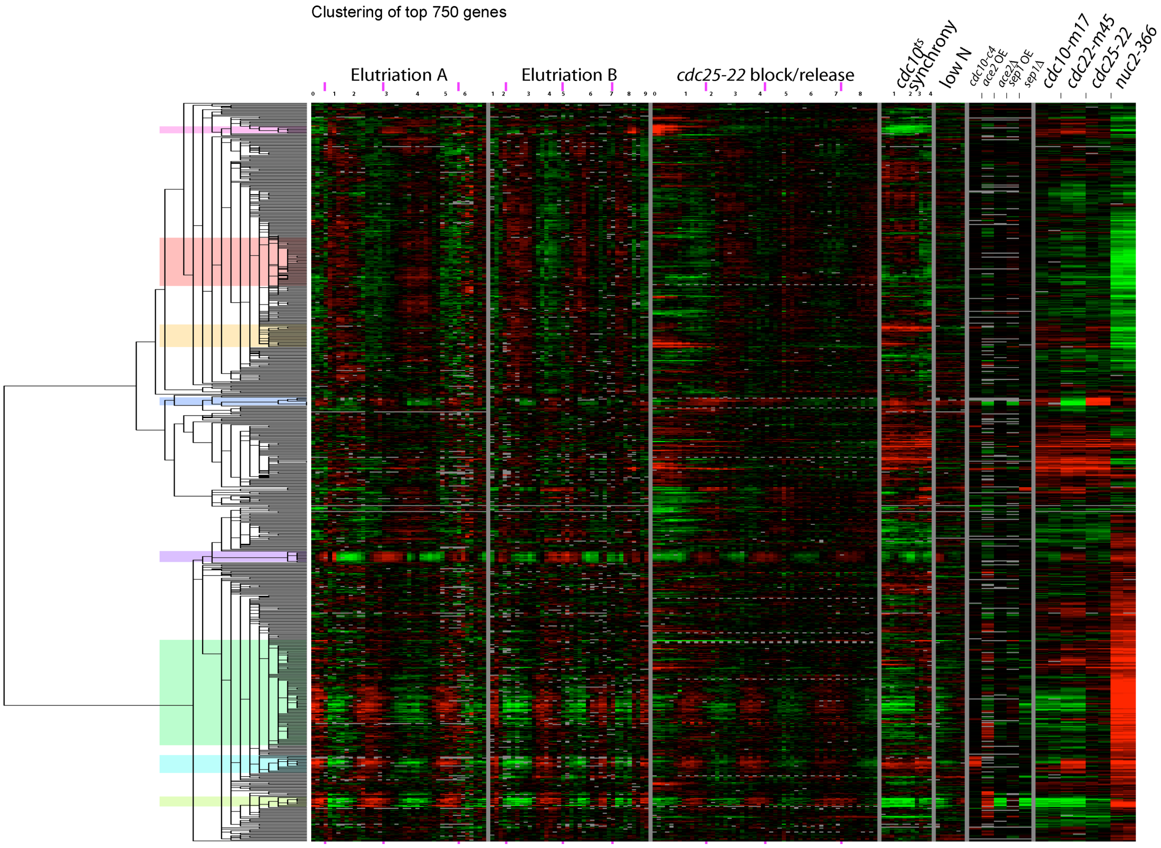


Figure 11: Cluster Analysis of the Top 750 Cell Cycle–Regulated Genes

Gene expression data from all experiments were clustered by a hierarchical method (Eisen et al., 1998). Every row represents a gene; every column represents an array. Red signifies up-regulation (i.e., an experiment/control ratio greater than one); green signifies down-regulation (i.e., an experiment/control ratio less than one). Black is a ratio close to one, and grey is missing data. Dynamic range is 16-fold from reddest red to greenest green. The time in hours since the beginning of time-course experiments is shown in black numerals at the top of the Figure. Peaks in septation index are marked with purple rectangles at the top and bottom of the Figure. Clusters discussed in the text are marked with blocks of color. Data for the cdc10-C4 (asynchronous cells with the hyperactive allele cdc10-C4), ace2 OE (asynchronous cells over-expressing ace2), ace2Δ (asynchronous ace2Δ cells), sep1 OE (asynchronous cells over-expressing sep1), and sep1Δ (asynchronous sep1Δ cells) are taken from Rustici et al. (2004). cdc10 encodes a component of the MBF transcription factor; ace2 encodes the Ace2 transcription factor, and sep1 encodes a forkhead transcription factor. Other experiments are described in Materials and Methods.

clusters are particularly tightly co-regulated. Most of the clusters are named for one representative gene. The clusters are, from the late G2/M wave, the Cdc15, Cdc18, and Eng1 clusters; from S and early G2, the telomeric, histone, and Wos2 clusters; and from the early to mid G2 wave, the Ribosome biogenesis and Cdc2 clusters.

If the genes in each cluster are truly co-regulated, then the promoters of these genes will be bound by the same transcription factor, and therefore the promoters should share a common DNA sequence motif corresponding to the transcription factor binding site. We searched for such motifs upstream of the genes in each cluster. We used three motif search programs: AlignAce, a Gibbs-sampling algorithm (Hughes et al., 2000); SPEXS[12], a word-count algorithm (Brazma and Vilo, 2001; Vilo et al., 2000), and MEME[13], an expectation-maximization algorithm (Bailey and Elkan, 1995; Bailey and Gribskov, 1998). In general, all three programs found the same motifs.

In the study of Rustici et al., four clusters were found. It is difficult to compare the clusters of Rustici et al. with ours: The genes, experiments and clustering methods were different. However, in general, the clustering of Rustici et al. tended to produce fewer, larger clusters, and focused on time of expression as the main distinction between the clusters, whereas our method produced more, smaller clusters, and focused on regulatory mechanisms (as well as time of expression). Peng et al., like us, used hierarchical clustering and found eight clusters, some of which are quite comparable to ours. However, again, we put more emphasis on regulatory mechanisms as opposed to time of expression, and this generated some different clusters.

The M Clusters The wave of expression in the late G2 and M phases includes most of the strongly regulated genes. This wave contains three major clusters, which we call the Cdc15, Cdc18, and Eng1 clusters (figures 12, 13, and 15). Functionally, these

[12] http://www.egeen.ee/u/vilo/SPEXS/

[13] http://meme.sdsc.edu/meme/website/intro.html

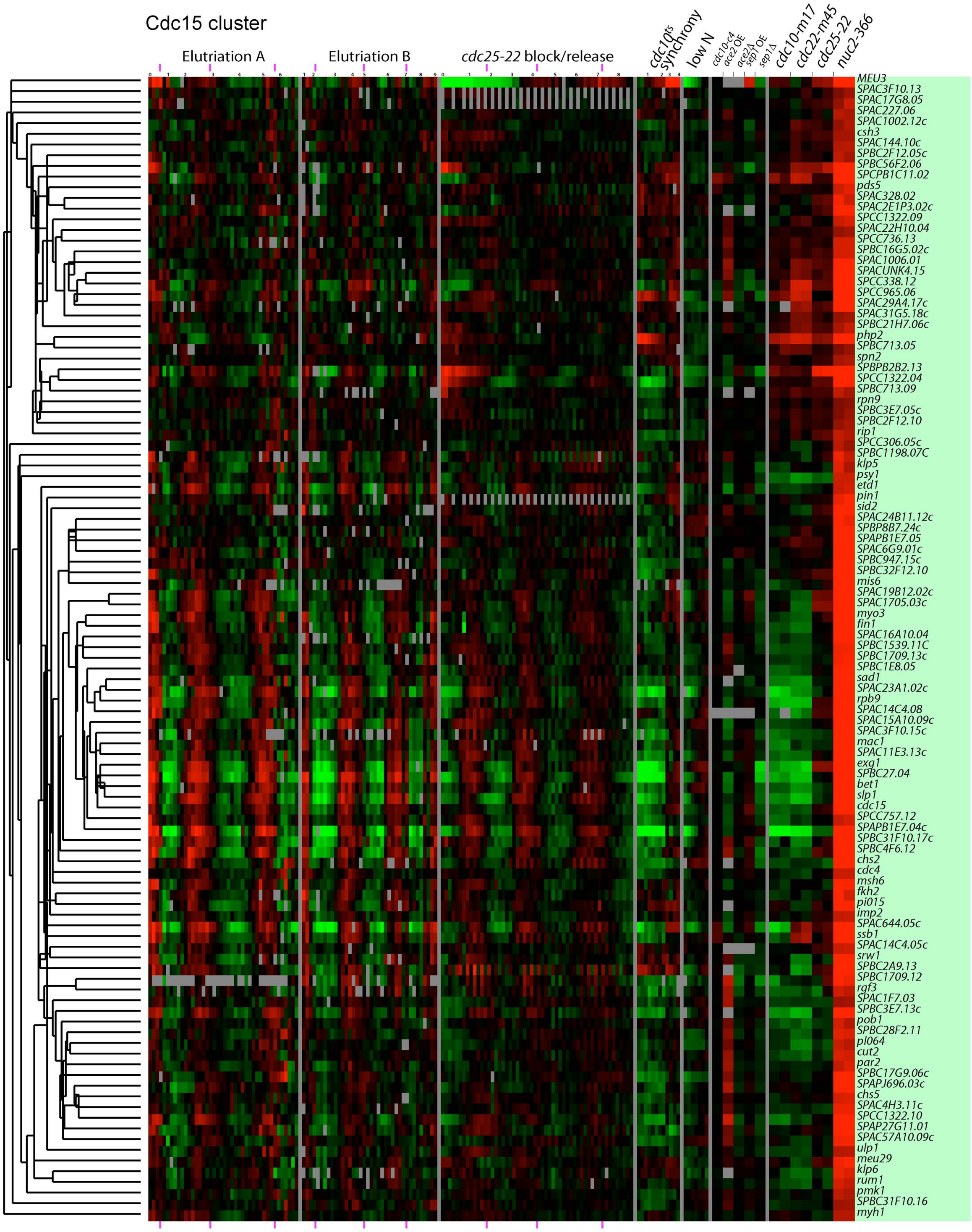


Figure 12: Cdc15 cluster

Clusters of apparently co-regulated genes were chosen from figure 11 on page 58. See legend to figure 11 for further information.

clusters are important for mitosis and cell separation, DNA synthesis, and cytokinesis, respectively. Genes of the Cdc15 and Cdc18 clusters peak almost simultaneously with anaphase (see figures 10 and 14), whereas the Eng1 genes peak slightly later.

The Cdc15 cluster (figure 12 on the preceding page) is the largest of the three clusters and contains over 100 genes. These are involved in mitosis and mitotic exit, cytokinesis and septation, vesicle trafficking, cell wall remodeling, and other functions. Genes involved in mitosis and mitotic exit include the APC adapter subunits srw1 and slt1, the prolyl isomerase pin1, spo12, the Cdk inhibitor rum1, five genes related to ubiquitination, four microtubule-related genes including kinesins klp5 and klp6, and four genes for chromosome segregation.

Cytokinesis/septation fuctions can be ascribed to at least 13 genes including the key SH3 domain gene cdc15 and its paralog imp2, and a third SH3 domain gene, pob1. Also present are the kinases fin1 and sid2 and phosphatase subunit par2, which regulate the septation initiation network. mob1, which interacts with sid2, is also cell cycle regulated with similar timing, but lies outside the cluster as defined here. Other members likely involved in cytokinesis include genes for the rho family member rho4, the putative rhoGEF rgf3, the septin spn2, and the myosin myo3.

Construction of the septum involves synthesis of plasma membrane and deposition of proteins into that membrane. The Cdc15 cluster is rich in proteins involved in these processes. The cluster includes gwt1, likely involved in GPI anchor synthesis, and SPAP27G11.01, SPCC306.05c, and SPBC2F12.05c, linked with sterol functions. SPAC227.06 (a predicted Rab interactor), psy1 and bet1 (SNAREs), and SPBC31F10.16 are likely to function in vesicle transport. The budding yeast homolog of SPBC31F10.16, CHS6, is important for movement of chitin synthase from the trans-Golgi network/endosome to the plasma membrane. Other genes encode cell surface glycoproteins, such as the gene mac1, which is localized at poles and septum and is important for cell separation.

Genes for cell wall metabolism include two chitin synthase homologs, a putative chitin synthase regulator, six putative sugar/starch hydrolases, and the MAP kinase pmk1.

Finally, diverse other functions are represented. There are at least five genes involved in transcription, most notably the transcription factor fkh2, which may be one of the regulators of the Cdc15 cluster (Bulmer et al., 2004) (see below). There are also multiple genes involved in mitochondrial functions and in glycosylation.

The three motif search programs all found the consensus motif $GTAAACAAA$, easily recognizable as a binding site for a forkhead (FKH) transcription factor. Almost every gene in the cluster had such a motif. In *S. cerevisiae*, the main clusters of mitotic genes are also regulated (in part) by forkhead transcription factors. *S. pombe* has several forkhead transcription factors, but the two most likely to regulate the Cdc15 cluster are sep1 and/or fkh2. sep1 does not oscillate noticeably in our dataset, but it does have phenotypes that could be due to defects in the expression of genes of the Cdc15 cluster, and Rustici et al. have shown defects in cell cycle expression in sep1 mutants. fkh2 does oscillate, and is a member of the Cdc15 cluster. The fkh2 promoter contains two sites each for Forkhead, Ace2, and Cdc10. Interestingly, peak expression of fkh2 precedes the peak of 94% of the other genes in the cluster, consistent with the idea that it might help regulate these other genes. No direct binding of either Sep1 or Fkh2 to any of these promoters has been demonstrated, and we believe it is still an open question which protein regulates this cluster. It is possible that both proteins contribute. Because forkhead transcription factors can both repress and activate, and because they are regulated both transcriptionally and post-transcriptionally, the regulatory mechanisms could be complex.

The motif search programs also found $CCAGCC$ (Ace2 binding sites) and $ACGCG$

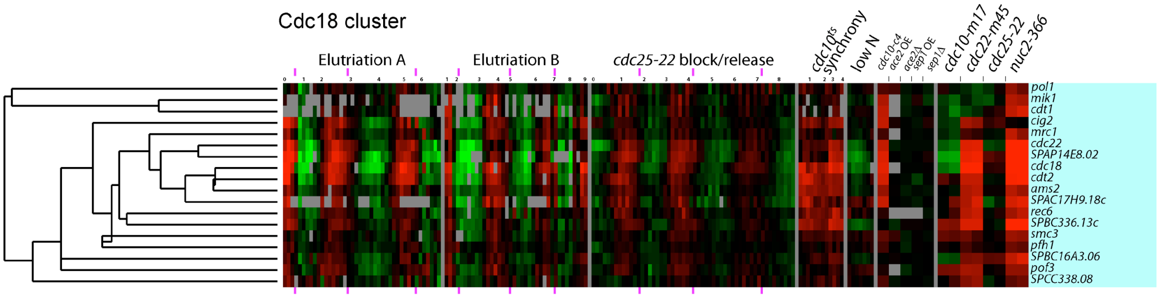


Figure 13: Cdc18 cluster

As figure 12 on page 60, cluster of apparently co-regulated genes were chosen from figure 11 on page 58. See legend to figure 11 for further information.

(MBF/Cdc10 binding sites) in a substantial minority of the genes of the Cdc15 cluster. Many genes (e.g., fkh2 and pds5) had all three kinds of sites. MEME (but not the other programs) also found the motif $\frac{A}{T}TGACAAC$. This is probably the same as the motif $CATG\frac{A}{T}CAAC$ found by Rustici et al. and named "New 1." To minimize confusion, we will refer to our version of the motif as "New1v" ("v" for variant).

MEME also found the motif $CC\frac{T}{A}CC\frac{T}{C}TCC$, and this may be a variant of the motif $\frac{A}{T}ACC\frac{T}{A}CGC\frac{T}{A}$ ("New 3") found by Rustici et al. We will refer to our motif as "New 3v." New 3v was found preferentially in front of genes for cell wall metabolism, such as hydrolases, glycoproteins, chitin synthases, and their regulators. Other functionally related genes are found in the Eng1 cluster (see below), where they appear to be regulated by Ace2. Interestingly, the consensus site for Ace2 ($CCAGCC$) is reminiscent of the core of New 3v ($CCACGC$), suggesting that an unknown Ace2-like factor could be involved.

We did not find the "PCB" consensus ($GCAAC\frac{G}{A}$), previously implicated in the control of some of the genes of this cluster (Anderson et al., 2002; Buck et al., 2004a).

The Cdc18 cluster (figure 13) contains 18 genes involved in DNA replication. Included in the cluster arecdc18 (initiation of DNA synthesis), pol1 (DNA polymerase

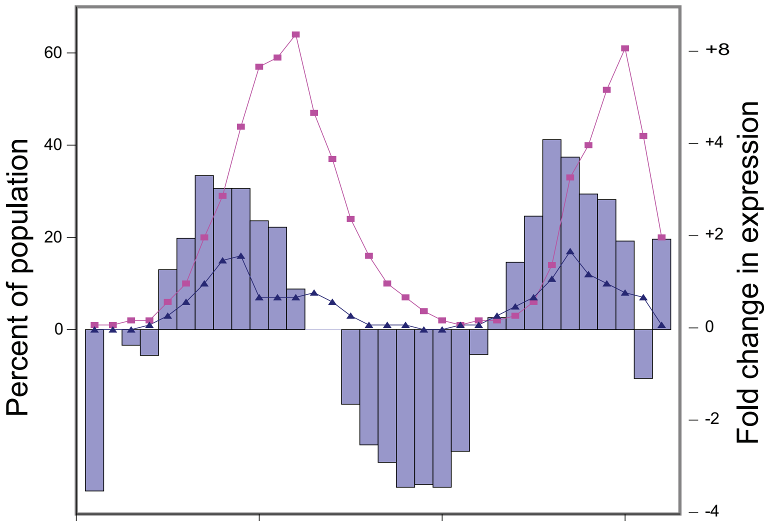


Figure 14: Oscillation of cdc18
The oscillation of the cdc18 transcript through two cell cycles in elutriation A is plotted as a histogram (right y-axis). Also shown are the binucleate (blue triangles) and septation indices (cyan squares) (left y-axis).

alpha), cdt1 (initiation of DNA synthesis), cig2 (S phase cyclin), mrc1 (S phase checkpoint), cdc22 (ribonucleotide reductase), cdt2 (DNA replication), smc3 (cohesin), and pif1 (DNA helicase). These genes are strongly regulated. Peak expression occurs at about the same time as that of the Cdc15 cluster, and is essentially simultaneous with the peak in binucleates (e.g., figure 14).

The Cdc18 cluster has a very similar cluster in *S. cerevisiae*, called the CLN2 cluster. Both clusters contain genes involved in DNA replication, and both clusters appear to be regulated by the MBF transcription factor (see below). For the Cdc18 (*pombe*) and CLN2 (*cerevisiae*) clusters, many of the genes in the clusters are orthologs; e.g., mik1/SWE1, cig2/CLB5, mrc1/MRC1, cdc22/RNR1, andsmc3/SMC3. Thus the cell cycle clusters regulating DNA synthesis are very highly conserved, with the overall function of the clusters, the regulation of the clusters, and the genes in the clusters, all being quite similar from *S. cerevisiae* to *S. pombe*.

The three motif search programs found two motifs in the Cdc18 cluster: $ACGCG$, and $ACGCG\frac{A}{T}CGCG$. The first of these is easily recognizable as the binding site for the MBF transcription factor (also known as DSC1) (Tanaka et al., 1992; Lowndes et al., 1992; Reymond et al., 1992), whereas the second is a related motif that may be a tandem, double binding site for MBF, or for an MBF-like factor. Consistent with the idea that MBF is a major regulator of this cluster, the genes of the cluster are up-regulated by the cdc10-c4 mutation (see figure 13, and see Rustici et al.) which creates a constitutively active form of MBF. Furthermore, six of these genes are known to be regulated by MBF (cig2, cdt1, cdt2, cdc18, cdc22, and mik1; GeneDB, Sanger Centre).

S. cerevisiae has two MBF-like transcription factors. One is itself called MBF and consists of the DNA-binding protein Mbp1 complexed with the modulatory protein Swi6. The second factor is called SBF and consists of a second DNA-binding protein, Swi4, complexed with Swi6. *S. cerevisiae* MBF and SBF, with their related but distinct DNA-binding proteins, bind to related but distinct motifs, and control the cell cycle expression of partially overlapping sets of genes (Koch et al., 1993; Koch and Nasmyth, 1994a). In *S. pombe*, there is likewise one modulatory protein, Cdc10 (the ortholog of Swi6) and two DNA-binding proteins, Res1 and Res2 (possible orthologs of Mbp1 and Swi4) (Tanaka et al., 1992; Lowndes et al., 1992; Reymond et al., 1992; Ayte et al., 1995; Whitehall et al., 1999; Tahara et al., 1998). Some investigators believe that in *S. pombe*, there is a unique MBF transcription factor and that it contains Cdc10, Res1, and Res2 (Ayte et al., 1995; Whitehall et al., 1999; Zhu et al., 1997). However, other investigators believe that the situation is similar to that found in *S. cerevisiae* and that there may be two MBF-like factors, one containing Cdc10 and Res1, and the other containing Cdc10 and Res2 (Tahara et al., 1998; Sturm and Okayama, 1996). Although our results do not speak directly to these models, the fact

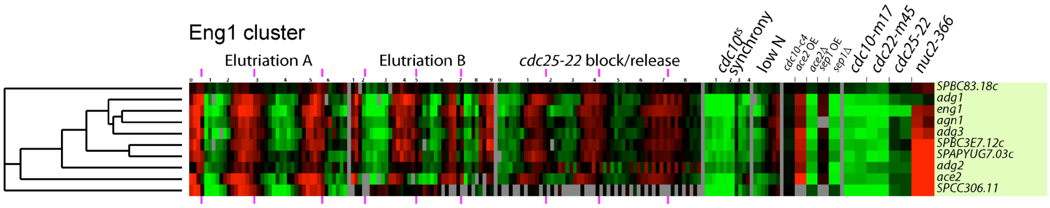


Figure 15: Eng1 cluster
Cluster of apparently co-regulated genes were chosen from figure 11 on page 58. See legend to figure 11 for further information.

that we find two kinds of motifs is easier to interpret in terms of a model with two different but related forms of MBF.

The Eng1 cluster (figure 15) contains nine genes, and these are involved in cell separation. The genes are adg1 and adg2 (cell surface glycoproteins), adg3 (β-glucosidase), agn1 and eng1 (glycosyl hydrolases), cfh4 (chitin synthase regulatory factor), mid2 (an anillin needed for cell division and septin organization), ace2 (a cell cycle transcription factor), and SPCC306.11, a sequence orphan of unknown function. The genes are very strongly cell cycle regulated. Peak expression of most of the genes occurs slightly later than the genes of the Cdc15 and Cdc18 clusters. Motif searches showed that each gene of the cluster has at least one binding site for the Ace2 transcription factor (consensus *CCAGCC*). In fact, eight of the nine genes contain multiple Ace2 binding sites. The exception is the ace2 gene itself, which contains only one Ace2 binding site, but multiple FKH binding sites. Interestingly, the ace2 gene is expressed earlier than the other genes of the cluster, consistent with the idea that it might regulate the other genes. The genes are up-regulated when ace2 is over-expressed, and down-regulated when ace2 is deleted (see figure 15 Rustici et al.). Ace2 was previously shown to be a regulator of eng1 (Martin-Cuadrado et al., 2003) and agn1 (Dekker et al., 2004).

The Eng1 cluster has a recognizably similar functional cluster in *S. cerevisiae*, the

SIC1 cluster (Spellman et al., 1998). This cluster also has many genes involved in cell separation (e.g., EGT2, an endoglucanase; CTS1, an endochitinase; YGL028c, a glucanase; DSE2, a glucanase; and CHS1, a chitin synthase), and the genes of the *S. cerevisiae* cluster are also regulated from Ace2 binding sites of the same consensus sequence (*CCAGC*). However, there is only one gene that is clearly present in the cluster in both species, the glycosyl hydrolase eng1 in *S. pombe*, and its ortholog DSE4 in *S. cerevisiae*. Thus the overall function of the cluster (cell separation), the nature of many of the enzymes in the cluster (carbohydrate hydrolytic), and the mechanism of gene regulation (binding by Ace2) have been conserved, even though the individual genes in the cluster have been largely shuffled. It is easy to understand why the individual genes are different, because the two species have cell walls containing different carbohydrates (and so requiring different hydrolytic enzymes), and because the modes of cell separation are very different (fission vs. budding). In fact, given these differences, it is remarkable that the mode of regulation and the functional cluster seem to have been conserved.

The S/Early G2 Clusters The relatively few genes that peak in late M, S, or early G2 fall into three small clusters: the telomeric cluster, the histone cluster, and the Wos2 cluster (16 on the following page, 17 on the next page, and figure 18 on page 69).

The telomeric cluster figure 16 on the following page contains eight tightly clustered genes found near telomeres. Peak expression is in early S. Two of the genes are at telomere 1L; two at 1R; two at 2L, and two at 2R. Interestingly, *S. cerevisiae* also has a cluster containing only telomeric genes (the Y´ cluster), and the genes of that cluster also peak in late G1 or early S.

The histone cluster (figure 17 on the following page) contains all nine histones of

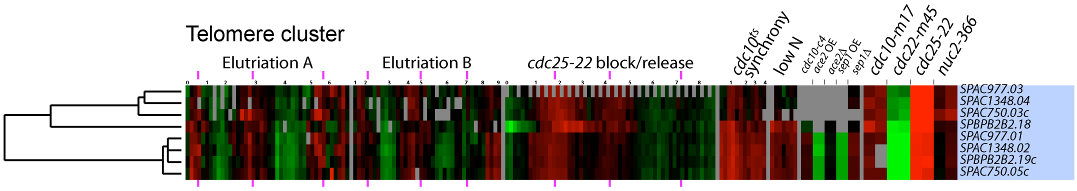


Figure 16: Teleomere maintenance cluster
Cluster of apparently co-regulated genes were chosen from figure 11 on page 58. See legend to figure 11 for further information.

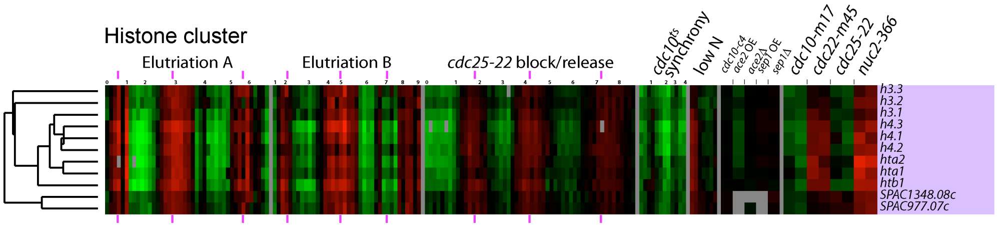


Figure 17: Histone cluster
Cluster of apparently co-regulated genes were chosen from figure 11 on page 58. See legend to figure 11 for further information.

S. pombe. These are tightly co-regulated and strongly periodic, and form a very tight cluster. Presumably, peak expression of the histone genes marks the time of S phase. These genes are expressed about 30 min after the DNA synthesis genes of the Cdc18 cluster. Surprisingly, the histone cluster contains two non-histone genes, SPAC977.07c and SPAC1384.08c. These two genes are near telomeres and are homologs of each other, but have no known function. Possibly they are actually co-regulated with the genes of the telomeric cluster, which peak just before the histone cluster.

Motif searches showed that all the histone genes (but not the two telomeric genes) had the motif $GGGTTAGGGTT\frac{T}{G}$. A degenerate second copy was sometimes also present. This motif has been noted previously (Matsumoto and Yanagida, 1985). In addition, six of the histone genes (and both telomeric genes) had a motif similar to an MBF binding site, $G\frac{C}{G}\frac{T}{G}ACGCG$.

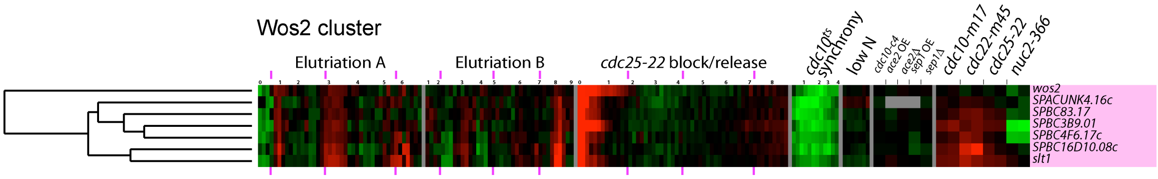


Figure 18: Wos2 cluster
Cluster of apparently co-regulated genes were chosen from figure 11 on page 58. See legend to figure 11 for further information.

In *S. cerevisiae*, the histone genes have at least three semi-redundant regulatory systems: First, they have the HIR gene system that represses histone expression outside of S (Spector and Osley, 1993; Spector et al., 1997). Second, they have regulated mRNA stability, such that the messages are only stable during S (Lycan et al., 1987). Third, they have a system for gene induction during S. Recently, it has been suggested that this positive system relies on the SBF transcription factor, possibly in combination with a forkhead transcription factor (Kato et al., 2004). The fact that an MBF motif is found in front of most of the *S. pombe* histone genes is consistent with the SBF motif found in front of most of the *S. cerevisiae* histones, and suggests that MBF may play a role, along with other mechanisms, in regulating histone expression in *S. pombe*.

