Genome-wide Characterization of the Complex Transcriptome Architecture of *S. cerevisiae* with Tiling Arrays

Marina Granovskaia

2008

Dissertation

Submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Diploma-Biochemist Marina Granovskaia
Born in Moscow, Russia

Oral examination: 15 February 2008

Genome-wide Characterization of the Complex Transcriptome Architecture of *S. cerevisiae* with Tiling Arrays

Referees: Dr. Blanche Schwappach, ZMBH, University of Heidelberg

Dr. Peer Bork, EMBL, Heidelberg

Acknowledgements

I would like to thank Dr. Lars Steinmetz for giving me the opportunity to work in his lab towards my PhD thesis. I am grateful for Lar's continuous support throughout my thesis work, his original ideas, and our intensive and constructive discussions, for his encouragements and understanding in the critical moments, for his contagious enthusiasm about new discoveries and passion about science. I value that Lars would never let me alone in the moments of hesitations, but would always promote my independence to develop my own ideas and fulfill them. I want to thank Lars for a wonderful opportunity to work at the Stanford University Genome Technology Center as part of the scientific collaboration in our lab. That has been an unforgettable scientific and personal experience for me and I want to thank all the SGTC team who made it as such. Working with you, Lars has always been a complete intellectual involvement and enjoyment!

I would like to thank Dr. Blanche Schwappach and Dr. Peer Bork for their continuous support, fruitful discussions and constructive advice in my thesis advisory committees and for agreeing to evaluate my thesis.

I would like to thank Dr. Wolfgang Huber for his advice on the bioinformatics part of the projects and for giving me the opportunity to work with excellent members of his team in a productive collaboration.

I would also like to thank Dr. Jochen Wittbrodt for agreeing to be on my defense committee and for giving me very helpful advice on the preparation for the defense.

I want to thank Dr. Paul Bertone for agreeing to be on my thesis defense committee.

I want to thank Dr. Martina Muckenthaler for being on my thesis advisory committees and for the constructive discussions therein as well as for her personal and professional advice.

I am particularly grateful to Dr. Matt Ritchie for constantly providing me with bioinformatics support, for thoroughly finding and trying new approaches to analyze my data. I thank Matt for great help in generating several figures for my thesis and for always supporting me in testing new things in the analysis of the data. It has been a pleasure working together with you!

I would like to thank Dr. Lars Juhl Jensen for an extremely fruitful collaboration on the cell cycle project, for producing an excellent and fast analysis of my datasets and sharing his broad expertise on the regulation of the cell cycle.

Very special words of appreciation go to Dr. Vladimir Benes for his great help and critical ideas throughout constant optimizations of hybridization procedure that I was doing in my PhD; for his readiness to support me in facing any intellectual challenges that I encountered and sharing his expertise

of the subject, for pushing me to think out of the established frames of the techniques in order to find new solutions. I also thank Vladimir for his endless support and encouragements.
Many thanks to Tomy Ivacevic for his help with technical details and other tips and tricks; thanks for his advice on hybridizations and for always being ready to devote his time. Thanks to all other members of the GeneCore facility for creating a pleasant atmosphere to work with and for their readiness to help.
I would like to thank all of the members of the Steinmetz Lab for making my years here an enjoyable time. They are: Sandra Clauder-Muenster, Ye Ning, Rui Wang, Zhenyu Xu, Himanshu Sinha, Eugenio Mancera, Fabiana Perocchi, Julien Gagneur, Stephanie Blandin, Raeka Aiyar, Wu Wei, Stefan Mockenhaupt, and Stefan Wilkening. Special thanks go to Sandra Clauder-Muenster and Ye Ning for their technical support and fruitful discussions on the new techniques and to Rui Wang for her support, encouragement and sharing wise life perspectives.
I would like to particularly like to thank Dr.Yury Belyaev and Dr.Arne Seitz from the ALMF facility for their diligent help with electron microscopy imaging and processing.
I would like to thank Bernadett Papp, Janina Karres, Philipp Gebhardt and Yevheny Vainshtein for their friendship and support. I want to thank Yevheny for his continuous help with computational stuff. I would like to thank a lot to Helio Roque for his great help as well.

Огромное спасибо моим замечательным родителям и бабушке за поддержку во время всех лет учёбы, за бесконечную веру в меня. Спасибо за Вашу любовь и дружбу и за критику, которая помогала мне осознавать поворотные моменты моей жизни.
Спасибо Боре за терпение и поддержку в течение этих лет и за то, что ты всегда пытался внести в мою жизнь внешний баланс, вырывая из круга бесконечных задач по работе.
Спасибо моему сыну Марку за то, что он стал ещё одним веским поводом и вдохновением добиться карьерного успеха, за то, что он дал мне понять, что как бы ни было сложно, в жизни есть непреходящие ценности, которые делают её счастливой и насыщенной. Спасибо за то, что ещё ничего не понимая, ты терпел моё отсутствие, когда я была нужна тебе больше всего!

Thanks to everyone else who is not mentioned here due to space limitations, but who has made this work possible and supported me throughout these years!

Table of Contents

List of Illustrations

List of tables

Summary

The wealth of information accumulated about most eukaryotic genomes over the past decade has driven the systems approach in biology, which focuses on extracting the complete functional information encoded in a genome, including the genomic, regulatory and structural elements and integrating it in genomic networks. Recent genome-wide transcriptome analysis in humans, *Drosophila*, *Arabidopsis* and yeast challenged the old notion of the fundamental aspects of gene regulation, providing evidence that protein-encoding genes are not the only agents controlling cellular processes. Non-coding RNAs comprising untranslated regions of protein coding genes, antisense transcripts of annotated genes, micro RNAs and small interfering RNAs present another tier in gene regulation, enabling integration and networking of complex suites of gene activity. Sophisticated RNA signaling networks operate in higher eukaryotes, enabling gene to gene communication and regulation of chromatin structure, DNA methylation, transcription, translation, RNA silencing and stability, and coordinate multiple tasks of the whole cellular system. Fundamental mechanisms and structure of such control architecture remained largely obscure due to limitations of available approaches, such as noise in the data, strand–unspecific transcription analysis and difficulties in functional follow-up opportunities in higher eukaryotes.