The Wos2 cluster (figure 18) contains seven genes expressed in late S or very early G2. Expression of the genes in the cluster responds strongly to the two experiments that involve temperature shifts (cdc25–22 synchrony and cdc10-M17 block-release; note that control cDNA for simple cell cycle–arrest experiments was made from wild-type cells similarly shifted to high temperature). Motif searches found repeats of the sequence $NGAAN$, a typical heat shock response element. The cluster contains wos2, encoding a chaperone activator interacting with Hsp90; SPACUNK4.16c, important in trehalose synthesis (trehalose is a thermo-protectant); SPBC16D10.08c, encoding a

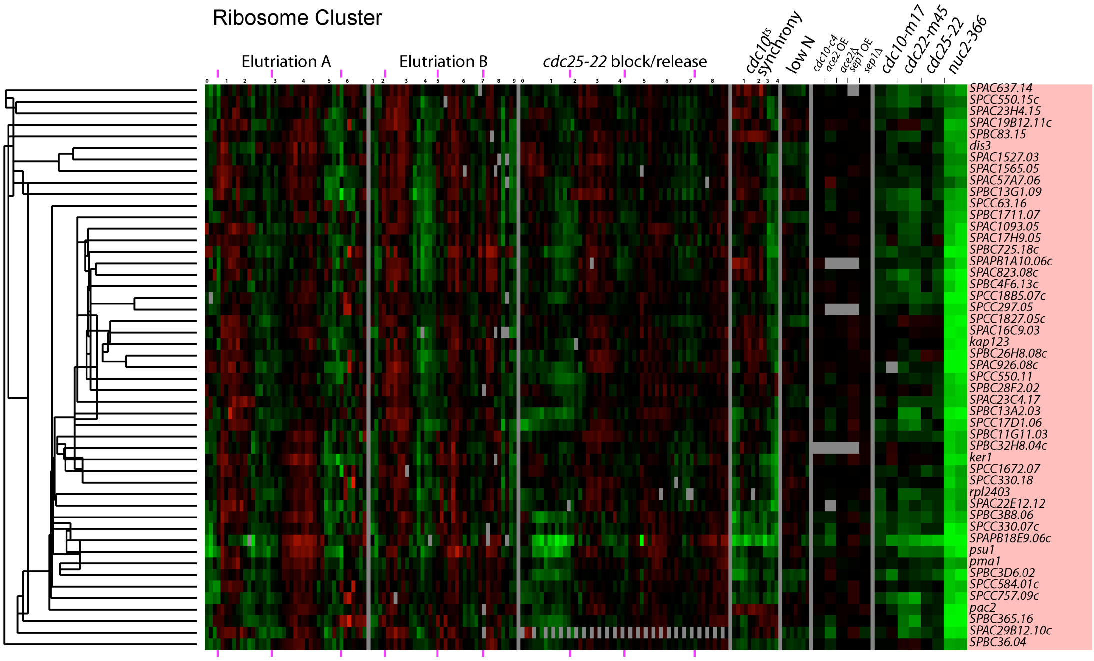


Figure 19: Ribosome biogenesis cluster
Cluster of apparently co-regulated genes were chosen from figure 11 on page 58. See legend to figure 11 for further information.

chaperone similar to *S. cerevisiae* Hsp104; and SPBC4F6.17c, similar to *S. cerevisiae* Hsp78, a mitochondrial chaperone.

The early to mid G2 genes: The Ribosome biogenesis and Cdc2 clusters

Although most of the strongly regulated genes peak near the G2/M transition, another large group of genes, 200 or more, peaks with a moderate amplitude at almost exactly the opposite side of the cell cycle, in early to mid G2 (figure 19). The expression of these genes does not respond to mutations in cdc10, ace2, or forkhead, and they are all strongly repressed at the nuc2 block. Near the center of this set of 200-plus genes is a sub-cluster of genes that is somewhat more tightly co-regulated than the rest. We have designated these the "Ribosomal biogenesis" cluster (figure 19). These genes include SPAC1527.03 (RNA-binding protein, LA-related), SPAC57A7.06 (processome component, involved in rRNA processing), SPBC13G1.09 (bystin family protein, associated with U3 and U14 snoRNAs, involved in rRNA processing); SPCC16A11.02 (WD-repeat protein, processome component, involved in rRNA processing); SPAC23C4.17 (tRNA methyltransferase of the NOL1/NOP2/sun family involved in methylation of cytidine to 5-methyl-cytidine [$\mathrm{m}^5\mathrm{C}$] at several positions in different tRNAs); SPBC11G11.03 (60S acidic ribosomal protein); rpl2403 (60S ribosomal protein L24–3); ker1 (interacts with RNA polymerase I); SPAC1093.05 (DEAD/DEAH box RNA helicase involved in rRNA processing); SPAC926.08c (Brix domain RNA-binding protein involved in ribosome biogenesis and assembly); and many others.

Other genes in the ribosome biogenesis cluster are involved in nuclear/cytoplasmic import and export. These genes include: nup61 (nucleoporin with a RanBp-binding domain), kap123 (karyopherin), SPCC550.11 (RanBP7/importin-beta/Cse1p family, RanGTP-binding protein involved in mRNA export), and mep33 (mRNA export protein).

It is not clear why such genes would be cell cycle regulated. However, Mitchison and colleagues (Mitchison and Nurse, 1985; Creanor and Mitchison, 1982, 1984; Mitchison et al., 1998; Sveiczer et al., 1996) have documented a cell cycle oscillation in the rate of growth and protein synthesis in *S. pombe*. In these studies, there seems to be an acceleration of protein synthesis, and a corresponding acceleration in cell growth rate, in mid G2. Furthermore, "NETO" (new end take off, the time when the new end begins to grow) occurs at about this time. The peak in expression of ribosome biogenesis genes we observe in early/mid G2 could lead to this slightly later peak of protein synthesis and growth rate. Sveiczer et al. (1996) suggest that the acceleration in protein synthesis is the "sizer" that leads to commitment to division; in terms of our findings, the peak in transcription of the ribosome biosynthesis genes would be an important component of the sizer.

We have recently found that many *S. cerevisiae* ribosome biogenesis genes are also cell cycle regulated. Expression peaks in G1, and so this peak could be important for the cell sizer, which in *S. cerevisiae* is in late G1. These genes also show a minor expression peak in early G2. The oscillation of these genes is seen in an elutriation experiment done in ethanol medium, but not in block-release experiments done in glucose medium. The reason for these different, experiment-specific results is unclear, but on the basis of the literature we believe that the oscillation may be under the control of cyclic AMP, and this cyclic AMP signaling does not occur in media with high glucose (Jorgensen et al., 2004; Muller et al., 2003).

Surprisingly, we found no DNA sequence motifs associated with the promoters of the genes in the ribosomal biogenesis cluster.

As one moves out from the center of the ribosome biogenesis cluster, one encounters many other genes peaking in G2 phase. These are of diverse function, but one interesting example is the pma1 gene, which encodes a proton pump. This pump is needed to maintain the proton gradient across the plasma membrane, affecting many processes, and so seems an unlikely candidate for a cell cycle–regulated gene. Nevertheless, it is cell cycle regulated both here and in *S. cerevisiae* (Spellman et al., 1998). The reason for the oscillation is unclear, but because Pma1 is an integral plasma membrane protein that must be inserted into the membrane at the time of synthesis, one possibility is that its synthesis matches the rate of plasma membrane production; in *S. cerevisiae*, this may reach a peak in G2, accounting for the peak in PMA1 transcription. A similar explanation could hold true in *S. pombe*.

Adjacent to the Ribosome biogenesis cluster is a cluster of 23 genes we call the Cdc2 cluster (see figure 20 on the following page). Like the ribosome biogenesis genes, these oscillate moderately with a peak in G2. Their regulation is distinguished

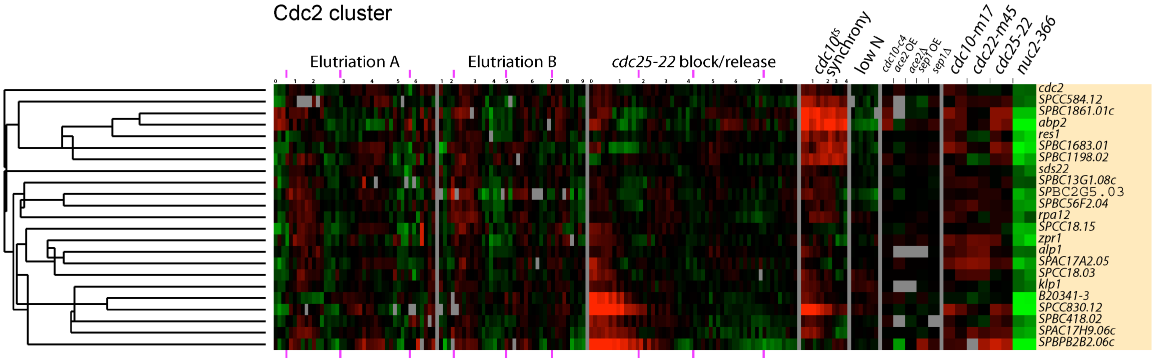
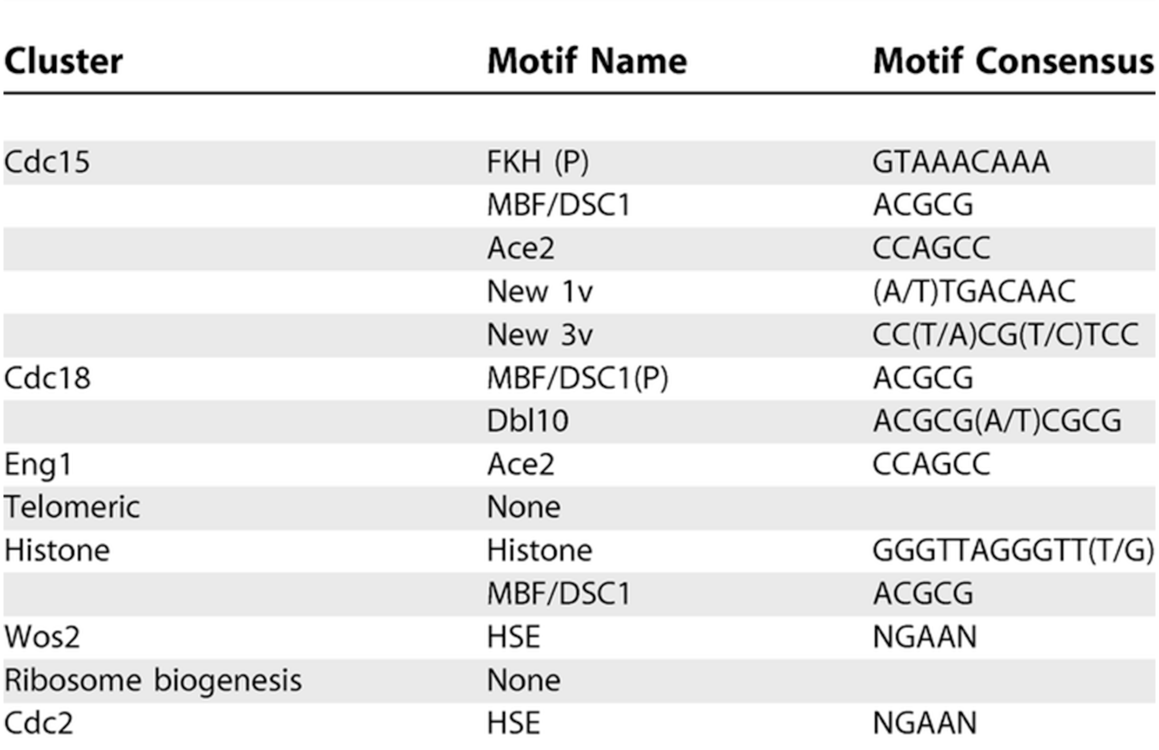


Figure 20: Cdc2 cluster

Cluster of apparently co-regulated genes were chosen from figure 11 on page 58. See legend to figure 11 for further information.

from the ribosome biogenesis genes by the fact that they are differentially regulated after heat stress. Motif search programs found heat shock motifs ($NGAAN$) associated with many of these genes. The Cdc2 cluster contains several interesting cell cycle genes, including cdc2 itself; SPBC1861.01c and abp2, which code for AT hook proteins thought to bind centromeric DNA and ARS DNA, respectively; res1, a key component of the MBF transcription factor; sds22, a protein phosphatase regulatory subunit known to be involved in the cell cycle; ash2, a member of the SET1 complex, and involved in lysine methylation of histone H3; alp1, a tubulin-specific chaperone; SPCC18.03, a putative transcriptional regulator; pkl1, a kinesin-like protein of the Kar3 family; and other genes. Other than the heat shock elements, no statistically significant motifs were associated with the promoters of these genes.

Characterization of Cell Cycle–Regulated Promoters Each cluster was searched for DNA sequence motifs. The most significant motifs are summarized in table 3 on the next page. However, the presence of these motifs in the upstream regions of the genes of a cluster says little about promoter structure. To investigate promoter structure in more detail, we used a program called SpikeChart (S. Pyne, B. Futcher,



Cluster	Motif Name	Motif Consensus
Cdc15	FKH (P)	GTAAACAAA
	MBF/DSC1	ACGCG
	Ace2	CCAGCC
	New 1v	(A/T)TGACAAC
	New 3v	CC(T/A)CG(T/C)TCC
Cdc18	MBF/DSC1(P)	ACGCG
	Dbl10	ACGCG(A/T)CGCG
Eng1	Ace2	CCAGCC
Telomeric	None	
Histone	Histone	GGGTTAGGGTT(T/G)
	MBF/DSC1	ACGCG
Wos2	HSE	NGAAN
Ribosome biogenesis	None	
Cdc2	HSE	NGAAN

Table 3: Cell cycle clusters and motifs

When a cluster contains several motifs, they are listed by decreasing order of statistical significance. When one of several motifs is clearly the most significant, it is noted as the "Primary" or "(P)" motif. The Cdc18 and Eng1 clusters also contain examples of the FKH, New1v, and New3v motifs, but these were not statistically significant in motif searches. The statistical significance of each motif varied widely between programs, as well as with different parameter settings, and is therefore not reported. However, significance rankings (e.g., FKH was more significant than New1v for the Cdc15 cluster) were fairly well preserved across programs and parameter settings. FKH is a forkhead motif and HSE is a heat shock element.

and S. Skiena, unpublished data) that finds and displays motifs in DNA sequences. SpikeChart uses a weight matrix to define a consensus motif, and it shows each occurrence of a motif as a spike of varying height depending on that motif's match to the consensus. For instance, a motif that matches the consensus motif exactly would be given a spike height of ten, whereas a motif with one or more mismatches to the consensus would be given a lower score, depending on the number and nature of the mismatches. (Weight matrices and scoring functions are shown in Table S2). SpikeChart can score many different kinds of motifs simultaneously, and can show the position of all scored motifs, so it is well suited to finding groups of motifs, whether they be of the same kind or different kinds. Initially, because we did not know where regulatory motifs might occur, SpikeChart was used to examine the first 200 base pairs (bp) of the open reading frame in question, and 2,000 bp upstream of the start codon (regardless of whether this region included the next open reading frame or not).

Groups of closely spaced, multiple motifs were usually visible, and these groups usually occurred in the upstream intergenic region (as opposed to within the open reading frame). These groups of motifs were striking for their complexity. There were often four to ten motifs per group, and often the motifs were of several different kinds. The groups of motifs usually occupied about 400 bp of DNA. SpikeChart confirmed that the Cdc15 cluster was dominated by FKH motifs; the Cdc18 cluster was dominated by MBF and DBL10 motifs; and the Eng1 cluster was dominated by Ace2 motifs. However, SpikeChart showed that in addition to the predominant motif, the genes of all these clusters often had other motifs as well. In particular, for M phase genes, it was very common to have at least one FKH motif and at least one other kind of motif. There was a weak to moderate correlation between the number of motifs upstream of a gene and the amplitude of that gene's cell cycle oscillation (data not shown).

We did not notice any cases where the group of regulatory motifs was inside an open reading frame (either the downstream or upstream open reading frame).

In long (>1 kb) intergenic regions, the group of motifs usually occurred within 800 bp of the start codon, but this was not always true; a substantial minority of regulatory motif clusters occurred more than 800 bp upstream (but still within the intergenic region). Because the median *S. pombe* intergenic region is only 900 bp, we wondered whether the cell cycle genes might have unusually long promoters. We measured the length of upstream intergenic regions versus cell cycle rank in our list of all 5,000 genes. The most strongly regulated 200 genes had upstream intergenic regions of about 1,200-bp median length, versus a genome-wide median length of 900 bp. Thus, the more strongly cell cycle–regulated genes have longer than average upstream regions. We have noticed the same phenomenon with the cell cycle regulated genes of *S. cerevisiae* (S. Pyne, S. Skiena, and B. Futcher, unpublished data). The longer-than-average promoters found for cell cycle–regulated genes suggests that these promoters might be above average in complexity.

Discussion

How Many Cell Cycle–Regulated Genes Are There? We have ranked *S. pombe* genes by the statistical significance of their oscillation, and we have discussed the most cyclic 750 genes. However, p-values and other evidence (see figure 10 on page 55) suggest there are at least 2,000 genes with weak oscillations. This number fits well with the combined results of our study and the studies of Rustici et al. and Peng et al.. The three cell cycle lists of 750, 407, and 747 genes, respectively, implicate a total of 1,373 genes. Each study has uniquely identified some genes, but in general these are not just errors, because the vast majority of the uniquely identified genes show some cyclicity in one or both of the other studies even though they do not rise above the cutoff in those studies. Thus we feel the three groups of investigators are each fishing 400 to 750 genes out of a pool of about 2,000 detectably oscillating genes. The three groups are in excellent agreement with respect to the strongly oscillating genes, but then diverge with respect to the more weakly oscillating ones (see figure 9).

However, at the same time, it seems unlikely that 2,000 genes would be directly involved in the cell cycle. There might be at least two kinds of reasons for the observed oscillations. First, an oscillation might be adaptive; i.e., there might be natural selection in favor of the oscillation. The DNA synthesis genes (e.g., cdc18, pol1, and cdc22) in the Cdc18 cluster are examples of genes in which it is easy to believe that the oscillation is adaptive. But second, some oscillations may be incidental. That is, there might be no selective advantage whatsoever to the oscillation, but instead the oscillation is a secondary or indirect effect. For example, chromatin condenses during mitosis. At least in multicellular eukaryotes, mitosis is associated with genome-wide repression of transcription. If there is a similar loss of transcription during mitosis in *S. pombe*, and if our microarray experiments are sufficiently sensitive, we will detect this decreased transcription as a cell cycle oscillation with a trough in mitosis for essentially all genes (preferentially the genes with a short mRNA half-life). But this cell cycle oscillation, though real, does not imply that the oscillation of any of these genes is beneficial; instead, it is a secondary consequence of mitotic repression and chromatin condensation, which presumably is beneficial. Incidental oscillation might also arise when two genes are adjacent to each other. One of the genes might oscillate for adaptive reasons, but the oscillation of this gene might carry over to adjacent genes, for which natural selection is perhaps indifferent to oscillation.

How can we distinguish adaptive from incidental oscillation? First, adaptive oscillations are likely to be large-amplitude oscillations, whereas incidental oscillations are likely to be small-amplitude oscillations. Our cutoff at 750 genes is a crude first screen to enrich for genes with adaptive oscillation. Second, one should consider the total oscillation of the gene's final activity. That is, the oscillation of a gene's transcript might be small. But if one finds that the same gene also has an oscillation in protein stability (e.g., because of regulated proteolysis), and also an oscillation in enzyme activity (e.g., because of phosphorylation), this suggests that the oscillation is adaptively significant. For example, in *S. pombe*, the cyclin transcripts oscillate only modestly, and yet the oscillation of the final product (Cdc2 protein kinase activity) is large. The modest oscillation of the transcript contributes in a significant, multiplicative way to the overall oscillation, and is undoubtedly adaptive. Third, one should consider co-regulated genes and the mode of regulation. If a gene is a member of a small cluster of genes, and the genes have related functions and are regulated by a specific cell cycle transcription factor, then the oscillation is almost certainly adaptive. But if the gene is co-regulated with hundreds of other genes all with very small oscillations, and there is no common function to the genes and no known cell cycle transcription factor, then the oscillation of the whole set of genes may be secondary to some effect such as chromatin condensation. Fourth, one should consider the chromosomal location. Genes adjacent to adaptively regulated genes could oscillate passively. In particular, genes in regions of special chromatin structures (e.g., near telomeres, centromeres, and silenced regions) could oscillate as a secondary consequence of cell cycle changes in the special chromatin structure.

In summary, we feel that a very large number of *S. pombe* genes, 2,000 or more, have at least very small cell cycle oscillations. But it is possible that in many cases this oscillation may be incidental and that only a smaller but unknown number oscillate for adaptive reasons. Sorting adaptive from incidental oscillations will require additional experiments.

Two Genome-Wide Waves of Transcription There were two large waves of transcription, one peaking in early/mid G2, and the other peaking in late G2 or M (see figure 10 on page 55). The early/mid G2 wave contains hundreds of genes, including many genes for ribosome biogenesis. Interestingly, Mitchison and co-workers (Sveiczer et al., 1996) have documented a cell cycle oscillation in protein synthesis, which peaks in mid G2, and may help trigger commitment to cell division. We believe that the early/mid G2 peak in ribosome biogenesis genes may lead to this slightly later peak in protein synthesis.

One property of these early/mid G2 genes is that they are deeply repressed at the nuc2 block in mitosis (see figures 19 on page 70 and 20 on page 73). This is reminiscent of mitotic repression, a phenomenon observed in multicellular eukaryotes in which the majority of transcription (Pol I, Pol II, and Pol III) is repressed during mitosis (Johnson and Holland, 1965; Gottesfeld et al., 1994; Gottesfeld and Forbes, 1997; Hartl et al., 1993; Leresche et al., 1996; Hu et al., 2004; Klein and Grummt, 1999; Spencer et al., 2000; Parsons and Spencer, 1997). Repression is especially well established for Pol I and Pol III polymerases, which are needed for transcription of ribosomal RNA and other RNAs required for protein synthesis. It is thought that highly active transcription may interfere with chromosome condensation, and so transcription is repressed to allow condensation.

A related observation is that in the 1970s and 1980s, metabolic labeling studies were done on synchronized cultures of *S. pombe*. These studies found “steps” of incorporation of labeled uridine into RNA (mostly ribosomal RNA) as a function of cell cycle phase. Around mitosis, incorporation was poor, then after mitosis, the rate of incorporation increased, and then flattened out again at the next mitosis, then increased, and so on. The interpretations of this step-like, cell cycle–regulated uridine incorporation were varied, and the subject disappeared from the literature without resolution (Elliott, 1983a,b; Fraser and Nurse, 1978, 1979).

Putting these observations together, we speculate that *S. pombe*, too, may have some degree of mitotic repression, perhaps important for chromosome condensation. Pol I accounts for the vast majority of the transcription in the cell. Mitotic repression of Pol I transcription of the ribosomal RNA genes would account for the pause in uridine incorporation seen in mitosis in the metabolic labeling studies. But if ribosomal RNA is not transcribed in M, and given that the components of the ribosome are tightly coordinated in their production, then genes for ribosomal proteins (as seen by Peng et al.), and genes for ribosome biogenesis, might also be repressed in M. Repression in M would account for the oscillation of the ribosome biogenesis cluster and its repression at a nuc2 arrest. Finally, if the ribosome biogenesis genes cluster together because they are subject to mitotic repression, this might explain why the cluster does not contain any characteristic 5´ motifs: Mitotic repression might not work through a particular upstream site-specific transcription factor. Indeed, in *S. cerevisiae*, ribosome biogenesis transcripts are controlled in part at the level of mRNA stability (Grigull et al., 2004). Thus, we suggest that *S. pombe* may have a form of mitotic repression and that this repression in mitosis may account for the oscillation of the ribosome biogenesis genes and other genes peaking in early/mid G2 phase and troughing in M.

The second large wave of gene expression peaks in late G2 and in M. This wave includes the Cdc15 cluster (which has many genes for mitosis), the Cdc18 cluster (DNA replication), and the Eng1 cluster (cell separation). There are many important cell cycle events in M and S, and these two phases are close together in rapidly-growing *S. pombe.* The many genes peaking in late G2 and M may simply represent the cell's efforts to prepare for the many activities of M and S. It will be of interest to see what happens to the timing of the Cdc18 cluster (DNA synthesis genes) in slowly growing cells with a long G1: Will they still be transcribed in mitosis, or will they now be transcribed in late G1?

If mitotic repression does exist in *S. pombe*, how is it that the Cdc15, Cdc18, and Eng1 clusters peak in M phase? Baum et al. (1998) have used nuclear run-on to show that cdc18 and some other members of the cdc18 cluster can be actively transcribed in mitosis at a time when histone H1 kinase activity is high and chromatin is presumably condensed. Our own results agree that essentially all the genes of the Cdc15, Cdc18, and Eng1 clusters are highly expressed at a nuc2 arrest, a time at which histone H1 kinase activity is high, and chromatin should be condensed. Our elutriation data suggest that in normal cells, the peak of expression of genes in the Cdc15 and the Cdc18 cluster is almost simultaneous with mitosis (see figures 10 and 12). The genes of these clusters may be specialized for transcription in mitosis. Interestingly, the Cdc15 cluster genes have binding sites for a forkhead transcription factor. Forkhead transcription factors have a "winged-helix" fold, a structure they share with histone H1. Like histone H1, forkhead proteins may be capable of binding to linker DNA in between nucleosomes, and seem to be capable of binding even to chromatin that is relatively condensed (Cirillo et al., 1998; Cirillo and Zaret, 1999; Cirillo et al., 2002; Carlsson and Mahlapuu, 2002). That is, perhaps forkhead is an enabler of transcription for genes in condensed chromatin, and so is particularly suitable for driving expression of genes in mitosis. The Cdc18 cluster depends on the MBF factor, and the MBF/SBF/E2F family of DNA-binding proteins also has a winged helix fold (Taylor et al., 2000). Finally, the Ace2/Swi5 family of transcription factors has been associated with the recruitment of chromatin remodeling enzymes and histone acetylases (Cosma et al., 1999). Even in mammals, which clearly do have mitotic repression, there are mitotic genes strongly expressed during mitosis (Whitfield et al., 2002b). Interestingly, at least some of these genes are thought to be regulated by forkhead transcription factors (Alvarez et al., 2001).

The more moderately expressed genes in the G2/M wave (i.e., genes not in the Cdc15, Cdc18, or Eng1 clusters) tend to be expressed in late G2 rather than in M (see panel C of figure 10 on page 55). Thus these genes may be subject to mitotic repression. Perhaps a large number of genes are expressed in late G2 because it the last chance to be expressed before M, a relatively inopportune time for transcription.

Comparison of Cell Cycle Genes in *S. pombe* and *S. cerevisiae* Of our top 200 ranked cell cycle–regulated genes, 72 (36%) had *S. cerevisiae* homologs that cycled, 68 had *S. cerevisiae* homologs that did not cycle significantly, and 60 did not have clear *S. cerevisiae* homologs. (A detailed comparison of the top 200 *S. pombe* genes and their *S. cerevisiae* homologs is available as Table S3).

Genes involved in core cell cycle processes such as DNA synthesis and mitosis were especially likely to cycle in both organisms. On the other hand, genes involved in budding (in *S. cerevisiae*) or fission (in *S. pombe*), or in cell wall carbohydrate metabolism, generally did not cycle in both organisms for the obvious reasons that the mechanism of cell separation, and the nature of the carbohydrates in the cell wall, are not conserved between the two yeasts.

There are many individual cases where a process is cell cycle–regulated in both organisms, but either the level of regulation (i.e., transcriptional or post-transcriptional) or the identity of the gene regulated varies between the two yeasts. One example is the activity of the cdc2/Cdc28 protein kinase. In *S. cerevisiae*, most of the cyclins are very strongly regulated at the transcriptional level (e.g., CLN1, CLN2, CLB5, CLB6, CLB1, and CLB2), but in *S. pombe*, the equivalent cyclins are only weakly or moderately regulated at the transcriptional level. Possibly compensating for this relatively weak transcriptional regulation, *S. pombe* has very strong post-translational regulation of Cdc2 kinase activity via Wee1/Mik1 inhibitory tyrosine phosphorylation of Cdc2, whereas the homologous system is relatively weak in *S. cerevisiae*. That is, both yeasts strongly regulate cdc2/Cdc28 activity through the cycle, but emphasize different mechanisms. A second example is provided by the gene products of dut1 (SPAC644.05c) and ung1. These proteins both work to exclude uracil from DNA, but by independent mechanisms. The Dut1 protein hydrolyses dUTP, whereas the Ung1 protein removes uracil from DNA by cleaving the glycosidic bond. In *S. cerevisiae*, dut1 is very weakly cell cycle regulated, whereas ung1 is moderately regulated. In *S. pombe*, dut1 (SPAC644.05c) is very strongly regulated, whereas ung1 appears not to be regulated at all. Thus both yeasts use cell cycle transcriptional control to exclude uracil from DNA, but the emphasis is on different genes.

Regulatory Networks and the Late G2 Bump In *S. cerevisiae*, there is a regulatory network governing the transcription of cell cycle genes. This network is organized as a circular cascade, such that transcriptional and post-transcriptional changes occurring during one part of the cycle seem to promote changes in the next part of the cycle, and so on around a circle (Zhu et al., 2000b; Simon et al., 2001; Futcher, 2002). In principle, *S. pombe* must also have a circular cascade of some kind to make the cell cycle repeat. However, fewer cell cycle regulatory mechanisms have been described in *S. pombe* than in *S. cerevisiae*, and so the wiring of the putative cascade is still unclear. In particular, it is unclear how extensive a role is played by transcriptional control.

Moreover, in *S. cerevisiae*, genes displaying large-amplitude cell cycle changes are distributed throughout the cycle (Spellman et al., 1998), consistent with the idea that transcriptional control contributes significantly to all phases of the cascade (Simon et al., 2001). However, in *S. pombe*, most large-amplitude genes are expressed in a window near the G2/M transition, whereas genes of moderate and low amplitudes are distributed throughout the cycle. This concentration of large-amplitude genes near M may suggest that transcriptional control is most important for only some portions of the cascade.

Within the G2/M window of high-amplitude transcriptional regulation, one can discern what may be part of the regulatory wiring diagram. The transcription factor gene fkh2 peaks in the earliest part of the late G2 window. Over 100 other genes in this window, including fkh2 itself, have FKH binding sites, so the up-regulation of fkh2 may contribute to this large wave of gene expression.

One of the critical targets of the Fkh transcription factor may be the gene for the Ace2 transcription factor. The ace2 promoter has multiple sites for Fkh binding. The ace2 promoter also has one site for Ace2, so, like the fkh2 gene, ace2 may be autoregulatory. The Ace2 transcription factor then induces a cluster of genes involved in cell separation and cell wall metabolism. Interestingly, a forkhead transcription factor is involved in turning on the ACE2 gene in *S. cerevisiae*, so this particular part of the cell cycle wiring diagram appears to be conserved in the two species.

Three of the major cell cycle transcription factors in *S. cerevisiae*, MBF/SBF, Fkh, and Ace2/Swi5, have homologous cell cycle transcription factors in *S. pombe*. The major exception is Mcm1, a MADS-box transcription factor. In *S. cerevisiae*, there are two paralogs of this gene, MCM1, and ARG80. Mcm1 is a transcription factor for cell cycle genes and mating genes, whereas Arg80 controls various metabolic processes. The best *S. pombe* orthologs are Map1 and Mbx1 (Buck et al., 2004a). There was no noticeable enrichment of an Mcm1-like binding motif in front of any cluster of cell cycle–regulated genes; i.e., there was no evidence for a binding site for Map1 or Mbx1.

In multicellular animals, the major well-characterized cell cycle transcription factor(s) are those of the E2F/DP family (Attwooll et al., 2004; Bracken et al., 2004). These typically control a cluster of genes expressed in late G1, and the genes are involved in DNA replication and commitment to the cell cycle. Functionally, the genes controlled by E2F/DP in animals are similar to the genes controlled by MBF in the two yeasts. E2F and DP proteins are not very similar in sequence to the proteins found in MBF, but it is also true than various E2F and DP proteins are not very similar to each other, though they are clearly related. E2F and DP recognize binding sites with a *CGCG* core, as does MBF. Furthermore, the DNA-binding domain of E2F/DP factors consists of a winged-helix fold (Zheng et al., 1999), as do the DNA-binding domains of Swi4 and Mbp1 (components of *S. cerevisiae* SBF and MBF, respectively) (Taylor et al., 2000; Zheng et al., 1999). Thus, despite the overall sequence dissimilarity, it is possible that MBF in the yeasts, and E2F/DP in animals, are cell cycle transcription factors that are related by descent and which have always controlled the cell cycle expression of genes involved in DNA replication.

Materials and Methods

Microarrays Microarrays were made by spotting unmodified, double-stranded PCR products onto glass slides coated with aminopropylsilane (Erie Scientific). Spotting was done using a robot of the DeRisi design[14] and ArrayMaker2 software[15]. PCR primers were designed using Primer3 (Whitehead Institute, Cambridge, Massachusetts, United States) and a shell script. Primers were designed against approximately 5,000 open reading frames and RNAs (excluding pseudogenes) as annotated by the Sanger Centre[16]. In general, PCR primer pairs were designed to give products 500 to 1,000 bp in length, because the yield of the PCR reaction decreased for

[14] http://cmgm.stanford.edu/pbrown/mguide/

[15] http://derisilab.ucsf.edu/arraymaker.shtml

[16] http://www.sanger.ac.uk/Projects/S_pombe/DNA_download.shtml

products longer than 1,000 bp. When the PCR product was small compared to the length of the gene, it was usually chosen from the 3´ region of the gene, so as to maximize representation in poly dT-primed cDNA synthesis. PCR products were amplified from genomic *S. pombe* DNA, and so in some cases the final product included introns, but the design parameters maximize contiguous exonic sequence. A fuller description of the microarrays will be published elsewhere. A full description of the primer pairs, and hence the features on the microarrays, can be found at http://www.redgreengene.com.

Cell cycle synchronizations Two methods of cell cycle synchronization were used, elutriation and a cdc25–22 block and release. Two independent elutriation experiments were carried out. For elutriations, 8l of h-972 cells (wild-type) were grown in YES (autoclaved, elutriation B or filter-sterilized, elutriation A) to early log phase (OD_{600} = 0.4) at room temperature (25°C). Cells were harvested by centrifugation, resuspended in approximately 100-ml YES, and sonicated, all at room temperature. For elutriation B, approximately half of the cell volume was reserved for the reference cDNA preparation. For elutriation A, the reference cDNA was prepared independently and the entire sample was used for elutriation. Cells were loaded into a Beckman elutriator rotor containing two 40-ml elutriation chambers connected in series. When two chambers are used in series, the bulk of the cells remain in the first chamber, but the smallest cells flow into the second chamber, and then, at higher pump speeds, some of these flow out of the elutriator for collection. This arrangement provides both high capacity and high resolution. The elutriator was used at 1,800 rpm at room temperature. After every increase in pump speed, a fraction of about 150 ml was collected, containing about 5×10^8 cells (elutriation B) or 3×10^9 cells (elutriation A). These were diluted to OD_{600} = 0.2–0.05 (greater dilution for samples harvested at late times) with conditioned (elutriation B) or fresh filter-sterilized (elutriation A) medium, and then sampled with time.

An entire three cell cycle time course was obtained from five elutriator fractions (elutriation B) or two fractions (elutriation A). We used adjacent fractions containing no ($< 0.5\%$) septated cells; the elutriator fraction with the largest cells (i.e., the last fraction collected) was used first, then the elutriator fraction with the next largest cells, and finally the elutriator fraction with the smallest cells (i.e., the first fraction collected). In general, the fractions were "overlapped," i.e., the last sample from one fraction and the first sample from the next fraction were collected at the same time. "Overlapped" fractions, though collected at the same time, were deemed to have been collected at slightly different times; the number of minutes by which overlapped fractions were offset was determined by the offset, in minutes, of the septation indicies for the two fractions. That is, for any pair of overlapped fractions, the smaller cells were deemed to have been collected earlier, by a time determined from the offset of the septation indicies of the two fractions. Note that elutration A used only two fractions, and so there was only one overlap. Samples were taken about 10 min. (elutriation A) or 15 min (elutriation B) apart; exact sampling times are given in the Treeview files 1, 2, and 3 (Dataset S1) and at http://www.redgreengene.com.

Cells (10^8 cells/sample) were harvested by centrifugation at 4°C and washed with ice-cold water, snap frozen, and stored at -70°C. For elutriation A, an equal volume of ice was added to the cell culture during harvest (harvest with ice). The reference sample for hybridizations was sonicated cells prior to elutriation (elutriation B), or h-972 grown to $\mathrm{OD}_{600} = 0.2$ in filtered YES at 25°C (elutriation A). Septation index was monitored by phase contrast microscopy of live cells during each experiment. In addition, frozen cell pellets were thawed and stained with DAPI and calcofluor to monitor anaphase ("binucleates") and septation for elutriation A. Cells were scored as "binucleates" if two nuclei were visible, but there was no septum.

For the cdc25–22 block release, the prototrophic strain JLP1164 h+ cdc25–22 was grown in filtered YES at 25°C to OD_{600} = 0.4 and then used to inoculate 4 × 500 ml filtered YES to an OD_{600} of 0.1 (flask 1), 0.08 (flask 2), 0.07 (flask 3), and 0.05 (flask 4). Cells were shifted to a water bath at 36.5°C for 4 h to arrest them in G2 (time = 0 h) and then shifted back to 25°C rapidly in an ice-water bath (26°C was achieved in approximately 5 min; cultures did not cool below 25°C). Samples were taken 10 min apart and harvested with ice as described above. The reference sample for hybridizations was JLP1164 h + cdc25–22 grown at 25°C to OD_{600} = 0.2 in filtered YES. Septation index was monitored by phase contrast microscopy.

Other microarray experiments To examine cells released synchronously from a cdc10 arrest, 8L of strain JLP1166 h− cdc10-M17 was grown at 25°C to OD_{600} = 0.5 in filtered YES, and then harvested and elutriated to obtain a fraction of G2 cells. These were diluted to 10^6 cells/ml, shifted to 36.5°C for 3 h 15 min, rapidly cooled to 25°C as described above (time = 0), and then sampled with time. Cells were harvested with ice. Samples were also collected and analyzed by flow cytometry to monitor DNA replication. The reference sample for hybridizations was JLP1166 h− cdc10-M17 grown to OD_{600} 0.2 in YES at 25°C.

To examine cells grown in low nitrogen, wild-type h-972 was grown in EMM lacking NH4 and supplemented with 20 mM phenylalanine (EMM-phe) to provide a limiting nitrogen source to expand the G1 window (Carlson et al., 1999). 8L of cells were grown at 25°C to OD_{600} = 0.4, collected by centrifugation at 4°C and kept on ice and sonicated on ice. Approximately half of the total cell volume (125 ml, total 2 × 10^8 cells) was reserved for reference cDNA synthesis and the remainder was elutriated at 4°C to fractionate the culture into 21 fractions ranging from small cells (50% G1) and then medium cells (G2) and finally to long, septated cells. Fractions were harvested immediately by centrifugation at 4°C. Fraction assignments were confirmed by flow cytometry analysis and high-quality hybridizations were obtained with fractions 2, 3, 5, 7, 10, 13, and 16.

To examine cells arrested at the cdc10, cdc22, cdc25, and nuc2 block points, four strains carrying these cell cycle mutants (cdc22-M45, nuc2–663, cdc25–22, and cdc10-M17) and a wild-type reference control were grown to OD_{600} 0.05–0.08 in YES at 25 °C and shifted to 36.5°C. After 4 h of arrest at this restrictive temperature, a sample was taken for microarray analysis. For each strain, the experiment was repeated with an independent single colony. Figures 10 on page 55, 11 on page 58, 16 on page 68, 17 on page 68, and 18 on page 69 show results (in different columns) from both single colonies. Strains used were wild-type PR109 h− leu1–32 ura4-D18 (obtained from P. Russell), and the cell cycle mutants (F84) OM591 h− cdc22-M45 (P. Russell), (F58) PR580 h− leu1–32 nuc2–663 (P. Russell), JLP1165 h+ cdc25–22 (this study), and JLP1166 h− cdc10-M17 (this study).

Microarray hybridization and processing Cell samples for RNA isolation were rapidly cooled by addition of an equal volume of ice (except for elutriation B in which samples were placed on ice) and then collected by centrifugation at 4,000 rpm. (3,300 × g) at 4°C for 3 min. Pellets were washed twice in ice-cold dH_2O, frozen in liquid nitrogen, and stored at −70°C. Total RNA was isolated using RiboPure Yeast (Ambion, Austin, Texas, United States) according to the manufacturer's instructions (elutriation A samples) or hot phenol essentially as described in Rustici et al. (2004)[17] with slight modifications according to the detailed protocols at http://www.redgreengene.com). Isolated RNA was further purified by RNAeasy

[17] http://www.sanger.ac.uk/PostGenomics/S_pombe/docs/rnaextraction_website.pdf

cleanup columns (Qiagen, Valencia, California, United States) and quantitated by absorption spectroscopy.

Microarray probes were prepared in two steps. First, cDNA was synthesized incorporating aminoallyl-dUTP (aa-dUTP). Purified aadUTP cDNA was then coupled with Cy3 or Cy5 fluorescent dyes according to protocols from the Institute for Genomic Research[18] with slight changes as follows: 20–25 µg of total RNA was used for cDNA synthesis with 4 µg of oligo-dT primer (not random hexamers), and reactions contained 300 µM aminoallyl-dUTP with 200 µM dTTP. RNA was destroyed using RNase instead of NaOH, and reactions were purified with a Qiagen PCR purification kit. Dye incorporation was determined by absorption spectra and was typically one fluor/20–30 nucleotides.

For hybridizations, cDNA with 50pmol Cy3 plus reference cDNA with 50pmol Cy5 was included in a 24µl total hybridization solution (25% v/v formamide, 5× SSC, 0.1% SDS, and 100µg/ml of sonicated salmon sperm DNA). Hybridizations were performed under 22 × 25mm lifter cover slips (Erie Scientific, Portsmouth, New Hampshire, United States) at 50°C in a humidified chamber for 16–20h. Hybridized arrays were washed by gently shaking as follows: twice briefly with 2× SSC/0.1% SDS (50°C), twice for 10 min with 2× SSC/0.1% SDS (50°C), and four times briefly with 0.1× SSC at room temperature. Arrays were dried by centrifugation.

Arrays were scanned using an Axon 4000B scanner, controlled by GenePix Pro 5.1 software with a pixel size of 5 µm and two-pass sequential line averaging. Laser power was set to 100%, and PMT gains were subjectively adjusted during prescan to maximize effective dynamic range and to limit image saturation. Lossless image files were stored for later analysis.

[18] http://www.tigr.org/tdb/microarray/protocolsTIGR.shtml

Data extraction and storage To extract data from microarray scans, previously stored image files representing all hybridizations were analyzed in parallel. Spot size, location, and quality were determined automatically by GenePix Pro algorithms. Dynamic spot resizing between 60 and 150 µm diameter was permitted based upon image examination and prior optimization. Misidentification of spot locations was corrected by manual adjustment of the map prior to automatic sizing and shifting. Only in cases of gross hybridization defect were spots/regions manually moved/resized or flags modified to "bad," permitting consistent spot calling. Following spot location, parameters and values for each spot were calculated by GenePix Pro and exported. No normalization was applied within GenePix Pro.

Raw data and images exported from GenePix Pro were used to populate a local installation of the Longhorn Array Database (Peter Killion, University of Texas at Austin. An SQL database based upon the Stanford Microarray Database, Stanford University). Initial data normalization was performed at the time of population. Briefly, spots were categorized as "*pombe*" or "other." *S. pombe* spots were further categorized into "normalization" (no bad, missing, absent, or not-found flags) or "non-normalization" (bad, missing, absent, or not-found flags). Only normalization spots were further considered for the normalization calculation. Finally, spots with greater than 5% saturation in either image channel were discarded from this group. The mean log_2 ratio of the median net intensities (Rm, foreground pixel − median of the local background) was calculated. This "normalization factor" represented the distance from a red/green ratio of one, and was used as a scalar modifier for the ratios of all spots in the hybridization; i.e., though only spots meeting a stringent "good" criterion were used to determine the normalization value, this value was subsequently applied to all spots, good or not. During retrieval of data, several further criteria were used to ensure high-quality data in downstream analysis. Only spots with non-negative flag values were retrieved (not bad, missing, absent, or not-found), and only spots with a regression correlation of pixel ratios (a metric of internal spot consistency) greater than 0.6 were used. Spot values were averaged (mean) when multiple independent spots representing a single PCR product were present as internal controls or otherwise. When analyzing multiple hybridizations, such as during time-course analysis, iterative gene and array centering was performed. Briefly, within an array of genes × arrays, the mean log_2 ratio of medians (R_m) was calculated and subtracted from each log_2 R_m, first along the gene axis, then along the array axis, until subsequent iterations varied by $< 0.001\%$.

Normalization There are some special red/green normalization issues relevant to the genome-wide waves of expression (see figure 10 on page 55). In general, fluorescence from mRNA from synchronous cells was normalized to total fluorescence from mRNA from asynchronous cells. If all mRNAs were equally reduced in abundance during mitosis (e.g., because of genome-wide mitotic repression), this normalization would obscure the effect. However, despite normalization, any such repression would still be somewhat visible as a relative loss of unstable mRNAs versus stable mRNAs. We presume that if the trough of peaks of expression in M/G1/S seen in figure 10 on page 55is indeed partially due to mitotic repression of transcription, then the individual genes troughing at this time are genes that produce unstable mRNAs, which are then noticeably repressed despite the red/green normalization because they are reduced relative to stable mRNAs. This argument suggests that stable mRNAs should tend to peak in the M/G1 interval. Indeed, some genes do peak at this time. These, however, would be fewer in number than the genes that fail to peak, because, in general, cells tend to express a small number of stable mRNAs to very high levels, comprising the bulk of mRNA, and a large number of unstable mRNAs to low levels

(Futcher et al., 1999).

A second normalization issue is that the oscillation of the strongly regulated genes would have an effect, via normalization, on the apparent expression of non-oscillating genes (i.e., genes that do oscillate would produce an artifactual, complementary oscillation in non-oscillating genes, via normalization). To side-step this artifact, pixel intensity data for the bottom 4,000 genes were extracted from the microarray data before the red/green normalization step, and then normalized and analyzed after extraction, so that the oscillation of the 1,000 most strongly cyclic genes would not interfere with normalization of the least cyclic genes.

Cluster analysis For cluster analysis, array- and gene- centered log_2 Rm data were hierarchically clustered along the gene axis by the agglomerative algorithm of Eisen et al. (1998). Data were visually presented using JavaTreeView[19]. Separation of the total dendrogram into subordinate clusters was performed subjectively.

Motif analysis To find DNA sequence motifs, nucleotides extending from -1 to the edge of the most 3´ proximal gene (stop or ATG, depending on orientation) with a maximum length of 12,000 bp were extracted genewise for each cluster and used as a target set for motif searching. Three different motif search programs were used: MEME, AlignAce, and SPEXS.

MEME (Multiple EM for Motif Elicitation) (Bailey and Elkan, 1995) was used to find motifs between five and nine nucleotides long present in any number of copies on either strand, weighted to find 1/3n to 3n total sites in the target set of n sequences. Parameters were set as follows: `$ meme (sequence name) -dna -minsites (n/3) -maxsites (n*3) -mod anr -minw 5 -maxw 9 -revcomp -nmotifs 10 -evt 0.1 -bfile (fifth order Markov model).` (Other param-

[19] http://jtreeview.sf.net, manual at http://jtreeview.sourceforge.net/manual.html

eters were also tried in additional searches.) The top ten motifs exceeding an E-value of 0.1 were generated using a background set consisting of the fifth order Markov model representing possible nucleotide pentuplets in all *S. pombe* upstream regions.

AlignACE (Hughes et al., 2000) uses a Gibbs sampling algorithm. Again, all *S. pombe* upstream regions were used as a background set.

SPEXS (Sequence Pattern EXhaustive Search) (Vilo et al., 2000), a word-search enumeration algorithm, was also used. Relative frequency of 1- to 9-mers was calculated, and compared between the target set and all *S. pombe* upstream sequences.

Identification of oscillating transcripts In general, identification of oscillating transcripts requires a method for finding oscillations in each experiment, and then a method for combining the results from different experiments. Here, we have used Fourier analysis to identify oscillating genes. A p-value for the hypothesis of oscillation was then established using Monte Carlo simulations on shuffled data. The p-values for different experiments were then combined using known statistical properties of the p-value. Finally, the p-values for each gene were ranked. Although we have ranked the p-values, these p-values are nevertheless closely correlated with the amplitude of the oscillation.

For each time series of observations in a single time course (e.g., a three cell cycle elutriation experiment), we calculated the Fourier sums A and B over the range of times, t, in the experiment:

$A = \sum sin(2\pi(\frac{t}{T})) * log_2(ratio(t))$

$B = \sum cos(2\pi(\frac{t}{T})) * log_2(ratio(t))$

Here, t is the time in minutes at which the sample was taken (where the beginning of sampling is zero time); T is the cell cycle period, i.e., the time in minutes required for a complete cell cycle; and ratio(t) is the ratio of experimental to control signal at time t.

We considered these two sums as a vector $C = (A, B)$, and then calculated the magnitude of the vector, $D_0 = \sqrt{(A^2 + B^2)}$. This magnitude, D_0, is our basic Fourier measure of whether a transcript oscillates. Note that there is no need to calculate phase.

However, random noise would generate some value of D greater than zero, and genes whose transcripts are relatively variable in abundance could generate relatively large values of D, even if these variations had no connection to the cell cycle. Therefore, as a second step, we randomly shuffled the series of observations for each gene in question, and calculated a new magnitude, D_R, for the randomized series. This randomization was repeated 1,000 times, generating 1,000 values of D_R. These represent the distribution of D for each gene, given that gene's actual variance in gene expression. Finally, we compared the original value of D_O from the unshuffled data to the distribution of D from the shuffled data, found how many standard deviations D_0 is from the mean of the distribution, and in this way calculated a z score for D_0.

This procedure was repeated for each gene and for each experiment. Thus, for each gene, there were three z scores, one per experiment (two elutriation experiments and the cdc25 block-release experiment). These three z scores were then combined by the method of Stouffer, yielding a single p-value for each gene. Genes were then ranked by p-value with the lowest p-value at the top of the list. In practice, a large amplitude of oscillation contributes tremendously to a low p-value, so the upper portion of the p-value list is almost exclusively occupied by genes with high-amplitude oscillations

Gene database In general, we have used the information in the GeneDB database[20] to describe the various genes studied; when a fact is given in the text about some gene

[20] http://www.genedb.org/genedb/pombe/index.jsp

but no reference is given, the information comes from GeneDB. When the primary literature has been consulted directly, the reference for the primary literature is given.

3 Abundant non-coding RNAs affect transcriptome structure in *S. pombe*

Yeast biologists were relatively slow to undertake whole-transcriptome profiling. As detailed above, systematic analyses of human and murine transcriptomes using high resolution arrays and EST/tag sequencing were first published in 2002 (Kapranov et al., 2002) and 2006 (Carninci et al., 2005) , respectively. These studies clearly and repeatedly demonstrated that the genomic "dark matter" of gene-free regions was being actively transcribed. Coding density of higher eukaryotic genomes is quite low, so it is perhaps unsurprising that exonic sequences made up only a fraction of processed cellular RNAs. The genomes of budding and fission yeast are more than an order of magnitude more gene dense than those of metazoa, making it unclear how findings in mouse and man would translate to smaller, more dense genomes.

Several instances of non-coding transcription had previously been identified in budding and fission yeast. In *S. pombe,* subtractive hybridization and cloning of meiosis-specific transcripts yielded several species which lacked an open reading frame, including structured RNAs and antisense transcripts (Watanabe et al., 2001). The presence of abundant processed non-coding transcripts in vegetative cells was alluded to shortly thereafter; nearly 7% of sequenced cDNA clones from mitotic cells lacked an identifiable ORF (Watanabe et al., 2002). Prior to high-resolution studies described below, numerous non-coding transcripts were known in *S. cerevisiae*, including several that conferred regulatory functions (outlined above on page 20). EST data had been generated for budding yeast and was invaluable for validation of predicted transcripts, but high genomic density made identification of non-coding species difficult (Velculescu et al., 1997). Despite the rich variety of gene expression studies which had been performed in yeast, transcriptome structure was not mapped until relatively "late" in comparison to other model organisms.

3.1 Transcriptome mapping and a glimpse of RNA-seq

In 2006, several tools developed for human transcriptome analysis trickled down to budding yeast. Ron Davis's group published a high-resolution snapshot of transcription in asynchronous *S. cerevisiae* using Affymetrix tiling arrays (David et al., 2006), providing strand-specific quantitation of transcript levels at 8nt spatial resolution. Novel computational tools were designed for analysis of these data, including an algorithm for segmentation of genomic intervals based on similarity of probe signals (Huber et al., 2006). This approach provided a list of genomic regions whose edges were based on differences in expression level. Several months later, the Ito lab published its long-read sequencing snapshot of mitotic and meiotic transcription in budding yeast (Miura et al., 2006). These data included more than 50,000 full length cDNA sequences from two vastly different transcriptional programs, and provided both validation of and complementary information to the tiling array data (although the two papers have not been directly meta-analyzed). A critical aspect of the Ito lab's sequencing approach was the use of vector-capped libraries (Ohtake et al., 2004; Kato et al., 2005), ensuring that the cDNAs to be sequenced were full length, complementary up to and including the 5'- m^{7}G cap. These direct sequencing data permitted accurate identification of transcription start and polyadenylation sites. The data resulting from both of these studies demonstrated impressive complexity of the budding yeast transcriptome, and permitted the genome-wide analysis of 5' and 3' untranslated regions, validation of predicted splicing events, and the identification of novel transcripts, including many antisense species. While somewhat limited in biological breadth, these dense data provided fertile ground for analysis of transcript structure and novel feature discovery.

Previously unappreciated antisense transcription was a common finding of both David et al. (2006) and Miura et al. (2006). This was especially true for data derived from tiling arrays, where more than 1000 previously unannotated antisense transcripts were identified, including nearly 200 "high confidence" "filtered" species. Several of these antisense transcripts were validated through further experiments, including a case of regulatory transcriptional interference which was shown to govern meiotic entry (discussed here on page 20 and in Hongay et al. (2006); Uhler et al. (2007b)). However, many of the antisense transcripts reported in David et al. could not be validated by other methods, often to the frustration of researchers whose "favorite genes" were purported to have antisense counterparts.

Before the publication of David et al. (2006); Miura et al. (2006), I had personally purchased and implemented a complete Affymetrix workflow in the lab. Affymetrix tiling arrays for budding and fission yeast had been previously ordered for use in ChIP/chip experiments, but required expensive and inconvenient processing at the University DNA Microarray Facility (UDMF) (University Hospital, Stony Brook NY). The availability of in-house hybridization and processing permitted flexibility in chemistries used, sample preparation times, and most significantly, reduced cost. Proof of principle for both the wet biology (probe labeling and hybridization/staining/scanning) and downstream data analysis were performed by labeling genomic DNA from wild type and partially aneuploid (harboring Ch^{16}, (Niwa et al., 1986)) cells. This strain contains a centromeric minichromosome with ~500kb of

Chromosome 3, and had been my previous test subject of choice for validating relative copy number on our spotted arrays prior to replication timing work for Mickle et al. (2007). The data generated from the Affymetrix platform were immense in comparison to spotted array data; significant time and coding effort was put into the processing of raw intensity data from the array into mapped positions along the *S. pombe* genome (several of these issues are addressed above on page 36). Despite the growing pains of dealing with new higher density data sets, I was able to determine a two-fold change across the aneuploid region with high confidence.

After seeing the rich complexity of *S. cerevisiae* transcription and the quality of data generated from Affymetrix arrays, I decided undertake a similar analysis in *S. pombe*. Collaborating with Huei-Mei Chen, another graduate student in the lab, I set out to map transcription in fission yeast, with the main goals of refining transcript structure, especially 5' and 3' ends, and identifying/validating introns, with particular interest in regulated splicing events. To our good fortune, troubleshooting of cDNA labeling for tiling arrays took a bit longer to scale up than expected; we were able to generate and hybridize only a small number samples over the first few months. During that time, the Steinmetz lab published a follow-up to David et al. (2006) which addressed the presence of widespread (but largely unvalidateable) antisense transcription in budding yeast, and provided a simple change to the labeling protocol which greatly improved our eventual data.

The cDNA synthesis and probe labeling protocol used in David et al.was exceedingly simple (and is the fundamental basis for our protocol described below on page 125). MMLV-derived reverse transcriptase was used to generate cDNA from total RNA using random hexamers. Following hydrolysis of RNA and cleanup by phenol extraction and EtOH precipitation, the cDNAs were enzymatically sheared by brief DNAse I digestion into fragments of 50-100nt. Fragments were subsequently

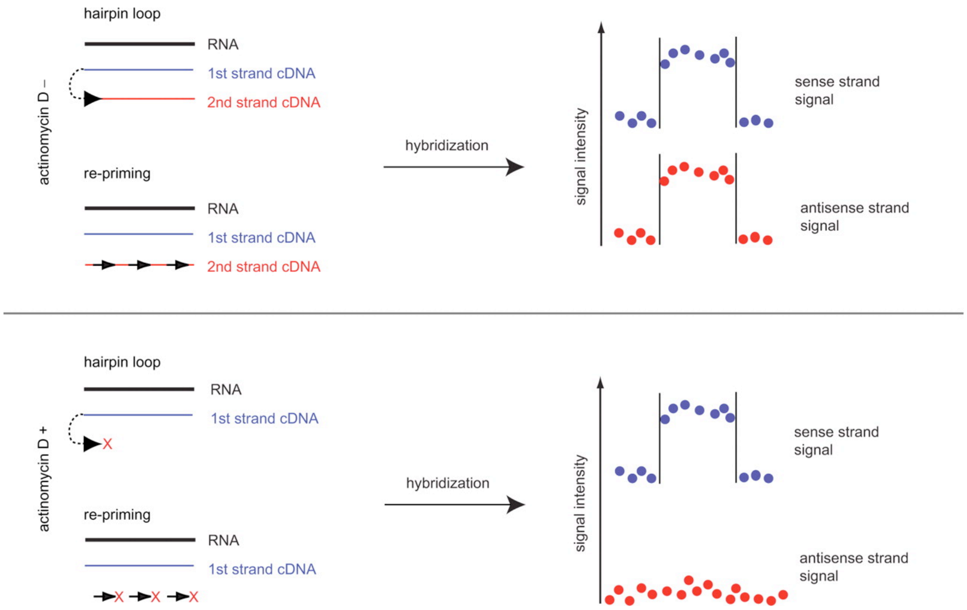


Figure 21: Actinomycin-D eliminates second-strand cDNA synthesis
Retroviral reverse transcriptases possess both RNA- and DNA-dependent DNA polymerase activity. In the viral life-cycle, the RNA genome is reverse transcribed to first-strand cDNA, the RNA genome is hydrolyzed by RNase-H activity of the RT, followed ultimately by second-strand cDNA synthesis. Actinomycin-D eliminates the DNA-dependent DNA polymerase activity, restricting synthesis to that of first-strand cDNA exclusively. Figure reproduced from Perocchi et al. (2007).

3' labeled by incorporation of biotin-N6-dATP using terminal deoxynucleotide transferase (TdT). This method was quite effective at generating labeled first-strand cDNAs; unfortunately, second-strand cDNAs were also produced and end-labeled. These second-strand cDNAs were indistinguishable from first-strand cDNAs from antisense, resulting in spurious identification of antisense transcripts (schematized in figure 21). The reason for this anomaly was recognized (but not addressed) by the Davis lab: MMLV (and AMV) reverse transcriptases normally generate first-strand cDNA, degrade the heteroduplex, and, acting as a DNA dependent DNA polymerase, generate second-strand cDNA (Spiegelman et al., 1970). These activities are clear in context of the viral lifecycle, where double stranded DNA is required for integration into the host genome.

The polypeptide antibiotic Actinomycin-D inhibits this second-strand synthesis activity of MMLV RT, an activity of unknown mechanism which has been recognized for nearly forty years (Müller et al., 1971; Ruprecht et al., 1973) . In 2007, Lars Steinmetz's group published a refinement of the method used by David et al. and compared their original *S. cerevisiae* transcriptome maps to new data derived from cDNAs generated in presence of Actinomycin-D (Perocchi et al., 2007). From this comparison, it became evident that antisense transcription was occurring in budding yeast, but to a lesser degree than previously proposed. Nearly half of the "antisense" transcripts observed in the original data disappeared from the new Actinomycin-D data, leaving antisense signal unchanged for roughly 200 genes (see examples in figure 23). Sense transcript level was shown to be strongly correlated with antisense signal from -Actinomycin-D hybridizations (see figure 22) , but the variability of antisense signal for a given sense transcript level precludes *a posteriori* deconvolution of such data. This is especially relevant in the context of "sampling transcription" which is proposed to occur from actively transcribed regions, resulting in the generation of antisense transcripts from many loci. At least in fission yeast, the majority of these incidental transcripts are not fully processed, and are the subject of further discussion below.