To address the complexity of transcriptome architecture we undertook the genome-wide transcriptome study in a simpler genome of *S.cerevisiae* with the help of a new tiling array. This array is unique in interrogating every single nucleotide of the yeast genome 6 times and it is the first time a whole eukaryotic genome is synthesized on a single array. Over three million probe pairs corresponding to sense and anti-sense strands of the standard laboratory strain, S288c, are staggered from each other by 4 bases. The relatively small genome and well-established genetics of *S. cerevisiae* offer the opportunities to rapidly test new hypothesis and characterize novel findings.

We have shown that 85% of the genome is expressed in rich media. Apart from expected transcripts, we found operon-like transcripts, transcripts from neighboring genes not separated by intergenic regions, and genes with complex transcriptional architecture where different parts of the same gene are expressed at different levels. We mapped the positions of 3' and 5' UTRs of coding genes and identified hundreds of RNA transcripts distinct from annotated genes. These non-annotated transcripts, on average, have lower sequence conservation and lower rates of deletion phenotype than protein coding genes. Many other transcripts overlap known genes in antisense orientation, and for these pairs global correlations were discovered: UTR lengths correlated with gene function, localization, and requirements for

regulation; antisense transcripts overlapped 3' UTRs more than 5' UTRs; UTRs with overlapping antisense tended to be longer; and the presence of antisense associated with gene function.

Overall our study revealed complexity of yeast transcriptional architecture calling for additional annotation of the genome and putting forward an important role for RNA-mediated regulation.

An attractive model for the study of the genome-wide RNA-mediated regulation of gene activity in yeast is mitotic cell cycle, which has been extensively studied over past decade and is therefore a well characterized system. Mitosis is associated with important physiological changes in the cell and diverse biological events depend on this periodicity. To ensure the proper functioning of the mechanisms that maintain order during cell division about 800 genes of diverse GO categories are coordinately regulated in a periodic manner coincident with the cell cycle. This includes genes involved in DNA replication, budding, glycosylation, nuclear division, control of mRNA transcription, responsiveness to external stimuli and subcellular localization of proteins.

Several genome-wide studies have been done to catalogue cell cycle-regulated genes with the help of early expression arrays. Given the high resolution of our technique, profiling genome-wide periodic expression with the tiling arrays allowed taking a step forward to prove the existence of RNA-mediated regulation of transcription.

Using two methods of synchronization, I have monitored cell-cycle dependent transcription for more than 3 complete cell cycles. I have identified about ~600 periodic ORFs. In consent with previous studies on transcriptional regulation during specific mitotic phases, I have shown prevalence of periodic expression of annotated genes in three distinct periods of cell cycle progression: late G1/S transition, G2/M transition and exit of M phase of mitosis.

Moreover, I have shown antisense transcription throughout the cell cycle phases. Out of ~260 antisense transcripts that we discovered, 37 display periodic patterns; half of them are expressed coincidentally with peak expression intensity of cell cycle-regulated ORFs, whereas the other half peaks at the periods of relaxation of the transcriptional machinery, which drives phase transition. Cycling antisense has been registered opposite several important cell cycle regulators.

Additionally, periodic novel isolated transcripts were detected in the dataset, which may represent non-annotated ncRNAs involved in regulation of mitosis or regulated by cell cycle controlling genes.

Zusammenfassung

Die Fülle an Informationen die sich über die meisten eukaryotischen Genome während der letzten Dekaden angesammelt haben, führte zu systematischen Analyse-Methoden in der Biologie. Diese haben zum Ziel die vollständigen funktionellen Informationen die in einem Genom kodiert sind, einschließlich der genomischen, regulatorischen und strukturellen Elemente, zu extrahieren und in genomische Netzwerke zu integrieren. Neueste genomweite Transkriptom-Analysen im menschlichen Genom, in *Drosophila*, *Arabidopsis* und Hefe, stellen die alte Auffassung von den fundamentalen Aspekten der Genregulation in Frage und liefern Beweise dafür, dass Protein-Codierende Gene nicht die einzigen Vermittler sind zelluläre Prozesse zu kontrollieren. Nicht-Kodierende RNAs welche untranslatierte Regionen Protein-Codierender Gene, Antisense-Strang Transkripte annotierter Gene, micro RNAs und 'small interfering' RNAs umfassen, stellen eine weitere Stufe der Genregulierung dar und ermöglichen die Integration und Vernetzung komplexer Abfolgen der Genaktivität.

In höheren Eukaryoten operieren ausgeklügelte RNA Signal-Netzwerke und ermöglichen ihnen eine Gen zu Gen Kommunikation und Regulation der Chromatin-Struktur, der DNA-Methylation, der Transkription, der Translation, 'RNA-silencing' und Stabilisierung und koordinieren zudem verschiedene Aufgaben des gesamten zellulären Systems. Fundamentale Mechanismen und Strukturen solcher Kontroll-Einrichtungen blieben weitgehend ungeklärt. Grund dafür waren Einschränkungen in vorhandenen Methoden wie das Störungs-Rauschen in Datensätzen, Strang-unspezifische Transkriptions-Analysen und Schwierigkeiten in funktionellen Folge-Untersuchungen in höheren Organismen. Um die komplexe Transkriptom-Architektur zu untersuchen führten wir mit Hilfe des neuen 'tiling array' eine genomweite Transkriptom-Studie des einfacheren Genoms der *S.cerevisiae* durch. Dieser Array ist der einzige der jedes einzelne Nukleotid des Hefe-Genoms sechs Mal abdeckt. Gleichzeitig ist es des erste Mal, dass ein gesamtes Genom auf einem einzigen Array synthetisiert ist. Über drei Millionen Sonden-Paare, entsprechend der Sense- und Antisense-Stränge des Standard Labor-Stammes, sind mit jeweils einem Vier-Basen-Versatz auf diesem Array vorhanden. Das relativ kleine Genom und die gut etablierte Genetik der *S. cerevisiae* bieten die Möglichkeit, schnell neue Hypothesen zu testen und neue Erkenntnisse zu charakterisieren.