All of the data which I've generated and analyzed come from cDNAs which were reverse transcribed in the presence of Actinomycin-D. The resulting data are implicitly strand-specific, and are uniquely able to identify bona fide antisense transcripts present in *S. pombe*. During the course of my tiling array analysis, two papers were published profiling the transcriptome of fission yeast (Wilhelm et al., 2008; Dutrow et al., 2008). Each of these manuscripts approached the general question of identifying all transcribed sequences in *S. pombe*, but failed to provide significant insight into

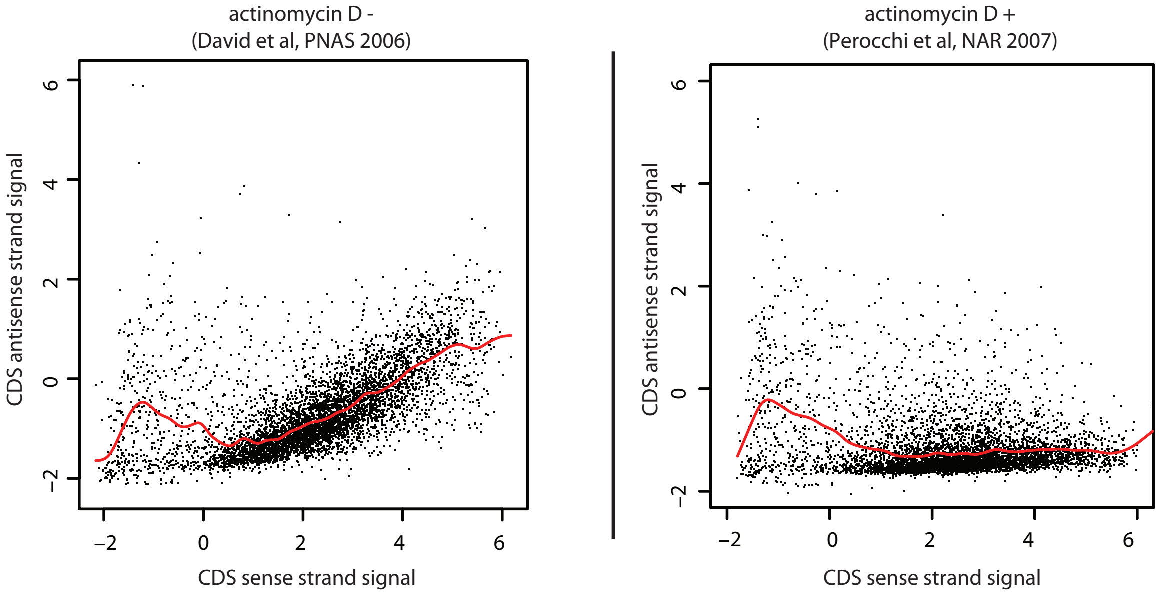


Figure 22: Spurious antisense signals are correlated with sense transcript level in budding yeast
Average sense and antisense signals across each *S. cerevisiae* ORF are shown. When reverse transcription is performed without Actinomycin-D, antisense signals are positively correlated with sense transcript level. When exclusively first-strand cDNAs are generated by addition of Actinomycin-D, the majority of these antisense signals disappear, suggesting that they are reverse transcription artifacts. Data obtained from David et al. (2006) and Perocchi et al. (2007).

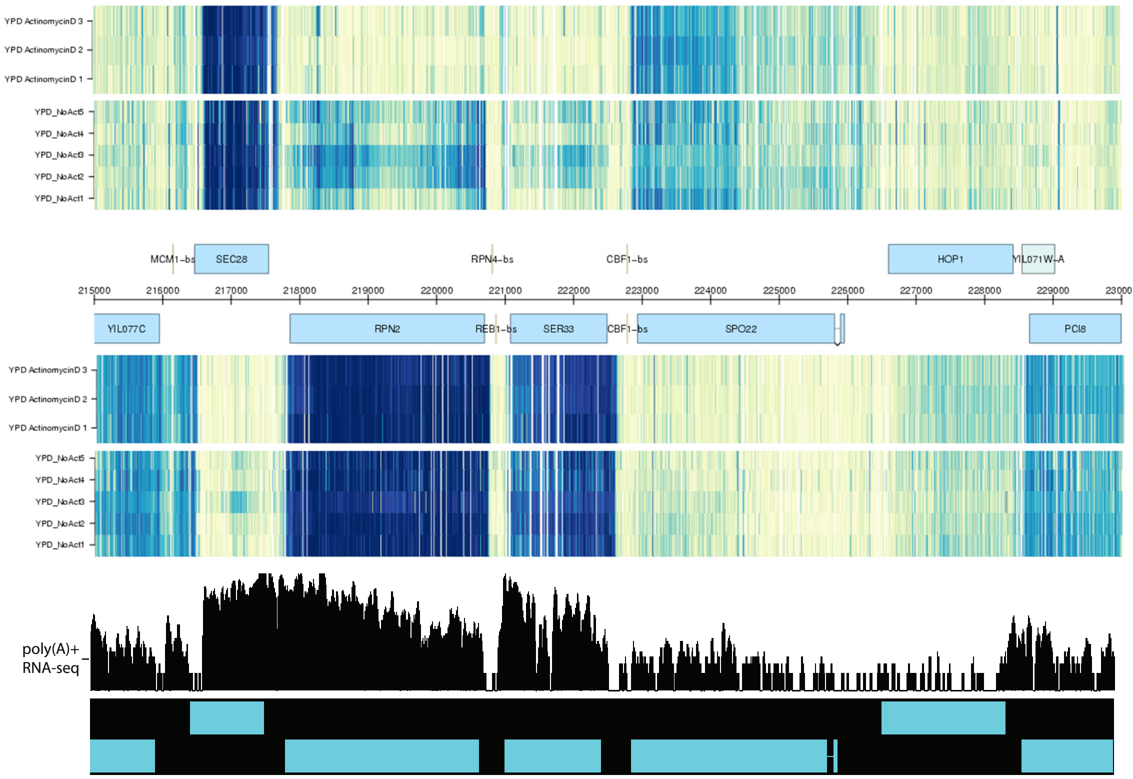


Figure 23: Spurious second-strand cDNAs mimic antisense transcripts

Many of the antisense transcripts identified by David et al. (2006) were artifacts of the DNA-dependent DNA polymerase activity of the reverse transcriptase used. Second-strand cDNAs were generated for a subset of all RNAs, resulting in antisense-like signals. Addition of Actinomycin-D suppresses this activity, excluding synthesis of all but first-strand cDNA (Figure 21 on page 100). Note signal on the non-coding strand (top) for both RPN2 and SER33, which disappears when cDNA synthesis is performed in the presence of Actinomycin-D. Antisense signal of SPO22 (also top) does not decrease in presence of drug, and lacks corresponding sense signal; SPO22 has a *bona fide* antisense transcript in vegetative cells.

Short read sequencing data from vegetative polyA+ RNAs (Nagalakshmi et al., 2008) are shown at the bottom of the panel. Strand-nonspecific data is the unable to determine the edges of the tightly packed transcriptional units present in budding yeast. Note the apparently contiguous signal of SEC28 and RPN2. Lack of strandedness precludes assignment of SPO22 and HOP1 signals as antisense.

Despite millions of mapped sequence reads, the short read data's dynamic range is far below that of the tiling arrays. While tiling array data indicates clear expression of both SPO22 and HOP1 antisense above background, sequencing coverage is poor. Average coverage across SPO22 was only 2.3 reads per exonic nucleotide, 0.7 for HOP1. Coverage is inconsistent, showing numerous Gaussian peaks of reads corresponding to fragmented cDNAs; standard deviations are 99% and 128% of respective means.

Publication	Technology	Stranded	resolution	Input material
Wilhelm et al.	cDNA on Affy	NO	20nt	total RNA
Wilhelm et al.	cDNA-seq	NO	“1nt”	fragmented polyA+ cDNA
Dutrow et al.	direct RNA on Agilent	yes	55nt	total RNA (1 polyA+)
This work	cDNA on Affy	yes	20nt	effectively polyA+ RNA

Table 4: Comparison of transcriptome studies in *S. pombe*

the nature of novel species. Furthermore, neither of these studies provided a combination of high resolution, strand specificity, and analysis of exclusively processed transcripts (detailed in table 4). While publication of these manuscripts preempted publication of my own simple “catalog” of transcribed sequences, they have provided an invaluable contrasts and supporting evidence for several mechanistic insights into global regulation of transcription in fission yeast and beyond .

3.2 Unstable RNAs are a hidden but universal feature of transcriptomes

In addition to stable transcripts which reach the cytoplasmic compartment, many RNAs are generated, fail to be packaged into functional mRNPs, and promptly degraded (Zenklusen et al., 2002). Several unstable non-coding species were serendipitously identified in budding yeast by Alain Jacquier’s group in 2005, when Affymetrix gene expression arrays were used to identify changes in transcript abundance in cells lacking functional Rrp6p in an attempt to profile maturation and turnover of partially processed transcripts (Wyers et al., 2005). The Affymetrix design used incorporated probes for a handfull of cryptic transcripts, signals for whom are normally low in wild-type cells. Several of these cryptic transcripts were stabilized upon depletion of Rrp6p, and were therefore termed “cryptic unstable transcript[s]”, or CUTs. In contrast to genome-wide increases in transcript levels following disruption of RNA turnover, focal changes were occurring and could be detected on microarrays.

rrp6 and *dis3* conditional mutants were already being examined in the lab in the context of regulated meiotic message turnover, and tiling microarray data was generated from these strains. Surprisingly, few changes occur in cells lacking *Dis3* , likely reflecting overlapping function with *Rrp6* (noted previously in budding yeast by Dziembowski et al. (2007)). Depletion of *rrp6*, however, stabilizes hundreds of hitherto unknown transcripts in fission yeast whose presence has significant implications in genome organization. A pair of detailed examinations of non-coding transcription in budding yeast was published during the course of my analysis (Neil et al., 2009; Xu et al., 2009), and suggest that many non-coding transcripts are the result of bidirectional transcription. As I have independently discovered in fission yeast, non-coding transcripts are frequently transcribed divergently from active promoters. The availability of these *S. cerevisiae* non-coding RNA maps has permitted the identification of cross-species patterns of genome organization and significantly broadened the applicability of my findings in *S. pombe.*

Abundant non-coding transcription is not limited to yeast. Tiling array analysis of exosome-depleted human cells demonstrated the active transcription of short upstream polyadenylated non-coding RNAs which were turned over in an hRrp40 dependent manner. This transcription was positively correlated with expression level of the primary message, suggesting a positive regulatory role for transcription of these species (Preker et al., 2008). Recent transcriptome-wide run-on experiments performed in John Lis's lab have demonstrated bidirectional initiation of transcription around active promoters, and observed sense/antisense spacing similar to what we observe in fission yeast (Core et al., 2008). Recent work in *Arabidopsis* has shown that depletion of the exosome leads to accumulation of a variety of different RNAs, including polyadenylated structural RNAs, microRNA precursors, and most intriguingly, heterochromatic repeat-associated transcripts (Chekanova et al., 2007). In fission yeast, heterochromatin formation has recently been shown to require *cid14* poly(A) polymerase activity (Bühler et al., 2007), an ortholog of the *Trf4/5* component of the TRAMP complex in budding yeast responsible for polyadenylation of exosome-bound transcripts. These findings reinforce the idea that heterochromatin is not transcriptionally silent, and suggest that the encoded transcripts are degraded in a co-transcriptional manner conserved from fission yeast to higher eukaryotes.

While generation of stable RNAs has been previously implicated in transcriptional interference mechanisms, little attention has been paid to the possible effects of transcription resulting in unstable species. As the majority of mechanisms proposed for positive and negative regulation by transcription do not require generation of a stable RNA, CUT transcription can engender regulatory activity. As I describe below, genome organization is strongly affected by transcription which results in both stable and unstable non-coding species.

3.3 CUTting apart the genome: non-coding RNAs separate transcriptional units in fission yeast

Introduction

As we peer deeper into the transcriptome, widespread transcription of much of the genome emerges as a universal feature of eukaryotic biology. Though early studies of expression were focused on discrete coding or structural RNAs, it has become clear through the development of unbiased transcript sequencing (Velculescu et al., 1995) that far more of the genome is transcribed than can be ascribed a functional role. The scope of this transcription is stunning. The recently completed ENCODE pilot project has provided evidence for transcription of 93% of bases interrogated within the human genome (ENCODE Project Consortium et al., 2007), most of which lack apparent coding potential, echoing early fine structure studies of transcription human cell lines which found that nearly 90% of the transcribed sequences fell outside of annotated exons (Kapranov et al., 2002). Over the last decade, many new classes of functional non-coding RNAs have been identified (reviewed in Hannon et al. (2006)), but despite these advances, a large fraction of the genome is still transcribed without clear purpose.

Advances in high-throughput detection and quantitation have enabled the cataloging and characterization of transcripts in a variety of model systems, from which it has become clear that widespread transcription is not limited to spacious metazoan genomes. Transcriptome studies of budding (David et al., 2006; Xu et al., 2009) and fission yeast (Dutrow et al., 2008; Wilhelm et al., 2008) have mapped widespread expression, refined the structure of known transcripts, and identified hundreds to thousands of novel transcripts whose presence had not been predicted *in silico.* While regulatory functions have been ascribed to a handful of non-coding transcripts in budding (Martens et al., 2004; Hongay et al., 2006) and fission (Hirota et al., 2008) yeast, these RNAs have been largely considered incidental byproducts of genic expression.

Results

We have examined the processed transcripts present in *S. pombe* using a high resolution strand-specific approach. Commercial tiling microarrays were used to interrogate long (>200nt) polyA+ RNAs from asynchronous mitotic cells grown in defined laboratory media (EMM). Sample preparation was carried out largely as described (David et al., 2006) with the notable addition of Actinomycin D during cDNA synthesis to avoid generation of confounding second strand cDNAs (Perocchi et al., 2007) (see Supplemental Methods). Technical replicate hybridizations were performed for all samples. These data represent processed, polyadenylated RNAs in vegetative *S. pombe*. Non Pol II transcribed RNAs are absent, as is much of the "sampling transcription" described previously (Figure 24) . Normalized probe intensities were computationally partitioned into genomic intervals of similar signals (Huber et al., 2006), and the resulting segments were analyzed further. These data represent the first high-resolution strand-specific examination of the *S. pombe* transcriptome.

Consistent with findings in mammals, we have identified processed RNAs corresponding to 90% of the genome from a single growth condition. During the initial sequencing of the fission yeast genome(14), a number of "gene free regions" were identified. We find that many of these intervals are actively transcribed into discrete non-coding RNAs and as long untranslated regions (UTRs) of adjacent open reading frames (Figure 25) . While gene free, these regions do not go untranscribed. In addition to the large amounts (~10%) of contiguous non-coding sequence in the form of 5' and 3' UTRs, we have found a striking number of discrete non-coding RNAs present in the *S. pombe* genome. We have identified hundreds of stable non-coding

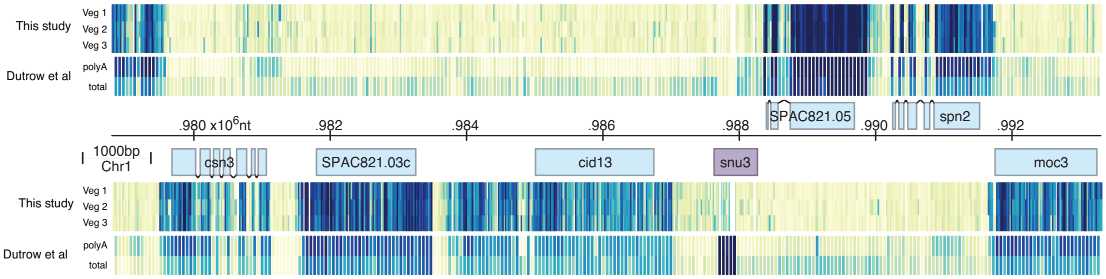


Figure 24: High resolution, strand specific quantitation of processed transcripts Our data represent a high-resolution, strand specific snapshot of the polyadenylated transcripts present in fission yeast. Shown above is data from this study and Dutrow et al. (2008) across a 15kb window of chromosome I. In this window, all signals are confined to a single strand corresponding to an annotated transcript. Intronic probe signals are substantially lower than flanking exons (see spn2 for example) and approach background signal levels. Our data lack the sampling transcription observed by Dutrow et al. in total RNA hybridizations, and more closely resemble their polyA+ enriched samples. A notable exception is seen at the snRNA snu3. This abundant non-polyadenylated transcript is virtually absent in our hybridizations (signal is within the first percentile of all annotated features), but was not depleted during polyA+ enrichment by Dutrow et al. (signal is above ninety-fifth percentile of annotated features).

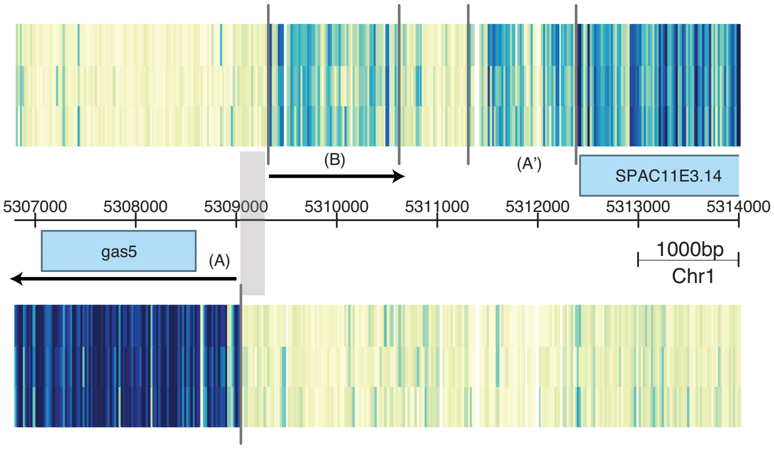


Figure 25: Gene free regions are actively transcribed
A number of gene-free regions were identified in the otherwise densely annotated *S. pombe* genome during sequencing (Wood et al., 2002). Although lacking open reading frames, we find that these regions are actively transcribed. Shown here is a portion of Chromosome I along which there is a 4kb region lacking annotated features. This region is actively transcribed in vegetative cells as both a divergent non-coding transcript (B) from the gas5 TSS (A) and a long 5' UTR of SPAC11E3.14 (A'). Similar unannotated transcripts are observed at nearly all other annotation-free regions.

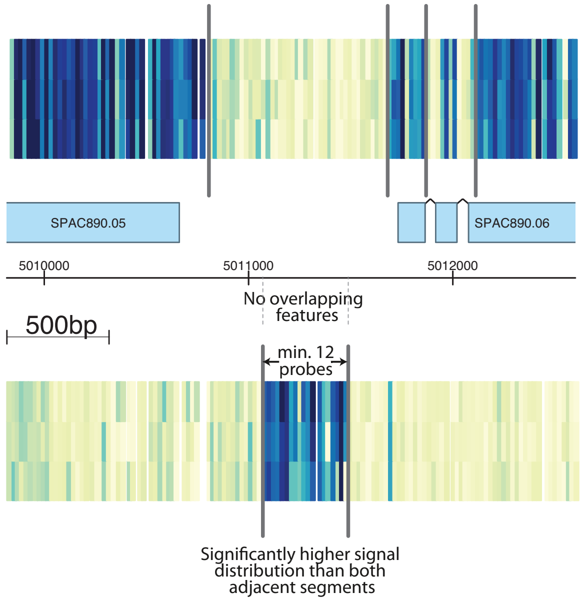


Figure 26: Definition of non-coding RNA used in this work
A conservative definition of non-coding RNAs was used in this work to minimize false discovery rate. Segments must satisfy multiple criteria to be called a novel feature. A minimum of 12 contiguous unique probes must be present, corresponding to a minimum feature size of 240nt or more (consistent with the minimum size recovered from RNA purification methods used). The novel feature must be expressed at a significantly higher level than both flanking segments and above background (as defined by a minimum difference in segment mean and one-sided KS test p-value). Finally, the novel feature must not overlap any current annotation.

RNAs using a conservative definition designed to minimize false positive calls (Figure 26) . This algorithm has identified more than 400 previously unannotated discrete features in addition to 100 previously described ncRNAs and 21 snoRNAs. These novel transcripts are expressed across a range similar to coding genes (Figure 27) , and have a median length roughly two-thirds that of the average *S. pombe* coding sequence. ncRNAs identified here are abundant in number and total length; these 503 well-defined non-coding transcripts account for 5% of the genome, more than that occupied by all identified introns.

Previous genome-wide studies of fission yeast have cataloged many non-coding

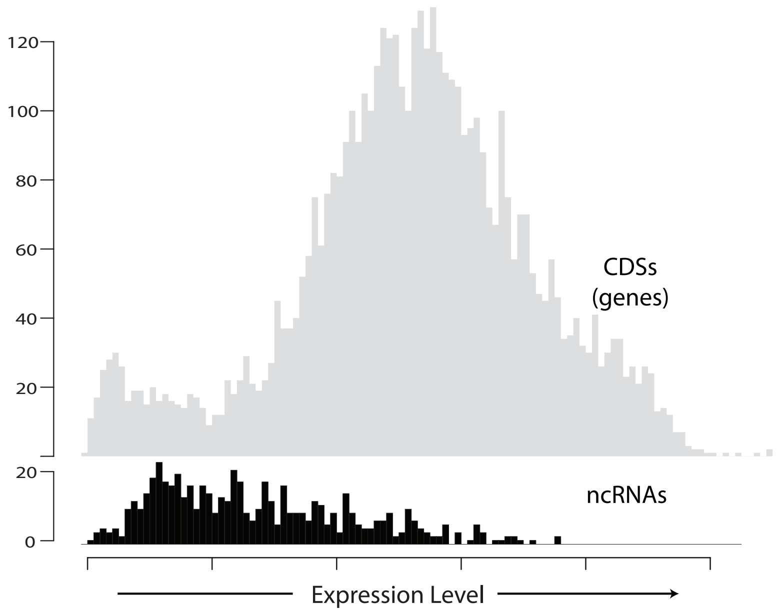


Figure 27: Non-coding RNAs are expressed across a range similar to coding transcripts
The non-coding RNAs identified in this study are expressed across a wide range of expression levels, and at levels similar to annotated coding sequences. Mean intensities of all probes comprising each non-coding feature are shown below the mean signal across coding bases (exclusively exonic probes) for all *S. pombe* genes. The majority of coding sequences are expressed well above background level, though transcript level varies across several orders of magnitude. Analysis of antisense transcription is presented below on page 131.

RNAs (Wilhelm et al., 2008; Dutrow et al., 2008) in the context of transcriptome mapping, but little effort has been focused on analyzing the function or formation of this abundant class of transcripts. The combination of strand specificity and high spatial resolution of our data afforded us a unique look at the orientation and positioning of these non-coding RNAs in the context of all processed transcripts. In vegetative cells, we find that 41% of stable ncRNAs are situated in a manner consistent with their being the result of bidirectional transcription from the promoter of known transcripts. In such cases, the 5' edge of the divergent ncRNA maps to less than 250bp from the transcription start site of the cognate forward gene (Figure 28), a finding mirrored by recent studies in budding yeast in which many identified non-coding RNAs appear to initiate from a single shared nucleosome free region upstream of a coding gene (Neil et al., 2009; Xu et al., 2009). Widespread divergent transcription has also been described in mammals, where high throughput sequencing of ES cells revealed a novel population of small RNAs that represent transcription of both strands around known transcription start sites (Core et al., 2008; Seila et al., 2008). In this context, our findings suggest that bidirectional transcription is a universal feature of eukaryotic genomes.

The presence of non-coding RNAs adjacent to transcribed genes could be due to sampling transcription initiating from accessible DNA near an active promoter. If, however, divergent ncRNAs are the result of bidirectional transcription from a single promoter, changes in forward (message) transcription would be mirrored by changes in the level of the associated divergent ncRNA. We tested this causal model by examining expression changes around a set genes whose expression is strongly regulated by the presence of thiamine (vitamin B1). Comparing transcript abundance between cells grown with or without thiamine, we find that only 17 transcripts change substantially in response to this nutrient (Figure 29) , and that in addition to gene expression level,

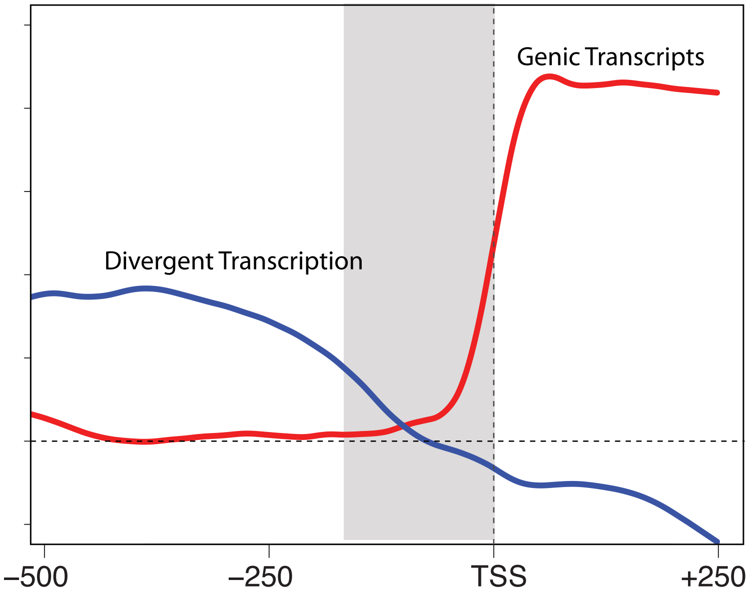


Figure 28: Divergent transcripts initiate within a short distance of the associated genic TSS
Initiation of a divergent transcript occurs within a narrow window upstream of the genic transcription start site. Each gene with an associated divergent non-coding RNA was aligned by segmentation-determined transcription start site, and probes 250nt 3' (towards the initiator codon) and -500nt (upstream) are shown. Divergent transcripts are represented by the blue line, whose turning point is present within a window consistent with a single nucleosome free region. Of note, the "opposite strand" signal is below background adjacent to the forward transcript. This is likely due to transcriptional interference from genic transcription excluding antisense "sampling" across this region.

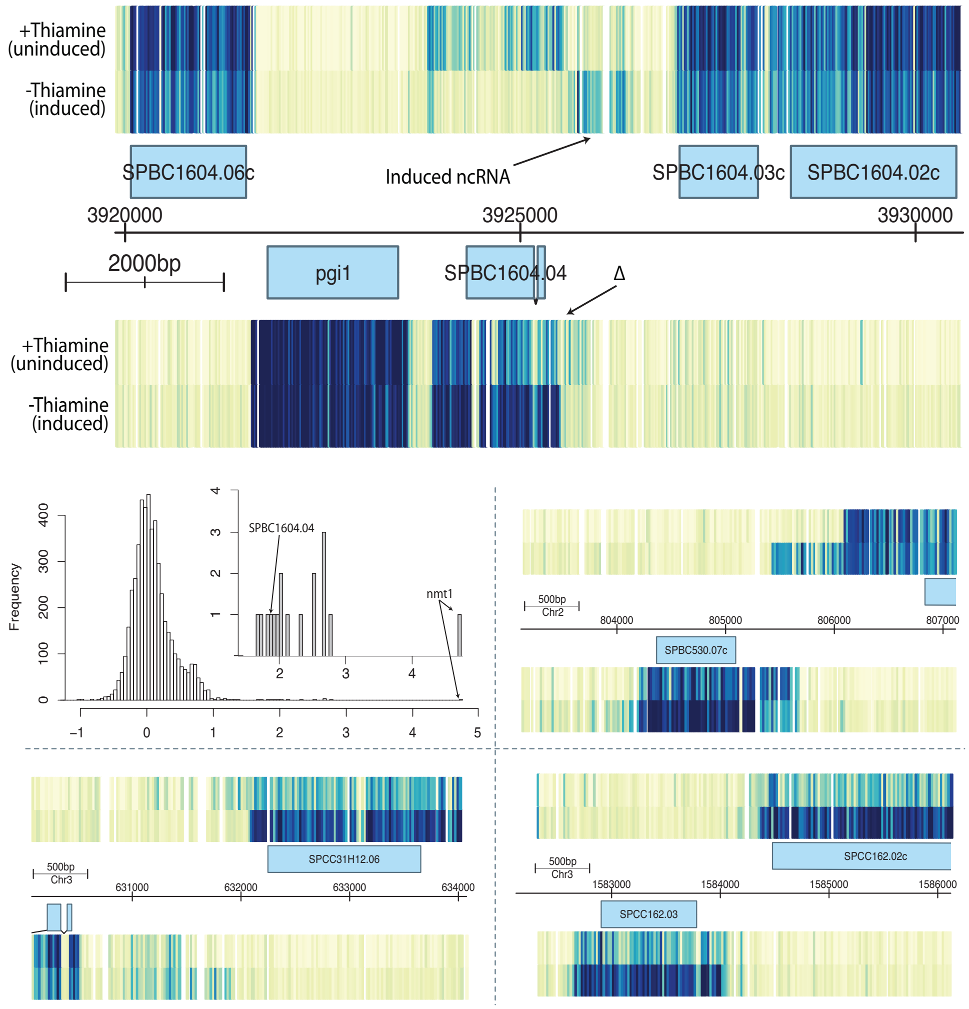


Figure 29: Thiamine starvation induces changes in transcript level and structure Differences in expression level between cultures grown in minimal medium with thiamine (+B1) or without (-B1) are shown. Shown above, divergent transcription occurs when the associated gene is induced. Antisense signal disappears with the onset of forward transcription, possibly suggesting an additional layer of antisense regulation.
Genome wide, only seventeen genes are robustly induced when cells are starved for thiamine (vitamin B1), shown at mid-left. As above, marked changes in transcript structure occur around nearly all induced genes. Diffuse 5' edges (present in the uninduced +B1 state) consolidate to a more discrete, shorter 5' UTR when expression is induced (top, mid right and bottom left panels). Divergent transcription is induced for the majority of these genes (12 of 17), including induction of non-coding RNAs and co-induction of divergent genes involved in thiamine biosynthesis (bottom right).

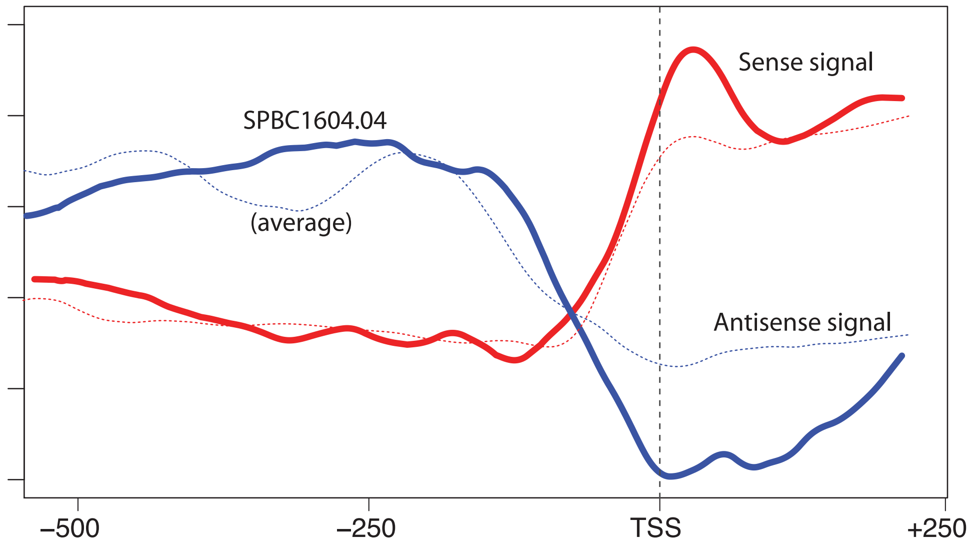


Figure 30: Bidirectional transcription occurs around B1 starvation induced genes Bidirectional transcription occurs around genes induced by B1 starvation. Instead of absolute transcription level as in figure 28, difference in signal intensity is shown.

transcriptome structure changes accompany induction. For more than two thirds of these regulated genes (12 of 17), transcription is induced bidirectionally, consistent with coupled regulation of both transcripts from a single initiating element (Figure 30). Though markedly reduced by its absence, transcripts for thiamine repressed genes are visible in cells grown in B1. In contrast, most associated non-coding RNAs are only evident under conditions which activate transcription.