Wir demonstrierten das 85% des Genoms in reichhaltigem Medium exprimiert ist. Neben den erwarteten Transkripten fanden wir Operon-ähnliche Transkripte, Transkripte von benachbarten Genen die nicht durch intergenetische Regionen getrennt sind und Gene mit komplexem transkriptionellem Aufbau bei denen verschiedene Teile eines Gens unterschiedlich stark exprimiert sind. Wir kartierten die Positionen

der 3’ und 5’ untranslatierten Regionen (UTRs) der codierenden Gene und identifizierten hunderte von RNA Transkripten die nicht zu annotierten Genen zählten. Diese nicht-annotierten Transkripte zeigen im Durchschnitt eine niedrigere Sequenzkonservierung und geringere Raten an Deletions-Phänotypen als Protein-Codierende Gene. Viele andere Transkripte überlappen Gene in antisense Orientierung. Für diese Paare entdeckten wir globale Korrelationen: Die UTR-Länge korreliert mit der Genfunktion, mit der Lokalisierung und mit den Anforderungen der Regulation; antisense Transkripte überlappen 3’ UTRs mehr als 5’ UTRs; UTRs mit überlappendem Antisense sind tendenziell länger; das Vorhandensein der antisense Transkripte ist mit der Genfunktion assoziiert. Insgesamt betrachtet offenbart unsere Studie einen komplexen Aufbau der Hefe- Transkription die eine zusätzliche Annotierung des Genoms einfordert und eine wichtige Rolle für die RNA-vermittelte Genregulation vorantreibt. Ein attraktives Model zur Studie der genomweiten RNA-vermittelten Regulation der Genaktivität in Hefe ist der mitotische Zellzyklus, der während der letzten Dekade intensiv erforscht wurde und daher ein gut charakterisiertes System darstellt. Die Mitose ist mit wichtigen physiologischen Veränderungen in der Zelle assoziiert und verschiedene biologische Ereignisse sind von dieser Periodizität abhängig. Circa 800 Gene der verschiedensten GO Kategorien sind koordiniert in einer periodische Weise reguliert, übereinstimmend mit dem Zellzyklus, um eine exakte Funktion des Mechanismus der die Ordnung während der Zellteilung gewährt, zu ermöglichen. Diese beinhalten Gene in der DNA Replikation, Knospung, Glykosilierung, Kernteilung, Kontrolle der mRNA Transkription, Empfindlichkeit für externe Stimuli und subzelluläre Lokalisierung von Proteinen.

Einige genomweite Studien wurden unternommen um Zellzyklus regulierte Gene mit Hilfe von frühen Expressions-Arrays zu katalogisieren. Bestimmt durch die hohe Auflösung unserer Technik, hat uns die genomweite Profilierung periodischer Expression mit den ‘tiling arrays’ es uns ermöglicht einen Schritt weiter zu gehen und die Existenz der RNA-vermittelten Regulation der Transkription zu beweisen.

Unter Anwendung von zwei verschiedenen Methoden der Synchronisierung habe ich die Zellzyklus abhängige Transkription für mehr als drei vollständige Zellzyklen aufgezeichnet. Ich habe ca. 600 periodische ORFs identifiziert. Übereinstimmend mit früheren Studien der transkriptionellen Regulierung während spezifischer mitotischer Phasen konnte ich eine Anhäufung der periodischen Expression von annotierten Genen in drei verschiedenen Perioden des Zellzyklus aufzeigen: Später G1/S Übergang, G2/M Übergang und Übergang von M Phase zur Mitose. Darüber hinaus zeigte ich antisense Transkription während aller Zellzyklus Phasen. Von ca. 260 antisense Transkripten die entdeckt wurden zeigen 37 ein periodisch wiederkehrendes Muster; die Hälfte von ihnen deckt sich mit der höchsten Expressions-Intensität von Zellzyklus regulierten ORFs, während die andere Hälfte die höchste

Expression zeigt während die Transkriptions-Maschinerie in einer Ruhephase ist, welches den Phasenübergang antreibt. Periodische antisense Transkription konnte gegenüber verschiedenen wichtiger Zellzyklus-Regulatoren festgestellt werden.

Zusätzlich konnten neue periodische Transkripte in dem Datensatz identifiziert werden. Diese könnten nicht annotierte nicht-codierende RNAs repräsentieren die in der Regulation der Mitose involviert sind, oder die durch Zellzyklus-kontrollierende Gene kontrolliert werden.

1. INTRODUCTION

I. Challenging Dogmas

The 20th century concept for the relationship between genetic information and its biological realization has been primarily rooted in the "one gene - one protein" hypothesis. The notion of DNA as a relatively stable, heritable source of template information for proteins, transduced through temporary and discrete RNA readout has been a matter of faith and has powerfully influenced researchers' ideas on the structure of genetic systems and their complex regulatory wiring (Mattick and Gagen 2001). Accordingly, myriads of proteins comprise structural and functional blocks of the cells and are the only agents, which execute intricate multi-switch control of cellular dynamics. This belief was mainly shaped by early experiments in prokaryotes, which defined our understanding of genes and their expression and regulation (Mattick 2004). Complete sequencing of many bacterial and archea strains has confirmed that their genomes predominantly comprise protein coding sequences, flanked by 5' and 3' *cis*-regulatory elements that operate to control the expression of these sequences at the transcriptional or translational level. The only exceptions constitute infrastructural tRNAs and rRNAs that are required for protein synthesis, and a small number of genes coding for non-translated RNAs with regulatory functions, which occupy no more than 1% of the genome sequence (Argaman *et al.* 2001; Wassarman *et al.* 1999).

Such architecture has been extended to eukaryotes, leading to the assumption that eukaryotic phenotypic complexity arises from multitasking of a core proteome of a limited size and is exerted via intricate combinatorics of its regulatory factors that intersect with more complex promoters (Levine and Tjian 2003). This view dwells on a biochemical rather than an informatics perspective, which gives insufficient consideration to the problem of how complex programs of gene activity in higher organisms could be integrated and regulated (Mattick and Gagen 2001). The validity of the "almighty proteome" view is in part bound to the question of how many regulatory inputs can sensibly be integrated to produce different outcomes.

The technological advancements of the last decade have brought about unprecedented sophistication in scrutinizing biological objects and promoted our knowledge to a much deeper and broader level, gradually shifting the veil from hidden tiers of an organism's complexity. The wealth of information accumulated about most eukaryotic genomes over the past decade due to available technologies, has shifted the focus of modern molecular biology to the systematic level of research, which deals with context dependent - varying according to physiological, developmental, pathological, etc. state of a cell or an organism - transcriptome, proteome and metabolome (Oliver 2002). Recent genome-wide transcriptome analysis in humans (Kampa *et al.* 2004; Penn *et al.* 2000; Schadt *et al.* 2004), *Drosophila*

(Hild *et al.* 2003; Stolc *et al.* 2004), *Arabidopsis* (Yamada *et al.* 2003) and yeast (David *et al.* 2006) were important scientific milestones that challenged the old notion of fundamental aspects of gene regulation, providing evidence that protein-encoding genes are not the only agents controlling cellular processes (Johnson *et al.* 2005). Non-protein-coding RNAs (ncRNAs) comprising untranslated regions of protein coding genes, antisense transcripts of annotated genes, micro RNAs present another tier in gene regulation, enabling integration and networking of complex suites of gene activity (Mattick 2003). Strikingly, non-coding RNAs constitute at least 70% of the genomic transcriptional output of higher eukaryotes, mounting to 95% in humans (Kenzelmann *et al.* 2006; Mattick 2001; 2003). Moreover the proportion of protein-coding genes declines as a function of an organism complexity. RNA signaling with sophisticated infrastructure operates in higher eukaryotes, enabling gene to gene communication and regulation of chromatin structure, DNA methylation, transcription, translation, RNA silencing and stability, and coordinates multiple tasks of the whole cellular system (Mattick and Gagen 2001; Wassenegger 2000).

RNA molecules are very versatile, mainly due to their chemical properties, which allow them to form complex tertiary structures that are capable of performing several roles, which were thought to be under the exclusive domain of proteins. RNAs are integral components of RNA-protein ribocomplexes, in which they serve a variety of functions. They can associate with specific DNA and/or RNA sequences, controlling several aspects of gene regulation and enhancing the plethora of molecular connections in eukaryotic cells. These transcripts are directly involved in control of many normal cellular processes and their deregulation (deletion, over-expression or down-regulation) may underlie or be a marker for the severity of complex diseases.

Fundamental mechanisms and structure of such control architecture remained largely obscure due to limitations of available approaches, such as noise in the data, strand–unspecific transcription analysis and difficulties in functional follow-up opportunities in higher eukaryotes. Having drastically transformed our outlook on specification of functional components of a given cellular system and control of their expression and regulation, i.e. on the flow of genetic information, efferent ncRNAs remain under a detailed scrutiny of many research groups.

II. Non-coding RNAs

Based on their functional relevance, ncRNAs can be roughly subdivided into two classes: (1) housekeeping ncRNAs and (2) regulatory ncRNAs. Housekeeping ncRNAs, like transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), small nuclear (snRNAs), snoRNAs, RNase P RNAs, telomerase RNA, etc., are mostly constitutively expressed and are required for the normal function and viability of the cell (Eddy 2001).

Many of these infrastructural RNAs are integral components of RNA-protein complexes, performing essential functions therein. To name a few, the U small nuclear RNA constitutes a catalytic center of the eukaryotic spliceosome, utilizing base-pairing interactions to identify splice sites. 4.5S and SRP RNA of prokaryotic and eukaryotic signal recognition particle, respectively, is a major structural component of the ribonuclearprotein complex, in which it resides. SRP forms an RNA-protein complex involved in exporting secreted proteins from the cell. The telomerase RNA serves a template for the synthesis of telomeres on the ends of chromosomes. Bacterial SsrA/tmRNA mimics tRNA and an mRNA, functioning to release ribosomes trapped in damaged mRNA molecules (Withey and Friedman 2003). Ribonuclease P is a uniquely conserved enzyme that cleaves a leader sequence from tRNA precursors.

Increasing number of regulatory ncRNAs comprise untranslated regions of a coding sequence, like introns (Lau *et al.* 2001; Llave *et al.* 2002a), or antisense transcripts, exerting their influence on their sense counterpart either in *cis*- or *trans*-, and finally small non-coding RNA molecules, such as small interfering RNAs (siRNAs) and microRNAs. In contrast to housekeeping RNAs, riboregulators are expressed at certain stages of development, during cell differentiation, or as a response to external stimuli, which can affect the expression of other genes at the level of transcription or translation (Prasanth and Spector 2007).

The current understanding of the roles of non-coding RNAs in eukaryotic cells and their involvement in gene organization, regulation, and etiology of disease is highlighted below.