To exclude the possibility that induced non-coding RNA transcription was limited to a single regulatory scheme, we expanded our analysis of induced expression to changes in expression during meiotic differentiation, a program which has previously been shown to involve regulation of hundreds of transcripts (Mata et al., 2002). RNA was extracted from synchronous pat1-114 cells undergoing the first meiotic division (MI) (Cervantes et al. (2000), see 4h time point below in figure 36 on page 135), at which time genes of the Mei4p-regulated mid-meiotic program are expressed. We examined expression changes around meiotically-induced Mei4p-dependent genes, and find that divergent transcription is induced for nearly 40% (120 out of 306 genes). The majority of these divergent transcripts represent non-coding RNAs, although several closely spaced pairs of divergently oriented coregulated genes were identified. Transcription of stable noncoding RNAs is coupled to expression of adjacent gene, and bidirectional activity is observed for both transcription factors tested .

In addition to the stable RNAs described above, much of the genome is transcribed into unstable species that are rapidly degraded. In higher eukaryotes, only an estimated 5% of Pol II transcribed ribonucleotides are exported to the cytoplasm as stable RNAs (Moore, 2002), while the majority are culled by nucleolytic surveillance mechanisms. At the center of RNA quality control is the exosome (reviewed in Schmid and Jensen (2008)), whose disruption leads to accumulation of normally degraded transcripts. Studies in budding yeast and *Arabadopsis* have described a new class of discrete unstable non-coding transcripts that accumulate following depletion or inactivation of the conserved nuclear exosome subunit Rrp6p (Wyers et al., 2005; Chekanova et al., 2007). These cryptic unstable transcripts (CUTs) are actively transcribed and, in budding yeast, gain a poly(A) tail via the TRAMP complex (Thiebaut et al., 2006). Despite transcription and adenylation, their rapid turnover normally precludes detection.

We used a conditional allele to inactivate rrp6 in mitotic cells and stabilize these normally degraded RNAs. The CUT containing extracts were interrogated as oligo(dT) primed cDNAs on tiling arrays. Probe intensities were compared to wild-type, and the differences between the CUT-enriched and wild type signals were partitioned and analyzed. Using identical criteria as above, we have identified 996 discrete unstable RNAs that are transcribed during growth in minimal media. These transcripts are shorter than their stable counterparts, with a mean length of 518nt vs 933nt,

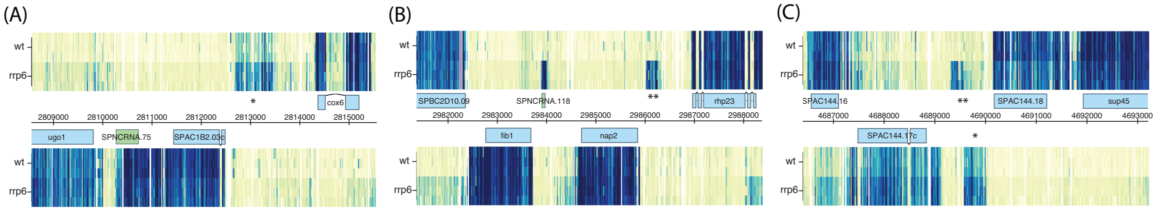


Figure 31: Stable and unstable ncRNAs are divergently transcribed from active promoters
Hundreds of novel non-coding transcripts which are divergently transcribed from promoters of known genes have been identified in this work; several examples are shown here. Both stable (*) and unstable (**) ncRNAs have been characterized, with the latter defined as non-coding transcripts that are visible only following inactivation of the nuclear exosome subunit Rrp6. Intensities of many "stable" non-coding RNAs increase when Rrp6 function is absent (A and C). Numerous divergent pairs have associated *pairs* of non-coding RNAs, as seen in (C). In addition to discrete non-coding RNAs, long diffuse 3' UTRs (possible readthrough transcription) are visible for several genes (eg, fib1 and nap2) when Rrp6 function is depleted.

and most, 56%, are transcribed divergently from the start of annotated genes (see panels B and C of figure 31) . As observed for stable ncRNAs, these CUTs are strongly associated with actively transcribed genes, reinforcing a causal relationship between genic and divergent non-coding transcription. Genes with divergent stable non-coding RNAs or CUTS are expressed at a significantly higher level than those without ($p<2.2*10^{-16}$, one-sided t-test), though expression level is no different between genes that give rise to stable versus unstable transcripts (p=0.318, two-sided t-test).

Non-coding transcription has broad implications on genome organization. Transcription through promoter regions has been shown to strongly modulate expression of the downstream gene, and overlapping transcription has implications in both transcriptional and RNA interference mechanisms (Ward and Murray (1979); Adhya and Gottesman (1982); Martens et al. (2004); Hongay et al. (2006); Martianov et al. (2007), reviewed in Mazo et al. (2007)). Divergent gene pairs regulated by a shared promoter or enhancers have been well characterized (Williams and Fried, 1986; Burbelo et al., 1988; Trinklein et al., 2004), and genome-wide, a disproportionate number of gene pairs in eukaryotes are arranged in a head-to-head orientation. Phylogenic analysis has demonstrated that these gene pairs are more strongly conserved than tandem-oriented genes (Li et al., 2006), suggesting widespread co-regulation of divergent transcripts. We have found numerous functionally related co-regulated gene pairs in fission yeast that respond to thiamine starvation or meiotic differentiation, but despite these examples of coordinated expression, expression of many divergent gene pairs is apparently unlinked. Distance between adjacent genes does not account for this, as there is only a slight negative correlation between inter-transcript distance and co-regulation. Since many of the identified non-coding RNAs are present between divergent genes, we postulated that non-coding transcripts affect coregulation of neighboring transcription units.

We turned to expression microarray data to examine the regulatory patterns of genes that give rise to these non-coding RNAs and to determine the effects of ncRNA transcription on adjacent genes. 863 hybridizations comprising the results of numerous publications and unpublished work from multiple labs including our own were analyzed. These data, representing diverse transcriptional responses covering a wide variety of environmental and developmental stimuli, were distilled to the standard scores of Pearson correlations between genes and the variances of expression ratios across all experiments; this analysis is described in detail below.

We find that divergent stable non-coding RNAs and CUTs are the result of fundamentally different modes of transcription. Genes that generate stable ncRNAs are extensively regulated, insofar as their relative expression changes widely across the experiments examined; expression ratios for genes with a stable divergent ncRNA vary significantly more than those of genes without ($p<0.0005$, one sided t-test). This finding is consistent with the widespread appearance of divergent stable transcripts adjacent to genes induced by B1 starvation and during mid meiosis. Conversely, unstable ncRNAs are associated with relatively invariant transcripts; CUT-associated expression varied less than other genes ($p<0.036$, one-sided t-test). We propose that actively regulated transcription, in contrast to constant basal expression, results in the formation of stable, polyadenylated divergent ncRNAs, consistent with the recruitment of 3' processing factors to upstream regulatory regions of genes bound by strong transcriptional activators (Uhlmann et al., 2007).

Confirming expectations from previous phylogenetic studies, our data support the notion that orientation of adjacent genes affects their co-regulation. Expression of divergent gene pairs is better correlated than that of tandem genes across the experiments examined ($p<5.2*10^{-9}$, one-sided t-test). Though it has been well established that enhancer activity can propagate over long genomic distances (Nasmyth, 1986), not all divergent gene pairs are co-regulated. We find that presence of a non-coding RNA between divergently oriented genes diminishes their co-regulation compared to pairs lacking such a transcript ($p<1.06*10^{-8}$, one-sided t-test). Non-coding RNA transcription de-couples genes comprising otherwise adjacent transcriptional units, and does so quite absolutely: divergent genes with an intervening ncRNA are no better correlated than tandem adjacent genes ($p=0.450$, two-sided t-test). Of note, tandem and de-coupled divergent gene pairs are more correlated than random pairs of distal genes, likely reflecting effects of DNA topology or the local chromatin environment.

Although stable ncRNAs and CUTs are generated from different types of promoters, they are equally able to reduce co-regulation of adjacent genes; we observe no difference in the de-coupling activity of these two classes ($p=0.7553$, two-sided t-test). This finding suggests that the act of transcription, more than the persistence of

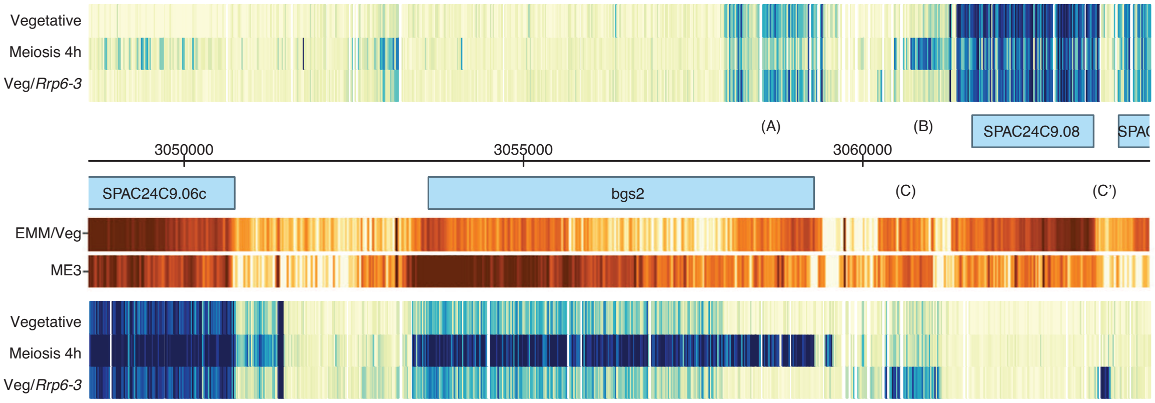


Figure 32: Complex transcriptional regulatory mechanisms are missed by strand non-specific sequencing
Tiling array data (this study) for a 15.5kbp region of chromosome I are presented in blue. Transcription of the meiosis-specific spore-wall synthesis gene bgs2 occurs from an internal transcription start site in vegetative cells. Bidirectional transcription from this site occurs, generating a 5' truncated bgs2 and an antisense transcript (A) which extends to roughly the full-length transcript's start. During meiosis, full length bgs2 is transcribed from an upstream promoter and levels of the internally initiated divergent transcript decrease. Bidirectional transcription also initiates from this new promoter, and generates a non-coding transcript which is absent in vegetative cells (B). In mei4Δ vegetative cells depleted of Rrp6 function, unstable non-coding RNAs (C and C') are visible as a divergent transcript from the adjacent genes, SPAC24C9.08 and SPAC24C9.09, respectively.
Sequencing data from Wilhelm et al. (2008) is shown in orange, and lacks both the strand specificity and consistency in signal of our tiling array data. Of note, the sequencing data are strongly biased toward the 3' end of transcripts measured, as seen for both induced bgs2 (Meiosis 4h for tiling data, "ME3" for sequencing) and vegetative SPAC24C9.08.

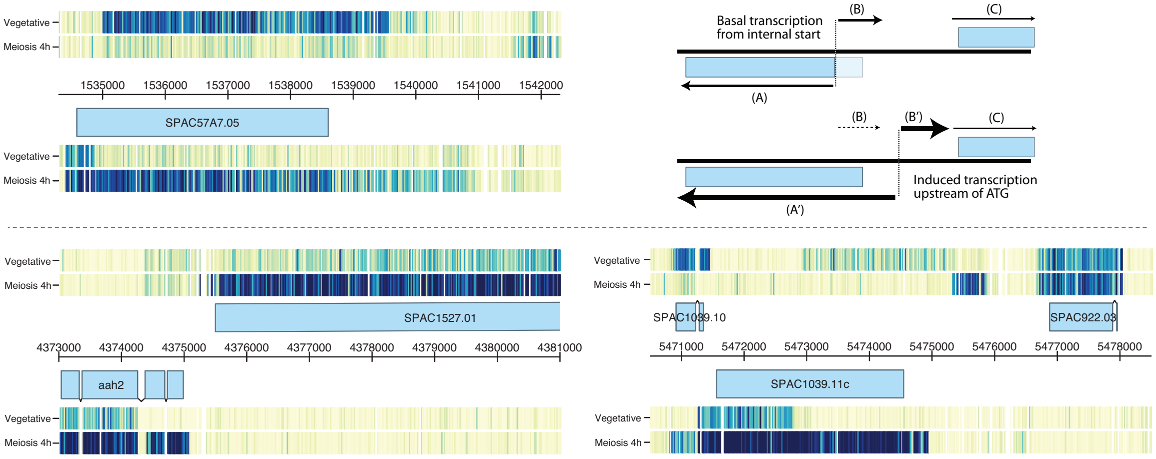


Figure 33: Internal divergent transcription regulates a functionally related group of genes

In addition to bgs2, shown on the previous page, several other related spore-wall specific enzymes have internal transcription start sites in vegetative cells. In each case, transcription from an internal start site generates a 5' truncated sense transcript (A) and divergent non-coding RNA (B) in vegetative cells. Upon meiotic induction, an upstream transcription start site is used, generating full-length sense (A') and a new divergent transcript (B'). For aah2, the divergent transcript is the strongly co-regulated mok11 gene (standard score of Pearson correlation from aah2, 2.51). For all other cases shown, a non-coding RNA (B') is produced, and isolates transcriptional activation of (A') from the adjacent coding gene (C). SPAC1039.11c and bgs2 (shown on the preceding page) are effectively decoupled from their adjacent genes (standard score of Pearson correlation -0.17 and -0.00, respectively).

a noncoding transcript itself, is responsible for this genome partitioning activity. Our findings in *S. pombe* are echoed in budding yeast. We examined the recently cataloged *S. cerevisiae* stable and unstable non-coding RNAs (Xu et al., 2009) in the context of 1728 published budding yeast expression hybridizations. To analyze a similar subset of transcripts in both yeasts, we excluded non-coding RNAs that overlap annotated coding sequence on either strand, and examined the remaining 553 stable and 505 unstable intergenic transcripts (65% and 55% of total, respectively). As in fission yeast, stable non-coding RNAs are associated with regulated genes whose expression varies more than those lacking non-coding transcripts ($p<3.4*10^{-9}$, one-sided t-test). Divergent gene pairs are more co-regulated than tandem genes ($p<1.7*10^{-10}$, one-sided t-test), and presence of a non-coding RNA de-couples adjacent genes compared to pairs that lack an intervening feature ($p<8.9*10^{-10}$, one-sided t-test). As in *S. pombe*, no difference was observed in the de-coupling activity of stable and unstable transcripts (p=0.586, two-sided t-test).

Communication between promoters and distal enhancers can occur over long genomic intervals. We propose that transcription of non-coding RNAs de-couples regulation of adjacent genes by acting as a “sink” for adjacent promoter/enhancer activity (Figure 34 on the next page). This isolating activity is engendered by the acts of transcription and termination of the non-coding transcript; stability of the resulting non-coding RNA is irrelevant, and we suggest that the fate of this RNA is determined by the nature of the factors loaded at the transcription initiation complex. This activity is conserved between two widely divergent model eukaryotes, and despite the larger intergenic distances present in multicellular organisms, we have no reason to expect this activity to be absent from such systems. Finally, we have identified a large number of novel non-coding transcripts present in fission yeast, provided the first genome-wide assessment of unstable transcripts in *S. pombe*, and suggest that

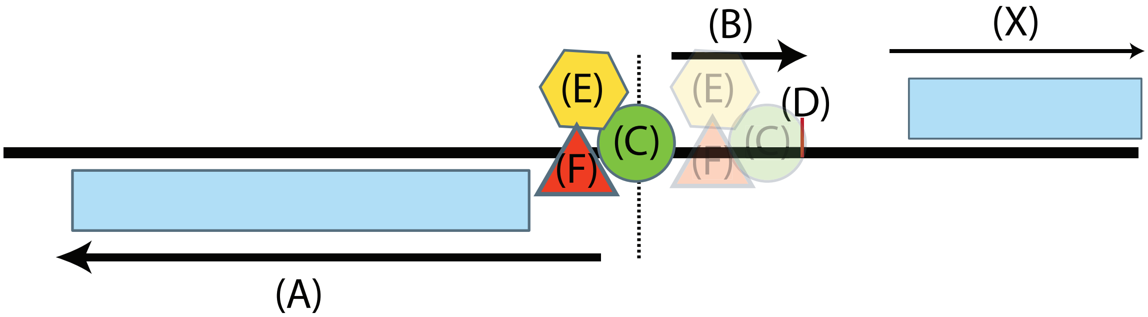


Figure 34: Model of non-coding transcription's isolating effects

Many non-coding RNAs are the result of bidirectional activity of a single promoter; transcription of (A) is concomitant with divergent ncRNA (B), whose stability is dictated by the nature of the transcriptional activator (C) present. Strongly regulated transcriptional units are associated with processed, stable ncRNAs, while relatively invariant genes tend to generate Rrp6-degraded CUTs., consistent with recruitment of 3' processing factors (F) to the site of transcription initiation by strong transcriptional activators (Uhlmann et al., 2007).

A coupled non-coding RNA (B) is transcribed and terminated by unknown sequences downstream (D), likely through a Nrd1-Nab3-like pathway, and avoids interference with the promoter of the adjacent gene (X). Transcription of (B) depletes any cis-scanning activity of transcriptional activators or enhancers bound upstream (E); through the transcription of a non-coding RNA (B), the adjacent coding gene (X) is isolated from factors regulating expression of (A).

stability of divergent non-coding RNAs is determined by the nature of the parent promoter in both fission and budding yeast.

Methods

Synchronous meiotic induction was performed largely as described (Cervantes et al., 2000). h^+/h^+ pat1-114/pat1-114 ade6-M210/ade6-M216 diploids (local strain F277, obtained from YGRC) were grown in YEL+A (yeast extract + adenine, 0.5% yeast extract, 3% glucose, 100μg/mL adenine) at 25°C to saturation. Cultures were diluted 1:100 into EMM2* (Modification of EMM, 15mM potassium hydrogen pthalate, 10mM dibasic sodium phosphate, 93.5mM ammonium chloride, 0.5% glucose, 1x each EMM salts, EMM minerals, EMM vitamins) + 75μg adenine and grown at 25°C to OD_{600} = 0.3. Cells were harvested by centrifugation, washed twice with water, and resuspended in original volume of EMM2*-Nitrogen (as EMM2* above, without ammonium chloride) + 10μg/mL adenine. After 14h at 25°C, G_1 arrested cells were harvested by centrifugation and resuspended in pre-warmed 34°C EMM2*-complete (as above, with 5g/L NH_4Cl and 75μg adenine) to OD_{600} = 0.3. Samples were collected for flow cytometry and RNA extraction every 2 hours.

RNA was extracted from flash-frozen cell pellets using a commercial glass-fiber adsorption kit (Ribo-Pure Yeast, Ambion, Austin TX). 30 OD_{600} units (effectively $3x10^8$ cells) were processed per column, as per manufacturer instructions. Post-elution removal of genomic DNA was performed by digestion with 40U of DNase I (Roche, Indianapolis, IN) for 1h. RNAs were recovered by column adsorption (RNeasy, Qiagen, Valencia, CA). Sample integrity was confirmed by microcapillary electrophoresis (Bioanalyzer 2100 RNA-nano, Agilent), and quantitated spectrophotometrically (Nanodrop ND1000, Thermo Scientific).

cDNAs were synthesized from polyadenylated messages. 400μg of total RNA was mixed with 37.5μg anchored oligo(dT) primers (equimolar (dT)$_{16}$-(dA/dG), (dT)$_{16}$-dC(dA/dG/dC)) in a final reaction volume of 300μl and incubated for 5' at 65°C, 2' at 0°C, followed by 2' at room temperature. To these primer-annealed RNAs, 90μl 5x First Strand Buffer (Invitrogen, Carlsbad, CA), 22.5μl 0.1M DTT, 4.5μl 600μg/μl Actinomycin-D (Sigma), 3μl RNasin, 12μl Superscript III RT (Invitrogen), and 18μl mixed 10mM dNTPs (including 2mM dUTP) were added for a total reaction volume of 450μl. Reverse transcription was performed at 42°C for 16h, followed by hydrolysis of RNA template by addition of 2μl 10mg/ml RNase A + 1μl RNase H (NEB, Ipswich, MA) and incubation at 37°C for 30'. Reaction was split into thirds and cDNAs were purified by column adsorption (PCR cleanup spin kit, Qiagen), followed by pooling of column eluates (total of 90μl, 10-15μg cDNA by spectrophotometry).

Purified cDNAs were fragmented and end labeled. To 85μl cDNA, 10μl 10x fragmentation buffer (Affymetrix, Santa Clara, CA), 2μl UDG (uracil DNA glycosidase), and 3μl APE 1 (human apurinic/apyrimidinic endonuclease) were added. Fragmentation was performed at 37°C for 1h followed by enzyme inactivation at 93°C for 10' and snap cooling to 4°C for 2'. Fragment size was determined by capillary electrophoresis (Bioanalyzer 2100 RNA-nano, Agilent), with a peak length of ~70nt. 93μl of fragmented cDNA was end labeled in a reaction including 30μl 5x TdT buffer, 3μl DNA labeling reagent (proprietary biotinylated dNTP, Affymetrix), and 16μl water. The 150μl reaction was incubated at 37°C for 1h, 70°C for 10', and cooled to 4°C for 10'.

Tiling array hybridizations were performed in technical triplicate. For each array cartridge, 150μl of hybridization cocktail was prepared, containing 5μg of fragmented end labeled cDNA (final concentration 33ng/μl), 2.5μl Control Oligo B2 (Affymetrix, final concentration 50pM), 75μl 2x Hybridization buffer, 10.5μl DMSO (final concentration 7%). Hybridization cocktail was mixed, briefly centrifuged, and denatured at

99°C for 5', followed by slow cooling in an air incubator to 45°C for 5'. Array cartridges (*S. pombe* tiling 1.0FR, Affymetrix) were loaded with 130μl of probe mixture and hybridized overnight at 45°C rotating at 60rpm. Following hybridization, probe cocktail was recovered, pooled, and stored at -20°C for future reuse.

Arrays were washed and stained as per manufacturer instructions. Hybridized arrays were filled with 160μl room temperature Wash Buffer A and mounted in a primed FS450 fluidics station (Affymetrix). Cartridges were processed as per manufacturer's FS450_0002 protocol. Following cartridge draining, arrays were washed with 10 cycles (2 drain/fill per cycle) of Wash Buffer A at 30°C and 6 cycles (15 drain/fill per cycle) using stringent Wash Buffer B at 50°C. Arrays were stained with SAPE for 10' at 35°C, followed by post-stain wash of 10 cycles (4 drain/fill per cycle) of Wash Buffer A at 30°C. Second staining was performed using biotinylated goat IgG anti-SAPE solution for 5' at 35°C, followed by an additional SAPE stain for 5' at 35°C. A final wash was performed using 15 cycles (4 drain/fill per cycle) of Wash Buffer A at 35°C. Cartridges were filled with 160μl Array Holding Buffer and immediately scanned on a GeneChip Array Scanner, model 3000-7G (University DNA Microarray Facility, Stony Brook University, Stony Brook NY).

Grids were placed and aligned to raw image files automatically (GeneChip Operating System 1.4, Affymetrix). The resulting cell level summary files (.CEL) were used for further analysis. Array probe sequences were obtained from Affymetrix synthesis map files and were mapped to a recent *S. pombe* genome revision (4/2007) using MUMmer (Delcher et al., 2002). Only probes which had a perfect genomic match were assigned positions, and of these, any probe which matched 23 or more nucleotides at multiple positions was flagged as potentially cross-hybridizing. All array features were mapped, including perfect match, mismatch, and control probes. In total, nearly 1.1 million features were unambiguously mapped to the genome (detailed above). Signals for all cDNA arrays were normalized using a variance stabilizing algorithm (Huber et al., 2002) against the mean probe values for all arrays analyzed. Background correction of variance stabilized values was performed using the mean probe signals of three genomic DNA hybridizations generated previously (Wilhelm et al. (2008), Array Express accession E-TABM-18).

Expression scores for genes and non-coding features were directly generated from normalized and background corrected $\log_2$ signal intensities. Coding feature scores were calculated as the mean intensity of non-crosshybridizing probes that mapped entirely within exonic sequence. For non-coding features, scores represent the mean of all non-crosshybridizing probes mapping entirely within feature boundaries. All scores represent strand-specific signals.

Normalized and background corrected intensities from asynchronous vegetative samples were positionally segmented using an algorithm derived from Huber et al. (2006). Transcript boundaries for annotated features were assigned computationally by identifying the segment edge nearest the annotation boundary which delineated a distribution of feature-representative probes from a significantly different probe population. A Kolmogorov-Smirnov (KS) test was performed between probes within the putative UTR and 1) those mapping entirely within the annotated feature (and only to exons, where appropriate) and 2) probes comprising the annotation-distal segment. High confidence transcript boundaries were called only for annotated features where segmentation demarcated a lower signal distributions adjacent to the feature (one-sided KS p-value $\leq 10^{-5}$), and where probes falling outside of the annotation boundary were from a distribution similar to those within the annotated feature (two-sided KS p-value $\geq 10^{-2}$) and difference of the means did not exceed 10%. Features adjacent to or overlapping regions of ambiguous or absent probe coverage were excluded from edge mapping.

Non-coding RNAs were empirically defined as intervals of significantly higher signal intensity than surrounding segments and overall background signal which do not overlap a previously annotated CDS. Contiguous regions of at least twelve unambiguous probes (nominally 240nt, consistent with the minimum length fragments recovered during RNA collection and cleanup) were considered novel features if signal intensities were significantly higher than both flanking segments by both one-sided KS test (p-value $\leq 10^{-6}$) and difference in signal means of at least 1.5 $\log_2$ units. Comparisons with both flanking intervals were made separately. Features were considered non-coding only if no overlap was observed with annotated coding regions on the sense strand. Non-coding features were further classified as "antisense" if any part of the interval overlapped a known coding feature on the opposite strand. tRNAs and rRNAs were implicitly excluded due to the ambiguous mapping of corresponding array probes. Other non-CDS features such as snoRNAs were removed post-hoc and served as an internal control for feature identification.

For all other samples examined, normalized intensities for the vegetative asynchronous hybridizations were subtracted from each additional experiments' values. The resulting difference in hybridization signal was segmented and categorized as above. Genes induced (compared to asynchronous vegetative cells as reference) have positive scores; regions of no relative change have a net signal of zero. Induced non-coding features were identified as above.

Figures were plotted using custom written code and modifications to software previously described (Huber et al., 2006). For plots displaying a single value/sample/probe, the arithmetic mean of all replicates is shown. All values plotted are normalized and background corrected signals; differences in expression between samples are not directly shown except where explicitly mentioned. For composite expression plots, probe intensities were mapped to each nucleotide flanking the transcription start sites (5' feature edge, as assigned above), and the average intensity across all relevant probes and features is shown. Overlap and smoothing between probes occurs, due to nominal 20nt spacing of 25nt probes. In contrast, heatmap plots represent the 3' most bases for overlapping probes; where overlap exists, the signal for only one probe is shown. Due to high probe density and the resolution at which figures are presented, this is largely irrelevant. Data from Dutrow et al. (2008) were obtained from the authors' publicly available website.

Analysis of expression data was performed on composite data sets for both *S. pombe* and *S. cerevisiae.* Fission yeast data was obtained from public data of the Fission Yeast Functional Genomics group (University College London, and Sanger Centre, UK) and data available from our own *S. pombe* hybridizations. Gene name ambiguities were resolved manually in consultation with GeneDB locus history (Sanger Centre, Hixton UK). *S. cerevisiae* expression data was obtained from the *Saccharomyces* Genome Database (Stanford University, Palo Alto, CA) curated experiment collection, public expression experiments from the Stanford Microarray Database (Stanford University, Palo Alto, CA), and a number of expression hybridizations performed locally. Gene name ambiguities were virtually absent from these disparate sets, with the notable exception of recently added dubious or uncharacterized short features, most of whom lack coverage on available expression arrays. These features were excluded from further analysis. Analysis of expression profiles was performed en masse. Pairwise Pearson correlations were calculated for all gene pairs (generating ~$6000^2/2$-6000 pairwise scores for *S. cerevisiae*, for example). Standard score of each gene's correlation its adjacent genes (both 5' and 3' proximal) was calculated using all Pearson correlations for a given gene as the background set. Variance of the original $\log_2$ expression score was calculated for each gene. All calculations were performed within the R platform for statistical computing.

4 Antisense transcription suppresses meiotic gene expression in fission yeast

A major reason for examining the transcriptome structure of *S. pombe* was to unravel the regulatory mechanisms controlling entry in and progression through the meiotic program. Meiosis has historically been a low-hanging fruit of transcription studies, and indeed served as one of the first expression programs examined on arrays for both budding (Chu et al., 1998; Primig et al., 2000) and fission (Mata et al., 2002) yeast . In both yeasts, meiotic functions can be separated into three transcriptional waves, consisting of 1) early genes are responsible for pre-meiotic S, including meiotic cohesins and factors involved in meiotic recombination, 2) middle genes are responsible for the meiotic divisions, and 3) late genes are largely responsible for assembly and maturation of the spore (Chu et al., 1998).

Previous work in the lab had focused on regulation of meiotic entry, and paralleling my work on transcription during the vegetative cell cycle was a study of transcriptional regulation of early meiotic genes. In 2005, the lab published a pilot study of meiotic genes which suggested extensive regulation of the meiotic program by differential splicing (Averbeck et al., 2005). The single-gene PCR based methods used were not amenable to high-throughput use, making systematic detection of regulated splicing difficult. Along with a fellow graduate student in the lab, I undertook a high-resolution tiling array analysis of fission yeast meiosis as a means for detecting differences in splicing through sporulation.

The results were stunning and quite unexpected. Rather than validating regulated splicing events and identifying a large number of new regulated genes, we identified extensive antisense transcription throughout the fission yeast genome, with a significant bias toward meiotically restricted genes. While quantitating splicing of the relatively short *S. pombe* introns was difficult, few changes appeared to be occurring. The implications of my initial findings were sweeping. Many of the unspliced signals identified by Averbeck et al. were amplifications of antisense transcripts, calling into question the significance and extent of regulated splicing in *S. pombe.* Antisense transcription was relatively widespread in *S. pombe,* but had previously been overlooked; most coding regions were assumed to generate a single transcript from a single strand. Finally, the genes with associated transcripts in vegetative cells were restricted to a relatively small number of biological functions, suggesting a possible regulatory role for antisense transcription.