1. Infrastructural RNAs

1.1 Transfer RNA (tRNA)

tRNA was first predicted by Francis Crick in his “adaptor” hypothesis to serve as a mediator between the triplet genetic code and the encoded amino acid (Eddy 2001). Crick argued that RNA would be evolutionary preferred over a protein as it is uniquely suited for a role of a small RNA recognition

molecule (Crick 1958). A small RNA chain (73-93 nucleotides) transfers a specific amino acid to a growing polypeptide chain at the ribosomal site of protein synthesis during translation. Aminoacyl tRNA synthetase catalyzes covalent attachement of the 3' terminal tRNA site to its specific amino acid. Its three base anticodon region base-pairs to the corresponding three base codon region on mRNA. Each type of tRNA molecule is suited to attach to a unique amino acid, but because the genetic code contains multiple codons that specify the same amino acid, tRNA molecules bearing different anticodons may also carry the same amino acid.

tRNA has the cloverleaf secondary structure, and a similar L-shaped tertiary structure that allows it to fit into the P and A sites of the ribosome (Clark 2006).

Organisms vary in the number of tRNA genes in their genome. For example, out of 19.000 genes in the nematode worm *C. elegans* 659 code for tRNA (Hartwell *et al.* 2004a); the budding yeast *Saccharomyces cerevisiae* has 275 tRNA genes in its genome. In the human genome, which according to current estimates has more than 25,000 genes in total, there are about 2000 non-coding RNA genes, which include tRNA genes. There are 22 mitochondrial tRNA genes (Hartwell *et al.* 2004b); 497 nuclear genes encoding cytoplasmic tRNA molecules and there are 324 tRNA-derived putative pseudogenes (Lander *et al.* 2001).

In eukaryotic cells tRNAs are transcribed by RNA polymerase III.

1.2 Ribosomal RNA (rRNA)

rRNA is synthesized in the nucleolus by RNA polymerase I and forms the ribosome. The function of the rRNA is to provide a mechanism for decoding mRNA into amino acids and to interact with the tRNAs during translation due to peptidyl transferase activity. Like tRNA, the ribosome also has 3 binding sites called A, P, and E. The A site in the ribosome binds to an aminoacyl-tRNA. In peptidyl transferase reaction the NH2 group of the aminoacyl-tRNA, which contains the new amino acid, attacks the carboxyl group of peptidyl-tRNA (contained within the P site) which carries the last amino acid of the growing chain. The tRNA that was holding on the last amino acid is moved to the E site, and what used to be the aminoacyl-tRNA becomes the peptidyl-tRNA (Yusupov *et al.* 2001).

Both 70S prokaryotic and 80S eukaryotic ribosome can be broken down into two subunits: 50S and 30S and 60S and 40S, respectively.

Ribosomal RNA Gene Structure

Bacterial 16S, 23S, and 5S rRNA genes are typically organized as a co-transcribed operon. There may be one or more copies of the operon dispersed in the genome (for example, *E.coli* has seven). The 3' end of the 16S rRNA binds to the Shine-Dalgarno sequence on the 5' end of mRNA, where translation starts.

In contrast, eukaryotes generally have many copies of the rRNA genes organized in tandem repeats; in humans approximately 300–400 rDNA repeats are present in five clusters (on chr. 13, 14, 15, 21 and 22). Mammalian cells have 2 mitochondrial (12S and 16S) rRNA molecules and 4 types of cytoplasmic rRNA. 28S, 5.8S, and 18S rRNAs are encoded by a single transcription unit (45S) separated by 2 internally transcribed spacer (ITS).

In all organisms, the mature rRNAs are generated by posttranscriptional processing reactions. In Bacteria and Archaea, the endonuclease RNAase III cleaves stem structures formed by complementary sequences that flank each of the mature rRNA sequences. The separated pre-rRNAs are 3’ processed by the 3’ to 5’ exoribonuclease RNase T and 5’ processed by the endonuclease RNase E. Processing occurs co-transcriptionally, but the requirement for the stem structures means that each mature RNA must be fully synthesized before its processing can commence.

In eukaryotes many of pre-rRNA processing enzymes remain to be identified. Processing is posttranscriptional, with the exception of the initial cleavage by RNAase III (Rnt1p in yeast) in the 3’-ETS which, at least in yeast, is cotranscriptional. Subsequent processing shows a strong 5’ to 3’ bias in the order of cleavage (Lafonataine and Tollervey 2001).

Importance of rRNA

rRNA is the target of several clinically relevant antibiotics: chloramphenicol, erythromycin, kasugamycin, micrococcin, ricin, sarcin, streptomycin, etc., which specifically bind to different catalytic sites of prokaryotic ribosomes and inhibit translation.

rRNA is the most conserved gene in all cells. For this reason, rRNA genes are sequenced to identify an organism's taxonomic group, calculate related groups, and estimate rates of species divergence.

Nucleolar dominance has been shown for rRNA genes. In some organisms, particularly plants, when two nuclei are combined into a single cell during hybridization the developing organism can 'choose' one set of rRNA genes for transcription. The rRNA genes of the other parent are suppressed and not generally transcribed, though reactivation of the suppressed or 'inferior' rRNA genes may occasionally occur. This selective preference of transcription of rRNA genes is termed nucleolar dominance.

1.3 Small nucleolar RNAs (snoRNAs)

snoRNAs are a class of small RNA molecules, which guide and, in association with specific proteins in snoRNP complex, catalyze the most common covalent modifications of rRNA - 2′-*O*-ribose methylation and pseudouridylation, which facilitate the folding and stability of rRNA (King *et al.* 2003; Ofengand 2002; Reichow *et al.* 2007). These modifications occur in the nucleolus where snoRNAs act on the pre-rRNA substrate (Boisvert *et al.* 2007; Kiss 2002) as well as on other important genes, like tRNAs and small nuclear RNAs (snRNAs). Modification sites cluster within functionally important regions of the ribosome, conserved throughout distant eukaryotes and are often located away from protein-binding sites (Decatur and Fournier 2002).

snoRNA genes are frequently encoded in the introns of ribosomal proteins, in particular in higher eukaryotes, and are synthesized by RNA polymerase II, but can also be transcribed as independent (sometimes polycistronic) transcriptional units. Processing of snoRNAs from introns of mRNAs serves as one of the first examples of the functional importance of intronic portions of pre-processed mRNAs thus blurring the boundaries of gene organization (Gingeras 2007). Each snoRNA molecule acts as a guide for only one (or two) individual modifications in a target RNA. snoRNA contains a 10-20 bp antisense element complementary to the sequence surrounding the base targeted for modification in the pre-RNA molecule, which enables the snoRNP to recognise and bind to the target RNA.

Two different families of snoRNPs - **antisense C/D box** (Fig. 1-1A) and **H/ACA box** snoRNAs (Fig. 1-1B), distinguished by the presence of conserved sequence motifs in the snoRNA, direct two different types of rRNA modifications. The class C/D snoRNPs are responsible for site-specific 20-O-methylation of ribose (Kiss-Laszlo and Hohn 1996; Nicoloso *et al.* 1996; Tycowski *et al.* 1996) while the H/ACA snoRNPs catalyze isomerization of uridine to C (Ganot *et al.* 1997; Ni *et al.* 1997).

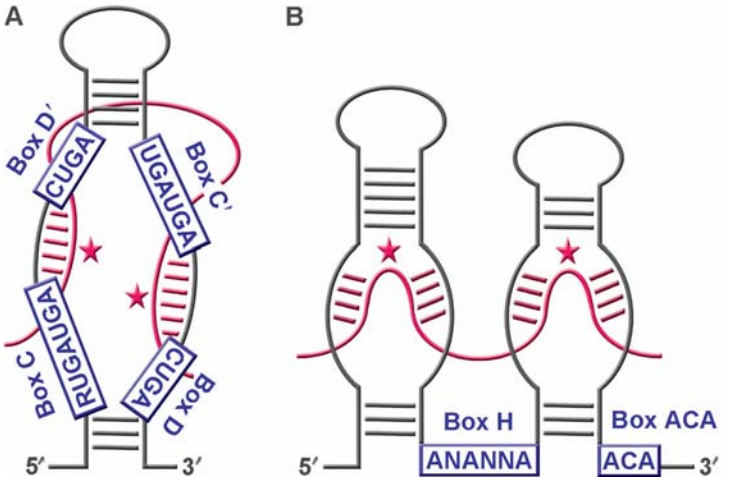


Figure 1-1. The guide snoRNAs and their target RNAs. The class **(A)** C/D and **(B)** H/ACA snoRNAs (grey) contain conserved and class-specific sequence motifs 'boxes' (blue) and unique guide regions that define their respective target RNA site(s) (magenta). The sites targeted for nucleotide modification are marked with a star (adapted from (Reichow *et al.* 2007)).

Composite H/ACA and C/D box snoRNA

An unusual guide snoRNA U85 was identified that functions in both 2'-O-ribose methylation and pseudouridylation of small nuclear RNA (snRNA) U5 (Jády and Kiss 2001). This composite snoRNA contains both C/D and H/ACA box domains and associates with the proteins specific to each class of snoRNA (fibrillarin and Gar1p, respectively). Composite snoRNAs accumulate in a subnuclear organelle called the Cajal body and are often referred to as Cajal body specific RNAs (scaRNAs) (Darzacq *et al.* 2002). This is in contrast to the majority of C/D box or H/ACA box snoRNAs which localise to the nucleolus. scaRNAs are proposed to be involved in the modification of RNA polymerase II transcribed spliceosomal RNAs U1, U2, U4, U5 and U12. Not all snoRNAs that have been localised to Cajal bodies are composite C/D and H/ACA box snoRNAs.

Another interesting example of a snoRNA is human **telomerase RNA** (TR). It is a scaRNA containing a chimeric 50-reverse transcriptase domain involved in enzymatic activity and a 30-H/ACA domain that recruits each of the core H/ACA snoRNP proteins (Mitchell *et al.* 1999). Although this domain is dispensable for *in vitro* activity of telomerase, it is required *in vivo* for the processing, stability and nuclear localization of TR (Lukowiak *et al.* 2001; Mitchell *et al.* 1999; Mitchell and Collins 2000).

Evolutional conservation of snoRNAs

The C/D and H/ACA snoRNPs are universally present in eukaryotes and have also been discovered in archaea, where they are referred to as small RNPs (sRNPs) (Dennis and Omer 2005; Omer *et al.* 2003). Archaeal sRNPs have both rRNA and tRNA modification activity (Clouet d'Orval *et al.* 2001). The archaeal RNA components of sRNPs (sRNAs) are typically smaller than their eukaryotic counterparts, possibly representing minimal structural units for both classes. Thus, the functional requirements of the core snoRNP proteins and the molecular mechanisms of core formation and 20-O-methylation are ancient and have been conserved throughout evolution.

snoRNA targets

The targets for snoRNAs are predicted on the basis of sequence complimentarity between putative target RNAs and the antisense elements or recognition loops in the snoRNA sequence. However, there are an increasing number of 'orphan' guides without any known RNA targets, suggesting that there more

proteins or transcripts might be involved in rRNA processing and/or that some snoRNAs might perform functions not concerning rRNA (Gingeras 2007).

Computational studies of the *Saccharomyces cerevisiae* genome identified many novel methylation - guide snoRNAs, involved in rRNA modification (Lowe and Eddy 1999; Schattner *et al.* 2004), indicating that although this is a well-established functional class of noncoding transcripts, its membership is still growing. Recent studies show that a number of snoRNA transcripts do not possess sequences that are fully complementary to rRNA targets (Jady and Kiss 2000; Li *et al.* 2005). This suggests that a larger network of cellular proteins and/or other transcripts outside of the rRNA complex may be required to assist snoRNAs in carrying out their functions and opens the possibility that snoRNAs may have functions beyond modification of rRNAs and spliceosomal RNAs (Heix *et al.* 1998; Hinsby *et al.* 2006). For instance, regulation of alternative splicing of a transcript encoded in *trans,* has recently been demonstrated for HBII-52 snoRNA (Kishore and Stamm 2006).