Results

Previous work in the lab examined a subset of meiotically induced genes and determined that corresponding unspliced transcripts were present in vegetative cells for 10% of the genes tested. As cells progress through meiosis, fully spliced transcripts predominate for these genes, suggesting differential splicing for a functionally related set of genes. The twelve genes identified as "splice regulated" peak in different waves of the meiotic program; three represent the early program, eight middle genes were identified, as was a single late meiotic gene. Using high-resolution strand specific tiling microarrays (Affymetrix), I am unable to validate unspliced sense transcript in vegetative cells for many of these genes, but find presence of significant antisense transcripts for all of the middle- and late- genes (figure 35). Transcripts originating from the opposite strand lack appropriate splicing signals, but are detected by the RT-PCR based approach used by Averbeck et al. and are indistinguishable from unspliced sense RNAs.

I have examined transcriptome structure in vegetative cells grown in minimal media and at 2h intervals throughout a pat1-114 synchronized meiosis (figure 36). In veg-

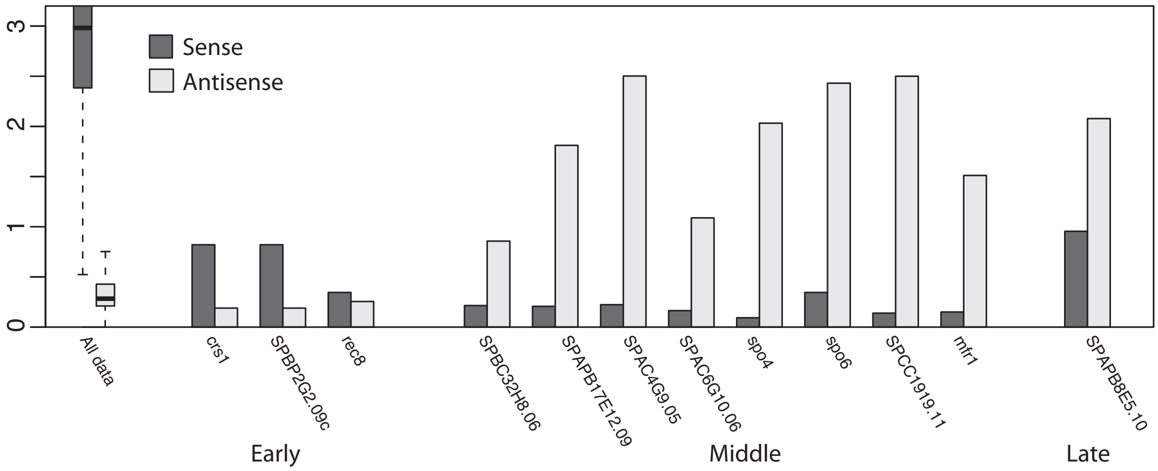


Figure 35: All mid-meiotic "splice regulated" genes have strong antisense transcripts in vegetative cells

Many genes previously identified as "splice regulated" by Averbeck et al. (2005) have strong antisense transcripts in mitotic (vegetative) cells. Strand-specific expression scores from high-resolution tiling microarray data are shown for each gene; shown at left are box-and-whisker plots representing the sense and antisense scores for all genes.

Early meiotic genes have significant sense transcript signal in vegetative cells; antisense transcripts for these genes are not significant. Conversely, all of the middle meiotic genes which were considered splice regulated are strongly transcribed from the antisense strand. Without exception, sense transcript level for these genes is at or near background level. Sense transcript for the sole late "splice regulated" gene identified, SPAPB8E5.10, is present in vegetative cells, but overshadowed by strong antisense signal which persists throughout meiosis. Induction of sense transcript occurs at 6-8h, consistent with the appearance of "spliced" message at that time.

etative cells, antisense transcripts are present for a large number of genes. I find that genes with strong corresponding antisense transcripts span a bimodal distribution relative to sense expression level. A large number of low-sense, high antisense transcripts have been identified, and at the opposite extreme, processed antisense transcripts are relatively abundant for the most strongly expressed genes. Antisense transcripts corresponding to strongly expressed genes are perhaps inadvertently polyadenylated by a high local concentration of Pol II associated cleavage and maturation factors.

Unlike budding yeast, *S. pombe* encodes an entire complement of RNAi machinery, including the RNA directed RNA polymerase, Rdp1. Processed antisense transcripts described herein are not generated by Rdp1, as antisense transcription in an rdp1Δ strain is largely unchanged from wild type cells (Pearson correlation 0.86, data not shown).

Previous examination of total RNA in fission yeast described a strong correlation between sense and antisense transcript levels, and furthermore an association between antisense abundance and Pol II occupancy (Dutrow et al., 2008). With the exception of the most strongly expressed genes, this correlation between expression from both strands is not observed in our analysis of processed transcripts (see figure 37 on page 136). Further examination of the data from Dutrow et al. supports the idea that sense-correlated expression from the antisense strand results in largely unprocessed species, as little correlation between sense and antisense signal is seen for poly(A) enriched RNAs (see figure 38 on page 137).

Genes with high antisense, low sense expression in vegetative cells disproportionately represent members of the Mei4 dependent mid-meiotic transcription program. 38% of the genes in this category have been previously identified as targets of Mei4p (Mata et al., 2002), compared to less than 9% of genes with substantial sense expression in vegetative cells. Genes involved in meiotic divisions are specifically over-

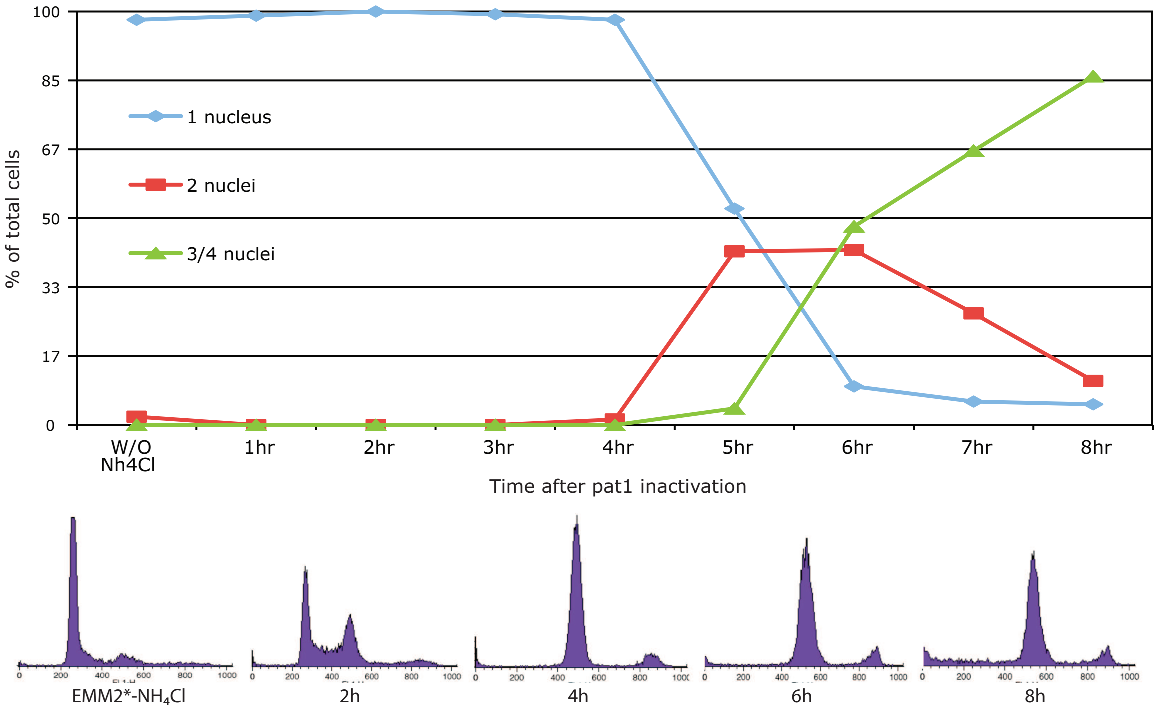


Figure 36: Meiotic timecourse progression

Progression through a pat1-114 induced meiosis is quantitated by DAPI-stained nuclei counts. Synchronous meiosis was induced in cells arrested in G_1 by nitrogen starvation. Cells were re-fed nitrogen containing medium pre-warmed to restrictive temperature at T_0.

At 2h, pre-meiotic S is occurring. 4h, beginning of first meiotic division. 6h, second meiotic division. At 8h, spore packaging is underway.

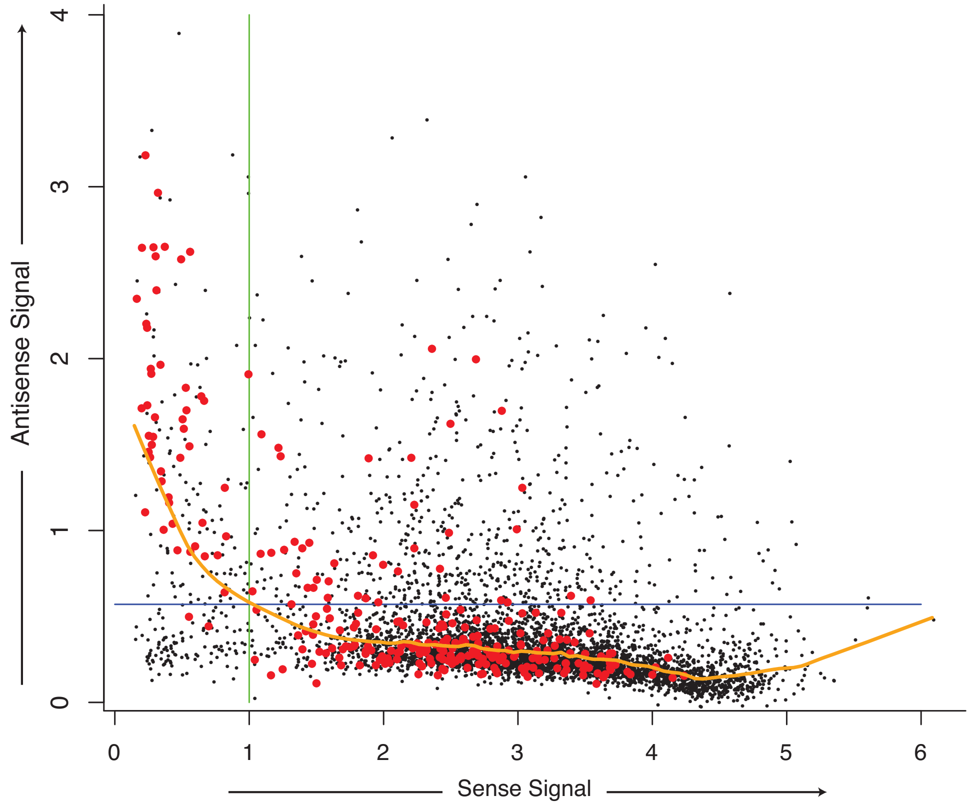


Figure 37: Many genes with strong antisense signal are Mei4-responsive meiotic transcripts

Strand specific expression signals from asynchronous vegetative cells are shown. Sense expression levels vary across a wide range, although sense transcripts are present well above background for most genes. Antisense transcripts are abundant at both extrema of expression; many genes with weak sense expression are abundantly transcribed on the antisense strand. Local trendline is shown in orange.

Significant antisense transcripts are present for many genes, and a disproportionate number of genes with low sense expression have high antisense signals (upper left). Blue bar represents 4σ above antisense background.

Previously identified Mei4 dependent mid-meiotic genes (Mata et al. (2002), shown as red points) are over-represented in the "low-sense, high-antisense" class, but are specifically absent from the "low-sense, low-antisense" class. A significant fraction of this group is unannotated, and may represent vestigial untranscribed ORFs.

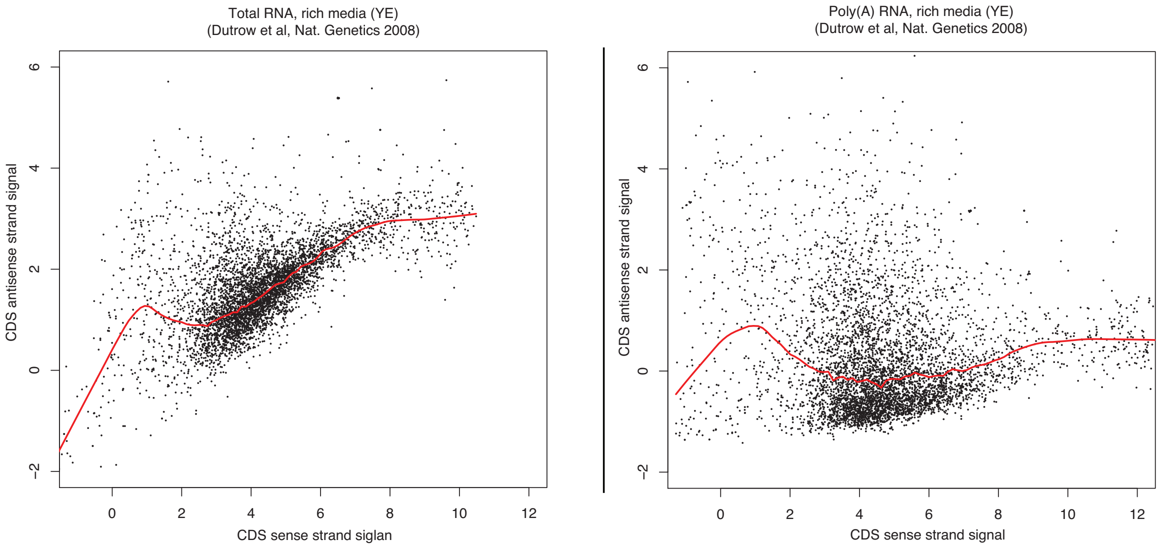


Figure 38: Previously identified sampling transcripts are unprocessed

A positive correlation between sense expression and abundance of antisense transcripts was noted in previously (Dutrow et al., 2008), and is shown (left). Increased chromatin accessibility or local concentration of components of the transcription apparatus may permit initiation of spurious transcripts in a sense-expression dependent manner.

These sampling transcripts are not polyadenylated, however, and are therefore excluded from my data (see figure 24 on page 109 for further details). The majority of antisense "sampling" transcripts are absent from a poly(A) enriched fraction (right), revealing a smaller set of processed transcripts generated from both strands.

represented in this group by ontology term analysis, with a p-value of $4.8 \mathrm{x} 10^{-13}$. Presence of genes in this class does not reflect an overall lack of expression in vegetative cells; meiotic genes are explicitly absent from the “low-sense, low-antisense” class (see figure 37).

Transcription from sense and antisense strands is mutually antagonistic. Given the overrepresentation of strongly regulated meiotic genes among strong-antisense transcripts, I postulated a negative regulatory role for opposite-strand transcription. This is supported by changes in transcript abundance during meiosis. Induced sense transcription is associated with a decrease in antisense transcript level. Conversely, a decrease in sense strand expression is correlated with an increase in antisense transcript level (see figure 39 on the next page). While not mechanistically demonstrative, this finding illustrates a negative correlation between expression of sense and antisense strands of a given locus. This effect is especially evident for Mei4p responsive genes, although this set is already enriched for transcripts which are induced across the interval shown.

Transcription of the antisense strand impedes expression of meiotic genes in vegetative cells. We have disrupted antisense transcription to demonstrate the mechanistic relevance of antisense transcription on repressing gene expression. spo6, the meiotic counterpart of Dfp1, is strongly induced in mid meiosis in a Mei4p dependent manner (Mata et al., 2002), and as a previously identified “splice-regulated” gene (Averbeck et al., 2005) is strongly transcribed on the antisense strand in vegetative cells (see figure 35 on page 133). spo6 antisense is generated as a long 3’ UTR of the convergently orientated adjacent sequence orphan, SPBC1778.05c. Truncation of this transcript and elimination of antisense transcription through spo6 is sufficient to permit spo6 sense expression in vegetative cells (see figure 40 on page 140).

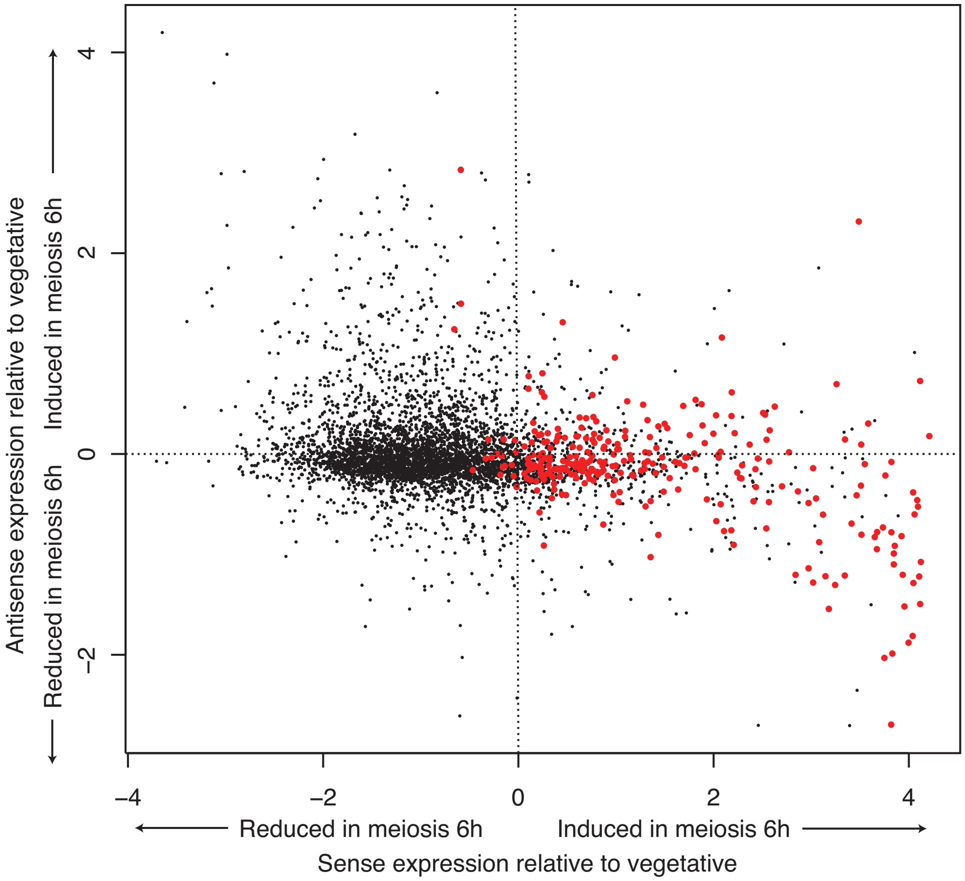


Figure 39: Antisense transcription antagonizes sense expression
Changes in sense expression are anticorrelated with changes in antisense transcript levels. Differences in strand-specific mean expression level between the 6h meiotic and vegetative samples are shown. Genes previously identified as Mei4p responsive, shown in red, are generally induced by 6h. Decreases in antisense transcript level accompany increases in sense expression for these genes; Pearson correlation for Mei4p responsive genes is -0.47. Increased antisense level is associated with decreased sense transcript, as seen in upper left; Pearson correlation for all other genes is -0.34.

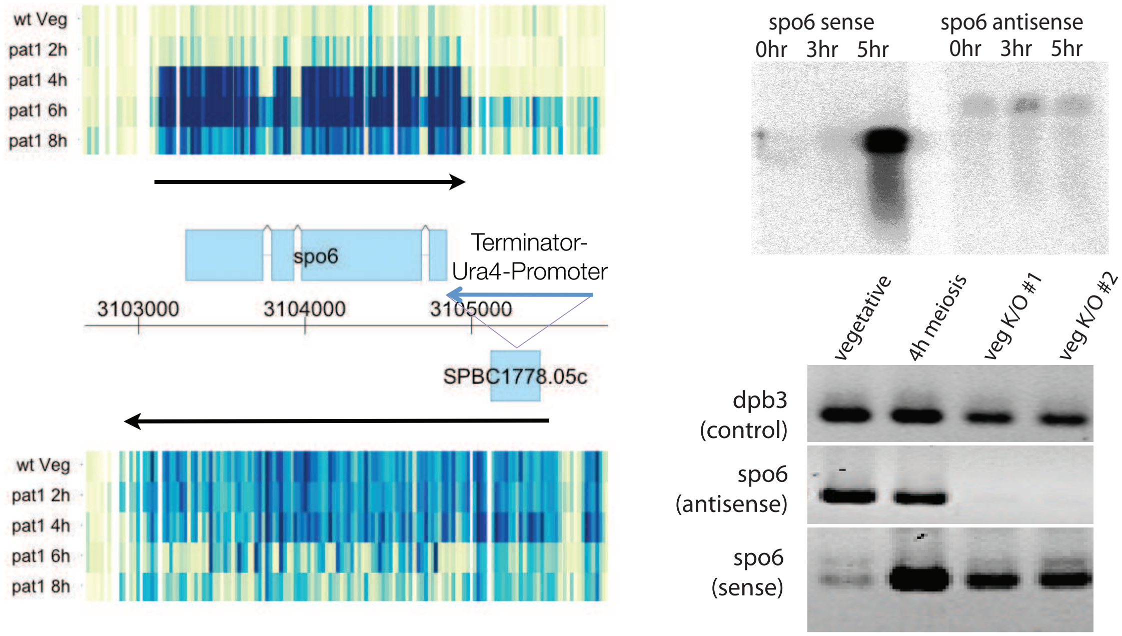


Figure 40: Disruption of spo6 antisense transcript permits ectopic sense expression
Antisense transcription through spo6 is the result of long readthrough from the adjacent sequence orphan, SPBC1778.05c. Northern blot (top right) using single-stranded riboprobes against spo6 demonstrates that contiguous antisense transcript observed on tiling arrays (left) is a single transcript, and is longer than the spo6 sense transcript. In vegetative cells, spo6 antisense transcript is relatively abundant, while spo6 sense transcript is absent. Induction of spo6 sense transcription occurs between 2h and 4h in this synchronized meiosis, and continues well through 6h. Results from tiling array data are recapitulated by strand specific PCR (bottom right) for vegetative and meiotic cells. The relatively invariant dpb3 is shown as a control.
Insertion of a terminator-flanked Ura4$^+$ cassette into the adjacent sequence orphan precludes transcription of spo6 antisense. In absence of antisense transcription, spo6 sense-strand expression is increased. Isolates from two independent integrations of the terminator cassette are shown.

Conclusions

Although antisense transcripts to meiotic genes were serendipitously identified more than twenty years ago(Kishida and Shimoda, 1986), this work represents the first systematic strand-specific quantitation of processed transcripts. I have demonstrated that antisense transcripts are present for a large number of genes, and specifically enriched for a common functional class expressed in preparation for the meiotic divisions. I have further shown that the previously described mechanism of regulated splicing appears to be restricted to early meiotic genes, and that previously identified unspliced messages present in vegetative cells are antisense transcripts. The regulatory potential of antisense transcription is general has been supported by anti-correlation between sense and antisense expression changes, and mechanistic evidence has been generated for spo6, itself a key meiotic regulator.

Furthermore, I suggest that antisense transcriptional interference is one of several layers of regulation for meiotic genes. Removal of repressive antisense transcription does not result in expression of spo6 at meiotically induced levels. Regulation of spo6 occurs through several mechanisms, including competition for binding of upstream sequence motifs between the inhibitory vegetative forkhead transcription factor Fkh2 and the related meiotic transcription Mei4p (especially relevant since mei4 is present at low levels in vegetative cells). In fkh2Δ-mei4Δ double mutant cells, spo6 sense expression level is increased to levels greater than that of 2h into meiosis; this increase is seen in cells where spo6 antisense is being transcribed, and as such reflects exclusively the inhibitory role of Fkh2p presence. These overlapping mechanisms likely reflect an evolutionary impetus to avoid ectopic expression of genes involved in the meiotic cell cycle.

Antisense-mediated reduction of sense expression is not specific to spo6; disruptions of antisense for several additional genes has been performed, and suggests that this is a widespread regulatory mechanism. Examination of the combinatorial effects of vegetative transcriptional interference and transcription factor competition are ongoing, and will be published shortly in collaboration with a fellow graduate student in the lab, Huei-Mei Chen.

Methods

Induction of synchronous meiosis is described above on page 125. Samples were collected every 2h following shift to restrictive temperature, and cDNAs were prepared for tiling array hybridization in technical triplicate. Data were normalized and background corrected as above, and CDS expression scores for Affymetrix data were generated as described previously. In addition to the meiotic time course described, technical duplicate hybridizations were performed on poly(dT) primed cDNAs generated from asynchronous cells derived from additional strains, including rdp1Δ, mei4Δ/fkh2Δ, fkh2Δ, and dis3-54 (following 2h incubation at restrictive temperature).

5 Perspectives

5.1 Experimental conclusions and their relevance in the field

Analysis of the vegetative cell cycle of *S. pombe* was performed using arrays which were, at the time, cutting edge. From these data, hundreds of cell cycle regulated genes were identified, and DNA binding sites for many components of a mitotic transcriptional oscillator were identified or refined. Our examination of cell cycle transcription was predicated on several assumptions. First, we measured transcriptional almost exclusively from coding sequences; expression of unannotated regions of the genome and those without protein-coding potential was largely ignored. Second, the 3' end of the coding sequence was used as a proxy for what was assumed to be a monolithic transcript; edges of transcripts which could contribute to differential translation capacity were not identified, nor were possible regulated splicing events. Finally, expression measurements were not strand-specific in nature; antisense transcription was indistinguishable from sense expression. At the time, many of these limitations were inevitable. The generation of expression array probes was a monumental task for the lab, and limited array density dictated the number of features which could be packed onto a single substrate. The lack of strand specificity and inability to distinguish different transcript isoforms clearly did not, however, overwhelm the overall findings of our study; the utility of the arrays was self evident in the biologically relevant results obtained.

Subsequent studies in a number of organisms, including fission yeast, have demonstrated vast complexity of the transcriptome which is belied by traditional expression arrays. Regulated splicing, although perhaps not as widespread in *S. pombe* as previously believed, has been shown to play a key role in regulation of expression. Clearly, many potentially overlapping mechanisms have evolved to fine-tune cellular pools of translation-competent message. I have shown, above, the regulatory power of antisense transcription, and have identified many genes for which antisense transcripts exist. While well suited as a baseline for comparison to the meiotic transcription program, many details pertinent to the regulation of the mitotic cell cycle have likely been overlooked. My "vegetative" reference represents the composite expression patterns of asynchronous cells in all stages of the cell cycle; given the sweeping changes in transcriptional programs between various cell-cycle stages and the plethora of regulatory mechanisms used by the cell to ensure compartmentalization of expression and function, it seems nearly inevitable that transcriptional interference and antisense transcription will be among the tools employed. Expression of non-coding RNAs has similarly been studied exclusively in asynchronous cells. While I have provided strong evidence that nearly half of the non-coding species transcribed in *S. pombe* are byproducts of genic transcription, many of the possibly regulatory functions of hundreds of transcripts remain unexplored.

Cross-study comparisons and the bounty of meta-analysis

All of the analyses presented in this dissertation have incorporated significant amounts of data from outside sources. Due to the timing of our fission yeast cell-cycle transcription study, significant effort was put forth in the analysis of our results in the context of a similar parallel study. Far from providing repetition or simple confirmation, the data sets were complementary such that their integrated analysis permitted more powerful conclusions than could be made from any single study; several of the most informative cluster-discriminating arrays were obtained from previous studies.

Re-analysis of published data provides significant weight to my study of non-coding RNAs in *S. pombe*. My data, positionally dense, but only covering a few conditions, were far more informative in the context of more than two-thousand individual expression arrays which were re-analyzed. Similarly, high-resolution tiling array data from budding yeast was re-analyzed to answer my own biological questions, fundamentally different from those for which the data had been published. While my findings stand independent of these outside datasets, they are integral to drawing broadly applicable conclusions.

My analysis of antisense transcription in fission yeast also draws from outside datasets. Arrays which were present solely as supplementary data to a previous publication have provided insight into a different population of transcripts than I had set out to study; their orthogonal data provided information on material which would have gone unnoticed using my original methods. Integration of outside data has enabled not only the inclusion of experiments that would have been performed locally given infinite time and resources, but also tangential data that I would have never considered generating.

Similarly, data which I have generated are all deposited in public repositories, and provide a rich source for re-analysis by other labs. Data from our vegetative cell cycle experiments have been analyzed both in context of the other published fission yeast experiments (Marguerat et al., 2006) and across species in an attempt to define an evolutionarily conserved set of cycling genes (Lu et al., 2007). Continued requirement for public data deposition will ensure that similar meta-analyses can occur from a wide variety of data sets.

The importance of context in transcription

Eukaryotic genomes contain far more information than the sum of their included coding sequences. Genomic context plays multiple roles in the regulation of expression and transcription, in turn, shapes organization of the genome. The majority of large-scale transcription studies which have been performed have focused on individual transcripts as independent entities, where grouping of transcripts most often occurs by similarities in expression profiles and ignores critical spatial relationships. Recent studies, including the work presented here, have begun to examine transcription in spatial context of neighboring genes, and provide compelling reason to do so.

Recent work in budding yeast has illuminated an additional layer to genome organization by analyzing transcription in the context of nucleosome occupancy (Whitehouse et al., 2007; Neil et al., 2009). We have identified bidirectional transcription initiating from regions consistent with a single nucleosome free region, and nearly half of the discrete non-coding transcripts observed in vegetative fission yeast can be attributed to divergent transcription around a single promoter. More than half of the non-coding RNAs have no obvious source, but are demarcated by gene-like 5' edges. Although these may represent intentional transcription of sequences which have so far escaped functional assignment, another possibility is their genesis as unintentional byproducts of chromatin which has been made accessible for reasons other than genic transcription. Examination of non-coding RNAs in context of replication origins may provide significant insight into the chromatin changes which have been previously recognized at such sites (Zhou et al., 2005). Genome-wide analysis of nucleosome occupancy in fission yeast has not been published as of yet, but is underway in collaboration with Joel Huberman, Roswell Park Cancer Center.

It is unclear to what extent relatively simple yeast genomes can serve as a model for organization of mammalian transcription units. At a local level, similar conservation of divergently transcribed genes has been observed, and several recent studies have demonstrated generation of non-coding transcripts which may function in an isolating capacity similar to those described above (Core et al., 2008; Seila et al., 2008; Affymetrix ENCODE Transcriptome Project and Cold Spring Harbor Laboratory ENCODE Transcriptome Project, 2009). Structure of individual genes is far more complex in metazoa, adding multiple layers of potential regulation which are seemingly absent in yeast, such as trans splicing or extensive alternate exon usage. Despite these potential limitations, the roles of orientation and location dependent co-regulation presented above are likely conserved not only between yeasts, but to higher organisms as well.