The majority of snoRNAs are expressed constitutively. However a subset has a tissue-specific pattern. For example, several snoRNAs affected in the Prader-Willi/Angelman syndrome – a genetic disorder, where seven genes or a subset thereof in a region on chromosome 15 are missing or unexpressed due to imprinting error from parternal (in PWS) or maternal chromosome (Angelman) – are expressed only from paternal chromosome and are highly abundant in human brain tissue (Cavaille *et al.* 2000; Meguro *et al.* 2001).

1.4 Small nuclear RNA (snRNA).

snRNA is a class of small RNA molecules that are found within the spliceosome in the nucleus of eukaryotic cells. They are transcribed by RNA polymerase II or RNA polymerase III and are involved in a variety of important processes such as RNA splicing, regulation of transcription factors (7SK RNA and human U1 snRNA) or RNA polymerase II (B2 RNA), and maintaining the telomeres. They are always associated with specific proteins, and the complexes are referred to as small nuclear ribonucleoproteins (snRNP) or sometimes as snurps. These elements are rich in uridine content. Although the catalytic core of spliceosome has not been precisely identified recent studies provide direct evidence for the catalytic potential of spliceosomal snRNAs (Valadkhan 2007a; b).

2. Regulatory Non-coding RNAs.

Regulatory RNAs impact all steps in gene expression, acting either via base-pairing with other nucleic acids or through binding to and modifying the activity of a protein or protein complex (Storz *et al.* 2005). In the latter case they act as molecular decoys, which structure and/or sequence resembles that of the protein's natural cellular target, thus enabling competition for the protein binding to its target (Fig.1-2) (Gottesman 2002).

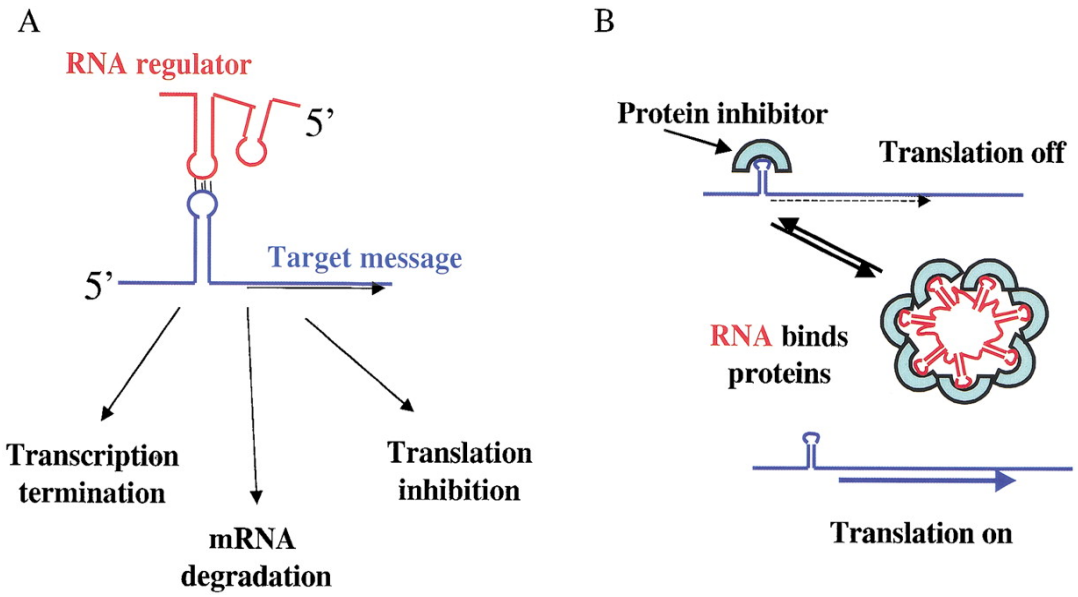


Figure 1-2. Categories of regulatory RNA action. (*A*) Regulatory RNA as antisense. Many small RNAs act as antisense regulators, pairing with a target, usually a messenger RNA, to change its behavior. In *E. coli*, many, if not all, such regulatory RNAS use the Hfq protein to stimulate pairing (discussed in text). The regulatory RNAs are shown in red. **(*B*)** Regulatory RNA as molecular mimic. One example of another class of small RNAs is shown at *right*. This is modeled on the CsrA/CsrB system, in which the protein CsrA is a translational inhibitor; inhibition is relieved by binding of multiple molecules of CsrA by CsrB (adapted from (Gottesman 2002)).

Thousands of naturally occurring antisense transcripts (NATs) of mammalian genes were identified during recent years in several independent genome-wide transcriptome studies, many of the NATs resulting from alternatively polyadenylated transcripts or heterogeneous transcription start sites (Dahary *et al.* 2005). *Cis*-NATs are expressed from overlapping loci on the opposite DNA strand, complementary to all or part of its sequence, whereas *trans*-antisense RNAs are transcribed from separate, non-overlapping loci that share complementary sequences with a particular region of the sense transcript that they regulate. Since *trans*-NAT pairs display imperfect complementarity they can therefore target many more sense transcripts to form complex regulatory networks (Li *et al.* 2006). Within a sense/antisense pairs (SAPs), *cis*-antisense form extended regions of perfectly matched double-stranded RNAs

(dsRNAs), while trans-antisense pairs, including miRNA usually form relatively short regions of base-pairing that are frequently interrupted by mismatches (Munroe and Zhu 2006). Although, sometimes the definition of sense and antisense is arbitrary, usually the antisense transcript is presumptive regulatory RNA, while the sense RNA is more abundant, more widely expressed or has a better characterized or more direct function (Chen *et al.* 2004a; Kumar and Carmichael 1998; Vanhee-Brossollet and Vaquero 1998; Yelin *et al.* 2003).