Transcriptional interference was observed early in the study of transcription, but has been largely ignored in genome wide studies. This is somewhat surprising given the regulatory potential of interference and the elegant examples of regulatory circuits, such as those controlling *IME4*, *SER33*, and *fbp1*, and regulation of the meiotic expression program shown above. These examples are likely just the “tip of the iceberg”, and strongly warrant re-examination of transcripts, and the action of transcription itself, in the context of adjacent genes.

The effects of new technologies on our understanding of transcription

Widespread transcription of both coding and non-coding genomic regions has been recognized for more than thirty years. It has not been until the last decade that advances in sequence detection and quantitation have permitted the in-depth determination of a genome’s output. Early descriptions of genomic architecture were based on the apparent output of discrete transcription units, whose functions could seemingly be dissected and recapitulated in chimeric constructs. Although gene organization into divergent, convergent, or tandem arrays was recognized from early studies with the potential for interference of adjacent transcripts described not long thereafter, transcriptional context and production of transcripts from both template strands were not fully appreciated until the last few years.

From afar, the reductionist view of a genome as a set of separate open reading frames is still valid. The ability to quantitate expression of every nucleotide in the genome does not reduce the utility of expression arrays; for many purposes, the focused data which can be generated from carefully planned “one probe per gene” designs is significantly easier to analyze and more cost effective to generate than brute-force tiling arrays. In this sense, tiling arrays and high throughput sequencing serve a complementary role to traditional expression microarrays. Novel features identified by these unbiased approaches should be included in expression array designs. In return, expression data encompassing a wide breadth of experiments can be mined in context of the genomic fine structure (as presented in 3.3 on page 107 above).

As new technologies become available, there is often a strong backlash against established competing tools; this has been especially true in the context of high throughput sequencing as an alternative to microarrays. The overall utility of next-generation sequencing is of no doubt, and its increasingly widespread adoption has permitted great strides in many areas including small RNA biology, novel transcript identification, and genome sequencing. Sequencing has not, however, superseded other forms of transcript detection and quantitation.

Despite ever-increasing throughput, “next-generation” sequencing is as of yet incapable of covering the same dynamic range as can an array. Although millions of sequence reads are generated from each instrument “run”, they are stochastically spread across transcripts ranging from less than one to thousands of copies per cell. Furthermore, samples to be run on current generation instruments require significant wet-lab processing, often including many rounds of PCR amplification. The results of less-than-ideal workflows can be clearly seen above in figure 23 and figure 32. Strand non-specific methods ignore known instances of transcription from both strands, and preclude the identification of unexpected transcript structure.

Proponents of RNA-seq often highlight the purported benefits of “digital” sequencing output over the “analog” results generated by microarrays. Serious biases generated by sample preparation notwithstanding, this comparison belies the complexity of the underlying issue. A digital clock which displays only the current hour is far less precise than the analog spring-driven sweeping second hand of a pocketwatch. Examined from another angle, the output from microarrays is effectively continuous, with subtle changes in concentration recorded as changes across large (2^{20}) potential values. Each element on an array is able to measure concentration of its complementary sequence; specificity is limited by probe length and density, individual elements respond independently. In stark contrast, sequencing provides a quantized readout of sequence abundance, but does so with a significant limitation. Ability to detect rare sequences is directly dependent on number of reads performed, and the complexity of the total sample. Several comparisons between sequencing and high-density array data have been presented above. In each case, the arrays provided a more consistent quantitation of expression for moderately expressed genes, and were able to more accurately quantitate lower abundance messages than possible by sequencing. Dynamic range and the ability to detect low-abundance messages is critical even in simple transcriptomes such as budding yeast, where expression varies greater than 2^{13} fold (Miura et al., 2008). Refinement and advancement of both tiling arrays and deep sequencing technology is nearly assured, and improvements to these complementary technologies will permit greater insight into transcription.

5.2 Future Directions

Detection of chromatin/protein interactions

In addition to measurement of transcriptional output, high throughput sequencing and tiling microarrays can be employed in the detection of protein-DNA binding interactions, as ChIP-seq and ChIP-on-chip. Although I have identified DNA sequence motifs enriched upstream of many clusters of cell cycle regulated genes, these sequences are often found at lower quality or density throughout the genome. Affinity tagged transcription factors can be selectively immunoprecipitated with bound DNAs, and a map of *bona fide* binding sites generated for each factor in each condition, cell cycle phase, for example. Analysis of one or a small number of factors would not only provide validation of predicted binding sites and determination of predicted although unused sites, but would also allow the refinement of sequence motifs by the examining relative enrichment of various sequence classes. When many (or eventually all) transcription factors' binding profiles have been analyzed, a far more informative combinatorial analysis can be performed; cooperative interactions between different factors, well known to occur by single gene studies, can be determined genome-wide. Although many spurious binding events of single factors likely occur throughout the genome, the requirement for factor-factor interactions may obviate selection against isolated sites.

High throughput protein-nucleic acid interaction data will also provide substantial insight into the origin and fate of non-coding RNAs. Association of 3' processing factors with promoters is expected to occur only upstream of stable non-coding species, and can be tested by genome-wide ChIP. Similarly, RNA-protein interactions of the TRAMP complex are expected at the 3' ends of unstable non-coding RNAs (and indeed, many other sites genome-wide). HiTS-CLIP (Licatalosi et al., 2008) could be performed to directly identify all TRAMP targets, complementing data which represent accumulation of targets following TRAMP inactivation. Far from supplanting detection of transcripts, these methods will provide orthogonal data which will permit further sub-classification of transcript classes.

Combining strands, space, and time: antisense transcription across the cell cycle

Cell-cycle oscillating transcripts have been characterized in *S. pombe* by three independent groups, as described above. From these studies, it has become clear that abundance of many, if not all, transcripts vary as cells cycle. I have also demonstrated several hitherto hidden layers of transcriptional complexity in fission yeast, namely the abundance of extragenic transcripts and the presence of many antisense transcripts with potentially widespread regulatory function. Advances in array manufacture and sample labeling since Oliva et al. are substantial. Using existing Affymetrix tiling arrays, one could examine the cell cycle at relatively high temporal resolution in a strand specific, high spatial resolution fashion. Due to input RNA requirements, sampling of only a few times across a synchronous cycle would be possible, but would provide invaluable snapshots of strand specific expression in the context of previously characterized transcription programs.

An alternative approach provides middle ground between transcriptome mapping and our published strand-nonspecific study. High-density arrays comprising hundreds of thousands of *ad hoc* strand-specific 60nt oligonucleotide probes are now the basis of our array facility. Arrays can be designed to target sense and antisense strands of previously identified genes and novel features (both those identified in this work, and any subsequent studies, as above), in an effectively unlimited iterative process. RNA requirements for these arrays are exceedingly low; fractions of a μg of total RNA is required, permitting analysis of limiting sample volume, or increased sampling density. Using such arrays, one could profile sense and antisense transcription of coding and non-coding transcripts throughout literally scores of times in the vegetative cell cycle using fewer total cells than a single Affymetrix tiling array. Decreased input requirements also permit examination of limiting material, for example several cell-cycles worth of samples from a single elutriation fraction.

Cross-species comparisons in the high-throughput age

Significant insights have been gained by comparisons across phylogeny both diverse and local. A central justification for model organism use is their relevance to other systems. Comparisons across wide evolutionary distances have been performed in the context of my work, such as in the conservation of modules of transcribed genes and their associated regulators, and the associations non-coding RNAs with various classes of promoters. Although not as far-reaching, significant insight can be gained by more “local” examination across species. Genome sequences for many *Saccharomycetes* and other closely related yeasts have been published, and have provided an invaluable resource for a variety of studies, not limited to identification promoter motifs (as pioneered in Kellis et al. (2003)), determination of neutral mutation rate and gene copy correction (Pyne et al., 2005), and functional elements of (non-coding) telomerase RNAs (Gunisova et al., 2009).

Genome sequences for other *Schizosaccharomyctes* are currently being finished by the Broad Institute, and will permit similar analyses for fission yeast. In addition to *in silico* approaches such as promoter motif discovery, the availability of custom-designed arrays and high throughput sequencing permits direct experimental investigation or validation using these other species. The once onerous task of array manufacture has been replaced by custom-designed tools on demand, with far lower total cost. In a similar hyperbolic example, one could sequence a new species’ genome and generate transcriptome data in the same instrument run with no prior information.

Cross species comparisons have direct applications to the projects listed above. For cell cycle regulated genes, functional sequence motifs will be more strongly conserved across species than will be vestigial or false positive sites. Comparison of cell-cycle regulated genes in *S. japonicus, S. octosporus*, or *S. cryophobus* would permit refinement of “cell cycle regulated genes” and moreover provide significant insight into direct and indirect oscillators. Analysis of non-coding RNAs across species will likely separate functional species with conserved sequence from incidental transcripts of irrelevant content. The significance of antisense transcription can also be well addressed by comparisons with evolutionarily close species. Adjacent genes engendering regulatory antisense transcripts should be more conserved than expected by the neutral model. Direct measurement of antisense transcript levels in related species will also help to distinguish active transcriptional interference from incidental expression of the opposite strand.

Transcript end mapping and measurement of transcription proper As evidenced by this work, expression and tiling arrays are powerful tools for the detection and classification of transcripts. Despite this great utility, some shortcomings exist, of which several have been addressed by new methods. Samples to be interrogated on arrays generally represent the sum total of transcripts in the cell at a given moment. Subcellular fractionation or isolation of polysomes can provide evidence for differential localization or active translation, respectively, but do not address a question underlying *regulation* of transcription. Arrays represent the steady state RNA levels in a cell, confounding measurements of transcriptional activation with the instantaneous background of extant transcripts. Transcription rate and elongation speed can be determined by nuclear run-on, but the procedure is laborious, low throughput, and inherently difficult to scale. A recent publication by the John Lis' lab has demonstrated an immuno-enrichment modification to the classic run-on procedure, which enables capture and subsequent quantitation of actively transcribed species (Core et al., 2008). Although the method as published generated sequencing-ready libraries, identification of nascent transcripts should be possible on arrays as well; two color arrays could be employed to compare instantaneous transcription rate to steady state, generating indirect RNA half-life measurements *en masse*.

A fundamental limitation of microarrays is the identification of sequence connectivity. While sensitive and accurate measurement of sequences corresponding to individual probes is possible, no information about connectivity *between* sequences is generated. In higher eukaryotes, the potential for alternative exon use and long-distance trans splicing make unambiguous assignment of a probe to a given transcript difficult (Kapranov et al., 2005; Willingham et al., 2006), and identification of individual (often overlapping) transcript ends is difficult, even in yeast. Traditional "long read" sequencing of vector-capped cDNAs has been performed in both budding yeast and mouse, but is a relatively low-throughput technology. Identification of individual 5' or 3' transcript edges is possible using standard high-throughput sequencing linker ligation, but discards information specific to each pair of ends. Paired end sequencing or full-length RNAs, while conceptually attractive, falls short due to strong amplicon-length biases during library preparation and sequencing. Paired end tag sequencing provides SAGE-like data for both 5' and 3' terminii of a single message in a homogeneous tag read length, and was the backbone of mouse transcriptome sequencing (see above). Development of high-throughput sequencing-formatted PET tag libraries will undoubtedly enable new efforts of transcript-end mapping by sequencing (Fullwood et al., 2009).

5.3 Summary

During my tenure in the lab, I have been fortunate to be involved in a wide variety of projects, adoption of new techniques and technologies, and have witnessed fundamental changes in the study of transcription. My attraction to yeast as a model system was largely due to the tractability of the system where, in the currency of the day, the entire genome could fit onto a single slide. Techniques for genetic manipulation of yeast were well beyond those available in higher eukaryotes, and the assumption that yeast served as an excellent model for inaccessible metazoan systems appeared sound. In the intervening years, huge strides have been made in both the genetics tools available for metazoa (with notable advances in retro/lentiviral vectors and various titrable inducible expression systems, shRNAs, siRNAs, and integrable miRNA knockdown systems), and the techniques with which we examine transcript structure and abundance (increased density of tiling arrays and the development of deep sequencing systems). Paralleling these advances have been a number of publications (referenced throughout this dissertation) which demonstrate mechanistic conservation across broadly divergent model organisms. Far from highlighting the differences between yeast and metazoa, progress during these years has underscored the continued relevance of both budding and fission yeast to the scientific gestalt.

The study of regulated transcription has expanded from analysis of discrete functional sequence elements to the examination of expression in both local and genomic context. Gene expression is controlled not only through production of functional transcripts, but also through transcription itself. Advances in expression measurement technology and the availability of large data sets have permitted integrated analysis of the actively transcribed genome as a whole.

References

Adhya, S. and Gottesman, M. (1982). Promoter occlusion: transcription through a promoter may inhibit its activity. *Cell*, 29(3):939–944.

Affymetrix ENCODE Transcriptome Project and Cold Spring Harbor Laboratory ENCODE Transcriptome Project (2009). Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. *Nature*, 457(7232):1028–32.

Albers, E., Laize, V., Blomberg, A., Hohmann, S., and Gustafsson, L. (2003). Ser3p (Yer081wp) and Ser33p (Yil074cp) are phosphoglycerate dehydrogenases in Saccharomyces cerevisiae. *J Biol Chem*, 278(12):10264–10272.

Alvarez, B., Martinez-A, C., Burgering, B. M., and Carrera, A. C. (2001). Forkhead transcription factors contribute to execution of the mitotic programme in mammals. *Nature*, 413(6857):744–747.

Alvarez, F. J., Douglas, L. M., Rosebrock, A., and Konopka, J. B. (2008). The Sur7 protein regulates plasma membrane organization and prevents intracellular cell wall growth in Candida albicans. *Mol Biol Cell*, 19(12):5214–25.

Alwine, J. C., Kemp, D. J., and Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A*, 74(12):5350–4.

Anderson, M., Ng, S. S., Marchesi, V., MacIver, F. H., Stevens, F. E., Riddell, T., Glover, D. M., Hagan, I. M., and McInerny, C. J. (2002). Plo1(+) regulates gene transcription at the M-G(1) interval during the fission yeast mitotic cell cycle. *EMBO J*, 21(21):5745–5755.

Attwooll, C., Lazzerini Denchi, E., and Helin, K. (2004). The E2F family: specific functions and overlapping interests. *EMBO J*, 23(24):4709–4716.

Averbeck, N., Sunder, S., Sample, N., Wise, J. A., and Leatherwood, J. (2005). Negative control contributes to an extensive program of meiotic splicing in fission yeast. *Mol Cell*, 18(4):491–8.

Ayte, J., Leis, J. F., Herrera, A., Tang, E., Yang, H., and DeCaprio, J. A. (1995). The Schizosaccharomyces pombe MBF complex requires heterodimerization for entry into S phase. *Mol Cell Biol*, 15(5):2589–2599.

Bailey, T. L. and Elkan, C. (1995). The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol*, 3:21–29.

Bailey, T. L. and Gribskov, M. (1998). Methods and statistics for combining motif match scores. *J Comput Biol*, 5(2):211–221.

Ball, C., Brazma, A., Causton, H., Chervitz, S., Edgar, R., Hingamp, P., Matese, J. C., Parkinson, H., Quackenbush, J., Ringwald, M., Sansone, S.-A., Sherlock, G., Spellman, P., Stoeckert, C., Tateno, Y., Taylor, R., White, J., Winegarden, N., and MGED Society (2004). Standards for microarray data: an open letter. *Environ Health Perspect*, 112(12):A666–7.

Ball, C. A., Sherlock, G., Parkinson, H., Rocca-Sera, P., Brooksbank, C., Causton, H. C., Cavalieri, D., Gaasterland, T., Hingamp, P., Holstege, F., Ringwald, M., Spellman, P., Stoeckert, Jr, C. J., Stewart, J. E., Taylor, R., Brazma, A., Quackenbush, J., and Microarray Gene Expression Data (MGED) Society (2002a). Standards for microarray data. *Science*, 298(5593):539.

Ball, C. A., Sherlock, G., Parkinson, H., Rocca-Sera, P., Brooksbank, C., Causton, H. C., Cavalieri, D., Gaasterland, T., Hingamp, P., Holstege, F., Ringwald, M., Spellman, P., Stoeckert, Jr, C. J., Stewart, J. E., Taylor, R., Brazma, A., Quackenbush, J., and Microarray Gene Expression Data (2002b). The underlying principles of scientific publication. *Bioinformatics*, 18(11):1409.

Barrientos, A. (2003). Yeast models of human mitochondrial diseases. *IUBMB Life*, 55(2):83–95.

Baum, B., Nishitani, H., Yanow, S., and Nurse, P. (1998). Cdc18 transcription and proteolysis couple S phase to passage through mitosis. *EMBO J*, 17(19):5689–5698.

Benito, J., Martín-Castellanos, C., and Moreno, S. (1998). Regulation of the G1 phase of the cell cycle by periodic stabilization and degradation of the p25rum1 CDK inhibitor. *EMBO J*, 17(2):482–97.

Berk, A. J. and Sharp, P. A. (1977). Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell*, 12(3):721–32.

Bracken, A. P., Ciro, M., Cocito, A., and Helin, K. (2004). E2F target genes: unraveling the biology. *Trends Biochem Sci*, 29(8):409–417.

Brazma, A. and Vilo, J. (2001). Gene expression data analysis. *Microbes Infect*, 3(10):823–829.

Breeden, L. L. (2003). Periodic transcription: a cycle within a cycle. *Curr Biol*, 13(1):R31–8.

Buck, V., Ng, S. S., Ruiz-Garcia, A. B., Papadopoulou, K., Bhatti, S., Samuel, J. M., Anderson, M., Millar, J. B. A., and McInerny, C. J. (2004a). Fkh2p and Sep1p regulate mitotic gene transcription in fission yeast. *J Cell Sci*, 117(Pt 23):5623–5632.

Bühler, M., Haas, W., Gygi, S. P., and Moazed, D. (2007). RNAi-dependent and -independent RNA turnover mechanisms contribute to heterochromatic gene silencing. *Cell*, 129(4):707–21.

Bulmer, R., Pic-Taylor, A., Whitehall, S. K., Martin, K. A., Millar, J. B. A., Quinn, J., and Morgan, B. A. (2004). The forkhead transcription factor Fkh2 regulates the cell division cycle of Schizosaccharomyces pombe. *Eukaryot Cell*, 3(4):944–954.

Burbelo, P. D., Martin, G. R., and Yamada, Y. (1988). Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer. *Proc Natl Acad Sci U S A*, 85(24):9679–82.

Carlson, C. R., Grallert, B., Stokke, T., and Boye, E. (1999). Regulation of the start of DNA replication in Schizosaccharomyces pombe. *J Cell Sci*, 112 (Pt 6):939–946.

Carlsson, P. and Mahlapuu, M. (2002). Forkhead transcription factors: key players in development and metabolism. *Dev Biol*, 250(1):1–23.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V. B., Brenner, S. E., Batalov, S., Forrest, A. R. R., Zavolan, M., Davis, M. J., Wilming, L. G., Aidinis, V., Allen, J. E., Ambesi-Impiombato, A., Apweiler, R., Aturaliya, R. N., Bailey, T. L., Bansal, M., Baxter, L., Beisel, K. W., Bersano, T., Bono, H., Chalk, A. M., Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R., Crowe, M. L., Dalla, E., Dalrymple, B. P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C. F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T. R., Gojobori, T., Green, R. E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T. K., Hirokawa, N., Hill, D., Huminiecki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S. P. T., Kruger, A., Kummerfeld, S. K., Kurochkin, I. V., Lareau, L. F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakauchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K. C., Pavan, W. J., Pavesi, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J. F., Ring, B. Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S. L., Sandelin, A., Schneider, C., Schönbach, C., Sekiguchi, K., Semple, C. A. M., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugiura, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S. L., Tang, S., Taylor, M. S., Tegner, J., Teichmann, S. A., Ueda, H. R., van Nimwegen, E., Verardo, R., Wei, C. L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S. M., Teasdale, R. D., Liu, E. T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J. S., Hume, D. A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J., Hayashizaki, Y., FANTOM Consortium, and RIKEN Genome Exploration Research Group and Genome Science Group (Genome

Network Project Core Group) (2005). The transcriptional landscape of the mammalian genome. *Science*, 309(5740):1559–63.

Cawley, S., Bekiranov, S., Ng, H. H., Kapranov, P., Sekinger, E. A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A. J., Wheeler, R., Wong, B., Drenkow, J., Yamanaka, M., Patel, S., Brubaker, S., Tammana, H., Helt, G., Struhl, K., and Gingeras, T. R. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, 116(4):499–509.

Cervantes, M. D., Farah, J. A., and Smith, G. R. (2000). Meiotic DNA breaks associated with recombination in S. pombe. *Mol Cell*, 5(5):883–8.

Chekanova, J. A., Gregory, B. D., Reverdatto, S. V., Chen, H., Kumar, R., Hooker, T., Yazaki, J., Li, P., Skiba, N., Peng, Q., Alonso, J., Brukhin, V., Grossniklaus, U., Ecker, J. R., and Belostotsky, D. A. (2007). Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. *Cell*, 131(7):1340–53.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., Sementchenko, V., Piccolboni, A., Bekiranov, S., Bailey, D. K., Ganesh, M., Ghosh, S., Bell, I., Gerhard, D. S., and Gingeras, T. R. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, 308(5725):1149–54.

Cho, R. J., Huang, M., Campbell, M. J., Dong, H., Steinmetz, L., Sapinoso, L., Hampton, G., Elledge, S. J., Davis, R. W., and Lockhart, D. J. (2001). Transcriptional regulation and function during the human cell cycle. *Nat Genet*, 27(1):48–54.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., and Herskowitz, I. (1998). The transcriptional program of sporulation in budding yeast. *Science*, 282(5389):699–705.

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K. S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell*, 9(2):279–289.

Cirillo, L. A., McPherson, C. E., Bossard, P., Stevens, K., Cherian, S., Shim, E. Y., Clark, K. L., Burley, S. K., and Zaret, K. S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J*, 17(1):244–254.

Cirillo, L. A. and Zaret, K. S. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell*, 4(6):961–969.

Collinge, J. and Clarke, A. R. (2007). A general model of prion strains and their pathogenicity. *Science*, 318(5852):930–936.

Core, L. J., Waterfall, J. J., and Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 322(5909):1845–8.

Cosma, M. P., Tanaka, T., and Nasmyth, K. (1999). Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell*, 97(3):299–311.

Costello, G., Rodgers, L., and Beach, D. (1986). Fission yeast enters the stationary phase G0 state from either mitotic G1 or G2. *Current Genetics*, 11(2):119–125.

Creanor, J. and Mitchison, J. M. (1982). Patterns of protein synthesis during the cell cycle of the fission yeast Schizosaccharomyces pombe. *J Cell Sci*, 58:263–285.

Creanor, J. and Mitchison, J. M. (1984). Protein synthesis and its relation to the DNA-division cycle in the fission yeast Schizosaccharomyces pombe. *J Cell Sci*, 69:199–210.

Culotti, J. and Hartwell, L. H. (1971). Genetic control of the cell division cycle in yeast. 3. Seven genes controlling nuclear division. *Exp Cell Res*, 67(2):389–401.

Daga, R. R. and Jimenez, J. (1999). Translational control of the cdc25 cell cycle phosphatase: a molecular mechanism coupling mitosis to cell growth. *J Cell Sci*, 112 Pt 18:3137–46.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., Davis, R. W., and Steinmetz, L. M. (2006). A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A*, 103(14):5320–5.

Dekker, N., Speijer, D., Grun, C. H., van den Berg, M., de Haan, A., and Hochstenbach, F. (2004). Role of the alpha-glucanase Agn1p in fission-yeast cell separation. *Mol Biol Cell*, 15(8):3903–3914.

Delcher, A. L., Phillippy, A., Carlton, J., and Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res*, 30(11):2478–83.

Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F., and Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. *Nature*, 432(7014):231–5.

DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278(5338):680–6.

Dutrow, N., Nix, D. A., Holt, D., Milash, B., Dalley, B., Westbroek, E., Parnell, T. J., and Cairns, B. R. (2008). Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping. *Nat Genet*, 40(8):977–86.

Dutta, C., Patel, P. K., Rosebrock, A., Oliva, A., Leatherwood, J., and Rhind, N. (2008). The DNA replication checkpoint directly regulates MBF-dependent G1/S transcription. *Mol Cell Biol*, 28(19):5977–85.

Dziembowski, A., Lorentzen, E., Conti, E., and Séraphin, B. (2007). A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nat Struct Mol Biol*, 14(1):15–22.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A*, 95(25):14863–14868.

Ekwall, K. (2004). The roles of histone modifications and small RNA in centromere function. *Chromosome Res*, 12(6):535–542.

Elliott, S. G. (1983a). Coordination of growth with cell division: regulation of synthesis of RNA during the cell cycle of the fission yeast Schizosaccharomyces pombe. *Mol Gen Genet*, 192(1-2):204–211.

Elliott, S. G. (1983b). Regulation of the maximal rate of RNA synthesis in the fission yeast Schizosaccharomyces pombe. *Mol Gen Genet*, 192(1-2):212–217.

ENCODE Project Consortium, NISC Comparative Sequencing Program, Baylor College of Medicine Human Genome Sequencing Center, Washington University Genome Sequencing Center, Broad Institute, and Children's Hospital Oakland Research Institute (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146):799–816.

Forsburg, S. L. (2003). Introduction of DNA into S. pombe cells. *Curr Protoc Mol Biol*, Chapter 13:Unit 13.17.

Fraser, R. S. and Nurse, P. (1978). Novel cell cycle control of RNA synthesis in yeast. *Nature*, 271(5647):726–730.

Fraser, R. S. and Nurse, P. (1979). Altered patterns of ribonucleic acid synthesis during the cell cycle: a mechanism compensating for variation in gene concentration. *J Cell Sci*, 35:25–40.

Fullwood, M. J., Wei, C.-L., Liu, E. T., and Ruan, Y. (2009). Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res*, 19(4):521–32.

Futcher, B. (1999). Cell cycle synchronization. *Methods Cell Sci*, 21(2-3):79–86.

Futcher, B. (2002). Transcriptional regulatory networks and the yeast cell cycle. *Curr Opin Cell Biol*, 14(6):676–683.

Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S., and Garrels, J. I. (1999). A sampling of the yeast proteome. *Mol Cell Biol*, 19(11):7357–7368.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y. H., and Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, 5(10):R80.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. (1996). Life with 6000 genes. *Science*, 274(5287):546, 563–7.

Gottesfeld, J. M. and Forbes, D. J. (1997). Mitotic repression of the transcriptional machinery. *Trends Biochem Sci*, 22(6):197–202.

Gottesfeld, J. M., Wolf, V. J., Dang, T., Forbes, D. J., and Hartl, P. (1994). Mitotic repression of RNA polymerase III transcription in vitro mediated by phosphorylation of a TFIIIB component. *Science*, 263(5143):81–84.

Gottesman, S. (2005). Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet*, 21(7):399–404.

Gould, K. L. and Nurse, P. (1989). Tyrosine phosphorylation of the fission yeast cdc2+ protein kinase regulates entry into mitosis. *Nature*, 342(6245):39–45.

Grigull, J., Mnaimneh, S., Pootoolal, J., Robinson, M. D., and Hughes, T. R. (2004). Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. *Mol Cell Biol*, 24(12):5534–5547.

Gunisova, S., Elboher, E., Nosek, J., Gorkovoy, V., Brown, Y., Lucier, J.-F., Laterreur, N., Wellinger, R. J., Tzfati, Y., and Tomaska, L. (2009). Identification and comparative analysis of telomerase RNAs from Candida species reveal conservation of functional elements. *RNA*, 15(4):546–59.

Hanahan, D. and Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, 100(1):57–70.

Hannon, G. J., Rivas, F. V., Murchison, E. P., and Steitz, J. A. (2006). The expanding universe of noncoding RNAs. *Cold Spring Harb Symp Quant Biol*, 71:551–64.

Hansen, K. R., Burns, G., Mata, J., Volpe, T. A., Martienssen, R. A., Bähler, J., and Thon, G. (2005). Global effects on gene expression in fission yeast by silencing and RNA interference machineries. *Mol Cell Biol*, 25(2):590–601.

Harigaya, Y., Tanaka, H., Yamanaka, S., Tanaka, K., Watanabe, Y., Tsutsumi, C., Chikashige, Y., Hiraoka, Y., Yamashita, A., and Yamamoto, M. (2006). Selective elimination of messenger RNA prevents an incidence of untimely meiosis. *Nature*, 442(7098):45–50.

Harris, P., Kersey, P. J., McInerny, C. J., and Fantes, P. A. (1996). Cell cycle, DNA damage and heat shock regulate suc22+ expression in fission yeast. *Mol Gen Genet*, 252(3):284–291.

Hartl, P., Gottesfeld, J., and Forbes, D. J. (1993). Mitotic repression of transcription in vitro. *J Cell Biol*, 120(3):613–624.

Hartwell, L. H. (1971a). Genetic control of the cell division cycle in yeast. II. Genes controlling DNA replication and its initiation. *J Mol Biol*, 59(1):183–194.

Hartwell, L. H. (1971b). Genetic control of the cell division cycle in yeast. IV. Genes controlling bud emergence and cytokinesis. *Exp Cell Res*, 69(2):265–276.

Hartwell, L. H. (2004). Yeast and cancer. *Biosci Rep*, 24(4-5):523–544.

Hartwell, L. H., Culotti, J., and Reid, B. (1970). Genetic control of the cell-division cycle in yeast. I. Detection of mutants. *Proc Natl Acad Sci U S A*, 66(2):352–359.

He, F., Li, X., Spatrick, P., Casillo, R., Dong, S., and Jacobson, A. (2003). Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Mol Cell*, 12(6):1439–52.

He, Y., Vogelstein, B., Velculescu, V. E., Papadopoulos, N., and Kinzler, K. W. (2008). The antisense transcriptomes of human cells. *Science*, 322(5909):1855–7.

Heckman, D. S., Geiser, D. M., Eidell, B. R., Stauffer, R. L., Kardos, N. L., and Hedges, S. B. (2001). Molecular evidence for the early colonization of land by fungi and plants. *Science*, 293(5532):1129–1133.

Hess, D. C., Myers, C. L., Huttenhower, C., Hibbs, M. A., Hayes, A. P., Paw, J., Clore, J. J., Mendoza, R. M., Luis, B. S., Nislow, C., Giaever, G., Costanzo, M., Troyanskaya, O. G., and Caudy, A. A. (2009). Computationally driven, quantitative experiments discover genes required for mitochondrial biogenesis. *PLoS Genet*, 5(3):e1000407.

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(7218):130–134.

Hong, S.-P., Leiper, F. C., Woods, A., Carling, D., and Carlson, M. (2003). Activation of yeast Snf1 and mammalian AMP-activated protein kinase by upstream kinases. *Proc Natl Acad Sci U S A*, 100(15):8839–8843.

Hongay, C. F., Grisafi, P. L., Galitski, T., and Fink, G. R. (2006). Antisense transcription controls cell fate in Saccharomyces cerevisiae. *Cell*, 127(4):735–745.

Hough, B. R., Smith, M. J., Britten, R. J., and Davidson, E. H. (1975). Sequence complexity of heterogeneous nuclear RNA in sea urchin embryos. *Cell*, 5(3):291–9.

Hu, P., Samudre, K., Wu, S., Sun, Y., and Hernandez, N. (2004). CK2 phosphorylation of Bdp1 executes cell cycle-specific RNA polymerase III transcription repression. *Mol Cell*, 16(1):81–92.

Huang, Y. (2002). Transcriptional silencing in Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Nucleic Acids Res*, 30(7):1465–1482.

Huber, W., Toedling, J., and Steinmetz, L. M. (2006). Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics*, 22(16):1963–70.

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18 Suppl 1:S96–104.

Hughes, A. L., Powell, D. W., Bard, M., Eckstein, J., Barbuch, R., Link, A. J., and Espenshade, P. J. (2007). Dap1/PGRMC1 binds and regulates cytochrome P450 enzymes. *Cell Metab*, 5(2):143–149.

Hughes, A. L., Todd, B. L., and Espenshade, P. J. (2005). SREBP pathway responds to sterols and functions as an oxygen sensor in fission yeast. *Cell*, 120(6):831–842.

Hughes, J. D., Estep, P. W., Tavazoie, S., and Church, G. M. (2000). Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae. *J Mol Biol*, 296(5):1205–1214.

Hunt, T. (2002). Nobel Lecture. Protein synthesis, proteolysis, and cell cycle transitions. *Biosci Rep*, 22(5-6):465–486.

Ideker, T., Thorsson, V., Siegel, A. F., and Hood, L. E. (2000). Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data. *J Comput Biol*, 7(6):805–17.

Iino, Y., Hiramine, Y., and Yamamoto, M. (1995). The role of cdc2 and other genes in meiosis in Schizosaccharomyces pombe. *Genetics*, 140(4):1235–45.

Imamura, T., Yamamoto, S., Ohgane, J., Hattori, N., Tanaka, S., and Shiota, K. (2004). Non-coding RNA directed DNA demethylation of Sphk1 CpG island. *Biochem Biophys Res Commun*, 322(2):593–600.

Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., and Speed, T. P. (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*, 31(4):e15.

Irwin, B., Aye, M., Baldi, P., Beliakova-Bethell, N., Cheng, H., Dou, Y., Liou, W., and Sandmeyer, S. (2005). Retroviruses and yeast retrotransposons use overlapping sets of host genes. *Genome Res*, 15(5):641–654.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, 3:318–56.

Jin, P., Gu, Y., and Morgan, D. O. (1996). Role of inhibitory CDC2 phosphorylation in radiation-induced G2 arrest in human cells. *J Cell Biol*, 134(4):963–970.

Johnson, T. C. and Holland, J. J. (1965). Ribonucleic acid and protein synthesis in mitotic HeLa cells. *J Cell Biol*, 27(3):565–574.

Jorgensen, P., Rupes, I., Sharom, J. R., Schneper, L., Broach, J. R., and Tyers, M. (2004). A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size. *Genes Dev*, 18(20):2491–2505.

Jourdain, I., Sontam, D., Johnson, C., Dillies, C., and Hyams, J. S. (2008). Dynamin-dependent biogenesis, cell cycle regulation and mitochondrial association of peroxisomes in fission yeast. *Traffic*, 9(3):353–365.

Kapranov, P., Cawley, S. E., Drenkow, J., Bekiranov, S., Strausberg, R. L., Fodor, S. P. A., and Gingeras, T. R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. *Science*, 296(5569):916–9.

Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S., and Gingeras, T. R. (2005). Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. *Genome Res*, 15(7):987–97.

Kato, M., Hata, N., Banerjee, N., Futcher, B., and Zhang, M. Q. (2004). Identifying combinatorial regulation of transcription factors and binding motifs. *Genome Biol*, 5(8):R56.

Kato, S., Ohtoko, K., Ohtake, H., and Kimura, T. (2005). Vector-capping: a simple method for preparing a high-quality full-length cDNA library. *DNA Res*, 12(1):53–62.

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(6937):241–54.

Kim, V. N., Han, J., and Siomi, M. C. (2009). Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol*, 10(2):126–39.

Kishida, M. and Shimoda, C. (1986). Genetic mapping of eleven spo genes essential for ascospore formation in the fission yeast Schizosaccharomyces pombe. *Curr Genet*, 10(6):443–7.

Kitamura, K., Katayama, S., Dhut, S., Sato, M., Watanabe, Y., Yamamoto, M., and Toda, T. (2001). Phosphorylation of Mei2 and Ste11 by Pat1 kinase inhibits sexual differentiation via ubiquitin proteolysis and 14-3-3 protein in fission yeast. *Dev Cell*, 1(3):389–99.

Klein, J. and Grummt, I. (1999). Cell cycle-dependent regulation of RNA polymerase I transcription: the nucleolar transcription factor UBF is inactive in mitosis and early G1. *Proc Natl Acad Sci U S A*, 96(11):6096–6101.

Koch, C., Moll, T., Neuberg, M., Ahorn, H., and Nasmyth, K. (1993). A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase. *Science*, 261(5128):1551–1557.

Koch, C. and Nasmyth, K. (1994a). Cell cycle regulated transcription in yeast. *Curr Opin Cell Biol*, 6(3):451–459.

Kshirsagar, M. and Parker, R. (2004). Identification of Edc3p as an enhancer of mRNA decapping in Saccharomyces cerevisiae. *Genetics*, 166(2):729–39.

Laub, M. T., McAdams, H. H., Feldblyum, T., Fraser, C. M., and Shapiro, L. (2000). Global analysis of the genetic network controlling a bacterial cell cycle. *Science*, 290(5499):2144–8.

Lawrence, C. E., Altschul, S. F., Boguski, M. S., Liu, J. S., Neuwald, A. F., and Wootton, J. C. (1993). Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science*, 262(5131):208–14.

Lazarow, P. B. (1995). Peroxisome structure, function, and biogenesis–human patients and yeast mutants show strikingly similar defects in peroxisome biogenesis. *J Neuropathol Exp Neurol*, 54(5):720–725.

Lee, M. G. and Nurse, P. (1987). Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. *Nature*, 327(6117):31–35.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 75(5):843–54.

Legras, J.-L., Merdinoglu, D., Cornuet, J.-M., and Karst, F. (2007). Bread, beer and wine: Saccharomyces cerevisiae diversity reflects human history. *Mol Ecol*, 16(10):2091–2102.

Leresche, A., Wolf, V. J., and Gottesfeld, J. M. (1996). Repression of RNA polymerase II and III transcription during M phase of the cell cycle. *Exp Cell Res*, 229(2):282–288.

Li, Y.-Y., Yu, H., Guo, Z.-M., Guo, T.-Q., Tu, K., and Li, Y.-X. (2006). Systematic analysis of head-to-head gene organization: evolutionary conservation and potential biological relevance. *PLoS Comput Biol*, 2(7):e74.

Licatalosi, D. D., Mele, A., Fak, J. J., Ule, J., Kayikci, M., Chi, S. W., Clark, T. A., Schweitzer, A. C., Blume, J. E., Wang, X., Darnell, J. C., and Darnell, R. B. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, 456(7221):464–9.

Liu, Q., Greimann, J. C., and Lima, C. D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell*, 127(6):1223–37.

Lo, H.-C., Wan, L., Rosebrock, A., Futcher, B., and Hollingsworth, N. M. (2008). Cdc7-Dbf4 regulates NDT80 transcription as well as reductional segregation during budding yeast meiosis. *Mol Biol Cell*, 19(11):4956–67.

Lowndes, N. F., McInerny, C. J., Johnson, A. L., Fantes, P. A., and Johnston, L. H. (1992). Control of DNA synthesis genes in fission yeast by the cell-cycle gene cdc10+. *Nature*, 355(6359):449–453.

Lu, Y., Mahony, S., Benos, P. V., Rosenfeld, R., Simon, I., Breeden, L. L., and Bar-Joseph, Z. (2007). Combined analysis reveals a core set of cycling genes. *Genome Biol*, 8(7):R146.

Lycan, D. E., Osley, M. A., and Hereford, L. M. (1987). Role of transcriptional and post-transcriptional regulation in expression of histone genes in Saccharomyces cerevisiae. *Mol Cell Biol*, 7(2):614–621.

Madeo, F., Herker, E., Wissing, S., Jungwirth, H., Eisenberg, T., and Frohlich, K.-U. (2004). Apoptosis in yeast. *Curr Opin Microbiol*, 7(6):655–660.

Marguerat, S., Jensen, T. S., de Lichtenberg, U., Wilhelm, B. T., Jensen, L. J., and Bähler, J. (2006). The more the merrier: comparative analysis of microarray studies on cell cycle-regulated genes in fission yeast. *Yeast*, 23(4):261–77.

Martens, J. A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature*, 429(6991):571–574.

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., and Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*, 445(7128):666–670.

Martin-Cuadrado, A. B., Duenas, E., Sipiczki, M., Vazquez de Aldana, C. R., and del Rey, F. (2003). The endo-beta-1,3-glucanase eng1p is required for dissolution of the primary septum during cell separation in Schizosaccharomyces pombe. *J Cell Sci*, 116(Pt 9):1689–1698.

Mata, J., Lyne, R., Burns, G., and Bähler, J. (2002). The transcriptional program of meiosis and sporulation in fission yeast. *Nat Genet*, 32(1):143–7.

Matsumoto, S. and Yanagida, M. (1985). Histone gene organization of fission yeast: a common upstream sequence. *EMBO J*, 4(13A):3531–3538.

Matthews, C. G. and Lott, F. E. (1889). *The Microscope in the Brewery and Malt-House*. Bemrose and Sons, 23 Old Bailey and Derby.

Mattick, J. S. (2005). The functional genomics of noncoding RNA. *Science*, 309(5740):1527–8.

Mazo, A., Hodgson, J. W., Petruk, S., Sedkov, Y., and Brock, H. W. (2007). Transcriptional interference: an unexpected layer of complexity in gene regulation. *J Cell Sci*, 120(Pt 16):2755–61.

McGovern, P. E., Zhang, J., Tang, J., Zhang, Z., Hall, G. R., Moreau, R. A., Nunez, A., Butrym, E. D., Richards, M. P., Wang, C.-S., Cheng, G., Zhao, Z., and Wang, C. (2004). Fermented beverages of pre- and proto-historic China. *Proc Natl Acad Sci U S A*, 101(51):17593–17598.

Meluh, P. B. and Koshland, D. (1997). Budding yeast centromere composition and assembly as revealed by in vivo cross-linking. *Genes Dev*, 11(24):3401–3412.

Mickle, K. L., Ramanathan, S., Rosebrock, A., Oliva, A., Chaudari, A., Yompakdee, C., Scott, D., Leatherwood, J., and Huberman, J. A. (2007). Checkpoint independence of most DNA replication origins in fission yeast. *BMC Mol Biol*, 8:112.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'–>5' exoribonucleases. *Cell*, 91(4):457–66.

Mitchison, J. M. and Nurse, P. (1985). Growth in cell length in the fission yeast Schizosaccharomyces pombe. *J Cell Sci*, 75:357–376.

Mitchison, J. M., Sveiczer, A., and Novak, B. (1998). Length growth in fission yeast: is growth exponential?–No. *Microbiology*, 144 (Pt 2):265–266.

Miura, F., Kawaguchi, N., Sese, J., Toyoda, A., Hattori, M., Morishita, S., and Ito, T. (2006). A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc Natl Acad Sci U S A*, 103(47):17846–51.

Miura, F., Kawaguchi, N., Yoshida, M., Uematsu, C., Kito, K., Sakaki, Y., and Ito, T. (2008). Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs. *BMC Genomics*, 9:574.

Mizuno, T., Chou, M. Y., and Inouye, M. (1984). A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc Natl Acad Sci U S A*, 81(7):1966–70.

Moll, T., Tebb, G., Surana, U., Robitsch, H., and Nasmyth, K. (1991). The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the S. cerevisiae transcription factor SWI5. *Cell*, 66(4):743–58.

Moore, M. J. (2002). Nuclear RNA turnover. *Cell*, 108(4):431–4.

Mouse Genome Sequencing Consortium (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915):520–62.

Muller, D., Exler, S., Aguilera-Vazquez, L., Guerrero-Martin, E., and Reuss, M. (2003). Cyclic AMP mediates the cell cycle dynamics of energy metabolism in Saccharomyces cerevisiae. *Yeast*, 20(4):351–367.

Müller, W. E., Zahn, R. K., and Seidel, H. J. (1971). Inhibitors acting on nucleic acid synthesis in an oncogenic RNA virus. *Nat New Biol*, 232(31):143–5.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881):1344–9.

Nasmyth, K. (1986). A U-turn in the regulation of transcription? *Trends in Genetics*, 2:115–116.

Neil, H., Malabat, C., d'Aubenton Carafa, Y., Xu, Z., Steinmetz, L. M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, 457(7232):1038–42.

Neuwald, A. F., Liu, J. S., and Lawrence, C. E. (1995). Gibbs motif sampling: detection of bacterial outer membrane protein repeats. *Protein Sci*, 4(8):1618–32.

Ng, P., Wei, C.-L., Sung, W.-K., Chiu, K. P., Lipovich, L., Ang, C. C., Gupta, S., Shahab, A., Ridwan, A., Wong, C. H., Liu, E. T., and Ruan, Y. (2005). Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nat Methods*, 2(2):105–11.

Niwa, O., Matsumoto, T., and Yanagida, M. (1986). Construction of a mini-chromosome by deletion and its mitotic and meiotic behaviour in fission yeast. *Molecular and General Genetics MGG*, 203(3):397–405.

Nurse, P., Thuriaux, P., and Nasmyth, K. (1976). Genetic control of the cell division cycle in the fission yeast Schizosaccharomyces pombe. *Mol Gen Genet*, 146(2):167–178.

Nurse, P. M. (2002). Nobel Lecture. Cyclin dependent kinases and cell cycle control. *Biosci Rep*, 22(5-6):487–499.

O'Conallain, C., Doolin, M. T., Taggart, C., Thornton, F., and Butler, G. (1999). Regulated nuclear localisation of the yeast transcription factor Ace2p controls expression of chitinase (CTS1) in Saccharomyces cerevisiae. *Mol Gen Genet*, 262(2):275–82.

Ohtake, H., Ohtoko, K., Ishimaru, Y., and Kato, S. (2004). Determination of the capped site sequence of mRNA based on the detection of cap-dependent nucleotide addition using an anchor ligation method. *DNA Res*, 11(4):305–9.

Oliva, A., Rosebrock, A., Ferrezuelo, F., Pyne, S., Chen, H., Skiena, S., Futcher, B., and Leatherwood, J. (2005). The cell cycle-regulated genes of Schizosaccharomyces pombe. *PLoS Biol*, 3(7):e225.

Outeiro, T. F. and Muchowski, P. J. (2004). Molecular genetics approaches in yeast to study amyloid diseases. *J Mol Neurosci*, 23(1-2):49–60.

Paddison, P. J., Caudy, A. A., Sachidanandam, R., and Hannon, G. J. (2004). Short hairpin activated gene silencing in mammalian cells. *Methods Mol Biol*, 265:85–100.

Paddison, P. J. and Hannon, G. J. (2002). RNA interference: the new somatic cell genetics? *Cancer Cell*, 2(1):17–23.

Parsons, G. G. and Spencer, C. A. (1997). Mitotic repression of RNA polymerase II transcription is accompanied by release of transcription elongation complexes. *Mol Cell Biol*, 17(10):5791–5802.

Patel, P. K., Kommajosyula, N., Rosebrock, A., Bensimon, A., Leatherwood, J., Bechhoefer, J., and Rhind, N. (2008). The Hsk1(Cdc7) replication kinase regulates origin efficiency. *Mol Biol Cell*, 19(12):5550–8.

Peng, X., Karuturi, R. K. M., Miller, L. D., Lin, K., Jia, Y., Kondu, P., Wang, L., Wong, L.-S., Liu, E. T., Balasubramanian, M. K., and Liu, J. (2005). Identification of cell cycle-regulated genes in fission yeast. *Mol Biol Cell*, 16(3):1026–1042.

Perocchi, F., Mancera, E., and Steinmetz, L. M. (2008). Systematic screens for human disease genes, from yeast to human and back. *Mol Biosyst*, 4(1):18–29.

Perocchi, F., Xu, Z., Clauder-Münster, S., and Steinmetz, L. M. (2007). Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res*, 35(19):e128.

Pramila, T., Miles, S., GuhaThakurta, D., Jemiolo, D., and Breeden, L. L. (2002). Conserved homeodomain proteins interact with MADS box protein Mcm1 to restrict ECB-dependent transcription to the M/G1 phase of the cell cycle. *Genes Dev*, 16(23):3034–3045.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M. S., Mapendano, C. K., Schierup, M. H., and Jensen, T. H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science*, 322(5909):1851–4.

Price, C., Nasmyth, K., and Schuster, T. (1991). A general approach to the isolation of cell cycle-regulated genes in the budding yeast, Saccharomyces cerevisiae. *J Mol Biol*, 218(3):543–56.

Primig, M., Williams, R. M., Winzeler, E. A., Tevzadze, G. G., Conway, A. R., Hwang, S. Y., Davis, R. W., and Esposito, R. E. (2000). The core meiotic transcriptome in budding yeasts. *Nat Genet*, 26(4):415–23.

Pyne, S., Skiena, S., and Futcher, B. (2005). Copy correction and concerted evolution in the conservation of yeast genes. *Genetics*, 170(4):1501–13.

Reymond, A., Schmidt, S., and Simanis, V. (1992). Mutations in the cdc10 start gene of Schizosaccharomyces pombe implicate the region of homology between cdc10 and SWI6 as important for p85cdc10 function. *Mol Gen Genet*, 234(3):449–456.

Ruprecht, R. M., Goodman, N. C., and Spiegelman, S. (1973). Conditions for the selective synthesis of DNA complementary to template RNA. *Biochim Biophys Acta*, 294(2):192–203.

Rustici, G., Mata, J., Kivinen, K., Lio, P., Penkett, C. J., Burns, G., Hayles, J., Brazma, A., Nurse, P., and Bahler, J. (2004). Periodic gene expression program of the fission yeast cell cycle. *Nat Genet*, 36(8):809–817.

Salditt-Georgieff, M., Harpold, M. M., Wilson, M. C., and Darnell, Jr, J. E. (1981). Large heterogeneous nuclear ribonucleic acid has three times as many 5' caps as polyadenylic acid segments, and most caps do not enter polyribosomes. *Mol Cell Biol*, 1(2):179–87.

Schmid, M. and Jensen, T. H. (2008). The exosome: a multipurpose RNA-decay machine. *Trends Biochem Sci*, 33(10):501–10.

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A., and Sharp, P. A. (2008). Divergent transcription from active promoters. *Science*, 322(5909):1849–51.

Shah, J. C. and Clancy, M. J. (1992). IME4, a gene that mediates MAT and nutritional control of meiosis in Saccharomyces cerevisiae. *Mol Cell Biol*, 12(3):1078–1086.

Sheth, U. and Parker, R. (2003). Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science*, 300(5620):805–8.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajska, A., Harbers, M., Kawai, J., Carninci, P., and Hayashizaki, Y. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A*, 100(26):15776–81.

Silva, J. M., Mizuno, H., Brady, A., Lucito, R., and Hannon, G. J. (2004). RNA interference microarrays: high-throughput loss-of-function genetics in mammalian cells. *Proc Natl Acad Sci U S A*, 101(17):6548–6552.

Simon, I., Barnett, J., Hannett, N., Harbison, C. T., Rinaldi, N. J., Volkert, T. L., Wyrick, J. J., Zeitlinger, J., Gifford, D. K., Jaakkola, T. S., and Young, R. A. (2001). Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell*, 106(6):697–708.

Simons, R. W. and Kleckner, N. (1983). Translational control of IS10 transposition. *Cell*, 34(2):683–91.

Spector, M. S. and Osley, M. A. (1993). The HIR4-1 mutation defines a new class of histone regulatory genes in Saccharomyces cerevisiae. *Genetics*, 135(1):25–34.

Spector, M. S., Raff, A., DeSilva, H., Lee, K., and Osley, M. A. (1997). Hir1p and Hir2p function as transcriptional corepressors to regulate histone gene transcription in the Saccharomyces cerevisiae cell cycle. *Mol Cell Biol*, 17(2):545–552.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell*, 9(12):3273–3297.

Spencer, C. A., Kruhlak, M. J., Jenkins, H. L., Sun, X., and Bazett-Jones, D. P. (2000). Mitotic transcription repression in vivo in the absence of nucleosomal chromatin condensation. *J Cell Biol*, 150(1):13–26.

Spiegelman, S., Burny, A., Das, M. R., Keydar, J., Schlom, J., Travnicek, M., and Watson, K. (1970). DNA-directed DNA polymerase activity in oncogenic RNA viruses. *Nature*, 227(5262):1029–31.

Sturm, S. and Okayama, H. (1996). Domains determining the functional distinction of the fission yeast cell cycle "start" molecules Res1 and Res2. *Mol Biol Cell*, 7(12):1967–1976.

Sugiyama, A., Tanaka, K., Okazaki, K., Nojima, H., and Okayama, H. (1994). A zinc finger protein controls the onset of premeiotic DNA synthesis of fission yeast in a Mei2-independent cascade. *EMBO J*, 13(8):1881–7.

Sveiczer, A., Novak, B., and Mitchison, J. M. (1996). The size control of fission yeast revisited. *J Cell Sci*, 109 (Pt 12):2947–2957.

Tahara, S., Tanaka, K., Yuasa, Y., and Okayama, H. (1998). Functional domains of rep2, a transcriptional activator subunit for Res2-Cdc10, controlling the cell cycle "start". *Mol Biol Cell*, 9(6):1577–1588.

Tanaka, K., Okazaki, K., Okazaki, N., Ueda, T., Sugiyama, A., Nojima, H., and Okayama, H. (1992). A new cdc gene required for S phase entry of Schizosaccharomyces pombe encodes a protein similar to the cdc 10+ and SWI4 gene products. *EMBO J*, 11(13):4923–4932.

Taylor, I. A., McIntosh, P. B., Pala, P., Treiber, M. K., Howell, S., Lane, A. N., and Smerdon, S. J. (2000). Characterization of the DNA-binding domains from the yeast cell-cycle transcription factors Mbp1 and Swi4. *Biochemistry*, 39(14):3943–3954.

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol Cell*, 23(6):853–64.

Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P., and Myers, R. M. (2004). An abundance of bidirectional promoters in the human genome. *Genome Res*, 14(1):62–6.

Uhler, J. P., Hertel, C., and Svejstrup, J. Q. (2007b). A role for noncoding transcription in activation of the yeast PHO5 gene. *Proc Natl Acad Sci U S A*, 104(19):8011–6.

Uhlmann, T., Boeing, S., Lehmbacher, M., and Meisterernst, M. (2007). The VP16 activation domain establishes an active mediator lacking CDK8 in vivo. *J Biol Chem*, 282(4):2163–73.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, 270(5235):484–7.

Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, Jr, D. E., Hieter, P., Vogelstein, B., and Kinzler, K. W. (1997). Characterization of the yeast transcriptome. *Cell*, 88(2):243–51.

Viladevall, L., St Amour, C. V., Rosebrock, A., Schneider, S., Zhang, C., Allen, J. J., Shokat, K. M., Schwer, B., Leatherwood, J. K., and Fisher, R. P. (2009). TFIIH and P-TEFb coordinate transcription with capping enzyme recruitment at specific genes in fission yeast. *Mol Cell*, 33(6):738–51.

Vilo, J., Brazma, A., Jonassen, I., Robinson, A., and Ukkonen, E. (2000). Mining for putative regulatory elements in the yeast genome using gene expression data. *Proc Int Conf Intell Syst Mol Biol*, 8:384–394.

Vitiello, S. P., Wolfe, D. M., and Pearce, D. A. (2007). Absence of Btn1p in the yeast model for juvenile Batten disease may cause arginine to become toxic to yeast cells. *Hum Mol Genet*, 16(9):1007–1016.

Ward, D. F. and Murray, N. E. (1979). Convergent transcription in bacteriophage lambda: interference with gene expression. *J Mol Biol*, 133(2):249–266.

Watanabe, T., Miyashita, K., Saito, T. T., Nabeshima, K., and Nojima, H. (2002). Abundant poly(A)-bearing RNAs that lack open reading frames in Schizosaccharomyces pombe. *DNA Res*, 9(6):209–15.

Watanabe, T., Miyashita, K., Saito, T. T., Yoneki, T., Kakihara, Y., Nabeshima, K., Kishi, Y. A., Shimoda, C., and Nojima, H. (2001). Comprehensive isolation of meiosis-specific genes identifies novel proteins and unusual non-coding transcripts in Schizosaccharomyces pombe. *Nucleic Acids Res*, 29(11):2327–37.

Weigel, D., Jürgens, G., Küttner, F., Seifert, E., and Jäckle, H. (1989). The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. *Cell*, 57(4):645–58.

Whitehall, S., Stacey, P., Dawson, K., and Jones, N. (1999). Cell cycle-regulated transcription in fission yeast: Cdc10-Res protein interactions during the cell cycle and domains required for regulated transcription. *Mol Biol Cell*, 10(11):3705–3715.

Whitehouse, I., Rando, O. J., Delrow, J., and Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. *Nature*, 450(7172):1031–5.

Whitfield, M. L., Sherlock, G., Saldanha, A. J., Murray, J. I., Ball, C. A., Alexander, K. E., Matese, J. C., Perou, C. M., Hurt, M. M., Brown, P. O., and Botstein, D. (2002b). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell*, 13(6):1977–2000.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, 75(5):855–62.

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J., and Bähler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature*, 453(7199):1239–43.

Williams, T. J. and Fried, M. (1986). The MES-1 murine enhancer element is closely associated with the heterogeneous 5' ends of two divergent transcription units. *Mol Cell Biol*, 6(12):4558–69.

Willingham, A. T., Dike, S., Cheng, J., Manak, J. R., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Duttagupta, R., Ganesh, M., Ghosh, S., Helt, G., Nix, D., Piccolboni, A., Sementchenko, V., Tammana, H., Kapranov, P., ENCODE Genes And Transcripts Group, and Gingeras, T. R. (2006). Transcriptional landscape of the human and fly genomes: nonlinear and multifunctional modular model of transcriptomes. *Cold Spring Harb Symp Quant Biol*, 71:101–10.

Winter, J., Jung, S., Keller, S., Gregory, R. I., and Diederichs, S. (2009). Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol*, 11(3):228–34.

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(6874):871–880.

Workman, C. T., Mak, H. C., McCuine, S., Tagne, J.-B., Agarwal, M., Ozier, O., Begley, T. J., Samson, L. D., and Ideker, T. (2006). A systems approach to mapping DNA damage response pathways. *Science*, 312(5776):1054–1059.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., Libri, D., and Jacquier, A. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell*, 121(5):725–37.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L. M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, 457(7232):1033–7.

Zenklusen, D., Vinciguerra, P., Wyss, J.-C., and Stutz, F. (2002). Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Mol Cell Biol*, 22(23):8241–53.

Zheng, N., Fraenkel, E., Pabo, C. O., and Pavletich, N. P. (1999). Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev*, 13(6):666–674.

Zhou, J., Chau, C. M., Deng, Z., Shiekhattar, R., Spindler, M.-P., Schepers, A., and Lieberman, P. M. (2005). Cell cycle regulation of chromatin at an origin of DNA replication. *EMBO J*, 24(7):1406–17.

Zhu, G., Spellman, P. T., Volpe, T., Brown, P. O., Botstein, D., Davis, T. N., and Futcher, B. (2000b). Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth. *Nature*, 406(6791):90–94.

Zhu, Y., Takeda, T., Whitehall, S., Peat, N., and Jones, N. (1997). Functional characterization of the fission yeast Start-specific transcription factor Res2. *EMBO J*, 16(5):1023–1034.

A Reagents and Media

Media: YEL+A

- 0.5% yeast extract, 3% glucose, 100μg/mL adenine, filter sterilized

Media: EMM2*

- 15mM potassium hydrogen pthalate, 10mM dibasic sodium phosphate, 93.5mM ammonium chloride, 0.5% glucose, 1x each EMM salts, EMM minerals, EMM vitamins

1x Spotted array hybridization buffer

- 25% v/v formamide, 5× SSC, 0.1% SDS, 100μg/ml of sonicated salmon sperm DNA

Wash Buffer 1 (spotted arrays)

- 2× SSC, 0.1% SDS

Wash Buffer 2 (spotted arrays)

- 0.1× SSC

2x hybridization buffer (Affymetrix)

- 200mM MES, 2M Na^+, 40mM EDTA, 0.02% Tween-20

Wash Buffer A (Affymetrix)

- 6x SSPE, 0.01% TWEEN-20

Wash Buffer B (Affymetrix)

- 100mM MES, 0.1M Na^+, 0.01% Tween-20

2x Stain Buffer (Affymetrix)

- 200mM MES, 2M Na^+, 0.10% Tween-20

1x Array Holding Buffer (Affymetrix)

- 100mM MES, 1M Na^+, 0.01% Tween-20

SAPE Stain Solution [Stain 1 & 3] (Affymetrix)

- 100mM MES, 1M Na^+, 0.05% Tween-20, 2mg/ml BSA, 10μg/ml SAPE

Antibody Stain Solution [Stain 2] (Affymetrix)

- 100mM MES, 1M Na^+, 0.05% Tween-20, 2mg/ml BSA, 0.1mg/ml goat IgG, 3μg/ml goat α-Streptavidin antibody (biotinylated)

Reagents/Supplies and their primary sources:

- Cy5 and Cy3 mono-reactive dyes, NHS ester (GE Life Sciences)
- 5-(3-aminoallyl)-dUTP (aa-dUTP) (Ambion)
- SuperScript II/III Reverse Transcriptase (Invitrogen)
- RiboPure Yeast RNA extraction kit (Ambion)
- LabChip/BioAnalyzer RNA Nano 6000 (Agilent Life Sciences)
- Superchip aminopropylsilane microarray slides (Erie)
- Microquill 2000 contact printing pins (Majer Precision)
- *S. pombe* Tiling 1.0FR arrays (Affymetrix)