The first antisense RNAs to be rigorously characterized were the ones that regulate replication of bacterial plasmids. Soon after, antisense transcripts that control expression of endogenous bacterial mRNA were identified (Eguchi and Tomizawa 1991; Itoh and Tomizawa 1980; Storz *et al.* 2005; Wagner and Simons 1994). Subsequently, a detailed understanding of antisense-mediated gene regulation in bacteria emerged (Wagner *et al.* 2002), showing that *cis*-encoded antisense RNAs in plasmids modulate expression of genes involved in replication and stable plasmid inheritance; some are associated with transposons and bacteriophages.

The regulatory action of endogenous NAT in eukaryotes remained unclear, with only a few notable exceptions (Kumar and Carmichael 1998; Lavorgna *et al.* 2004; Munroe 2004; Vanhee-Brossollet and Vaquero 1998; Werner and Berdal 2005). The importance of NATs in eukaryotes is now apparent (Knee and Murphy 1997; Williams and Fried 1986) and has been emphasized recently with the appearance of advanced technologies to unravel them. Several genome-wide studies illustrated the extent of potential SAP occurrence in different experimental organisms and in human cell lines.

In mice almost 29% (12,519) of all mapped transcriptional units (43,553) were found to overlap with a cDNA that mapped to the opposite strand, greatly exceeding any previous prediction (Katayama *et al.* 2005). In humans, Chen and colleagues reported that nearly 22% (5,880) of 26,741 transcriptional clusters form sense–antisense pairs (Chen *et al.* 2004a). These data overlap with the other published data sets by only 38%, indicating that the total number of human sense–antisense pairs might be even greater (Yelin *et al.* 2003). A genome-wide tiling array (Bertone *et al.* 2004) revealed 10,595 novel human transcripts, 1,187 of which lie antisense to annotated exons. Annotation of the *Drosophila* genome identified 1,027 sense–antisense pairs, representing 15% of the 13,379 genes (Misra *et al.* 2002). In plants, annotation of full-length rice cDNAs revealed 687 overlapping cluster pairs, amounting to around 7% of all cDNA clusters (Osato *et al.* 2003), and similar percentages were obtained in *Arabidopsis thaliana* (Wang *et al.* 2005a). Genome-wide antisense transcription in yeast was recently reported (Havilio *et al*, 2005) and further supported by identification of antisense transcripts for up to 1,555 of the genes of *Saccharomyces cerevisiae* genome (David *et al.* 2006).

Thus far the notion has been that eukaryotic ncRNAs are mainly involved in posttranscriptional regulation of diverse cellular processes, affecting RNA stability, nuclear processing, export or translation. Classically known examples of posttranscriptional regulation are mRNA degradation mechanism in plants by siRNAs, and block of translation imposed by microRNAs in animals. Some ncRNAs were shown to serve as guide molecules to DNA methyltransferases in histone methylation conferring active chromatin into silent state (Mathieu and Bender 2004; Pickford and Cogoni 2003). In the view of recent discoveries of antisense RNAs we can consider that those RNAs may elicit regulatory responses on transcriptional level as well. Understanding the integration of transcriptional and post-transcriptional regulatory mechanisms represents a major challenge for understanding antisense regulation in eukaryotes.

cis-NATs can be categorized according to their relative orientation and degree of overlap; head-to-head (5′ to 5′), tail-to-tail (3′ to 3′) or fully overlapping (Fig. 1-3). All genome-wide studies, except one (Katayama *et al.* 2005), have reported the tail-to-tail orientation to be the most prevalent. Although the underlying mechanism of antisense regulation is largely unknown, it has been suggested that at least for *cis*-NATs a relationship exists between the mechanism of antisense function and the transcriptional co-regulation of sense and antisense transcripts: some mechanisms of antisense action require co-expression with its target, some require a time delay between the onset of antisense and sense transcription, and other mechanisms manifest themselves in anti-correlated expression patterns of the regulator and its target (Lapidot and Pilpel 2006). This concept could be used to infer the regulatory mechanism of action, given the expression profiles of antisense and their sense targets.

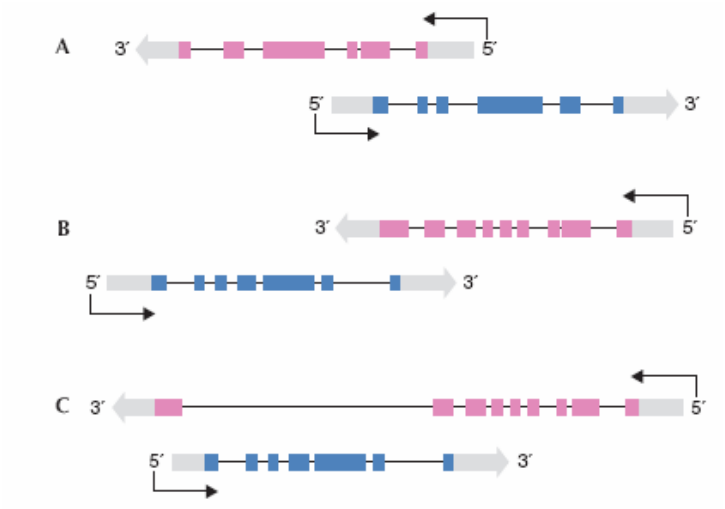


Figure 1-3. Relative orientation of cis-natural antisense transcript pairs. (**A**) Head-to-head (5′ to 5′) overlap involving 5′-untranslated regions and coding exons. (**B**) Tail-to-tail (3′ to 3′) overlap. (**C**) Fully overlapping (one gene included entirely within the region of the other).Coloured boxes represent exons, grey boxes represent untranslated regions (adapted from (Lapidot and Pilpel 2006)).

2.1 Bacterial stealth regulators.

2.1.1 *Cis*-ncRNAs

The plasmid-encoded ncRNAs have served as a paradigm for studying the functions of cis-encoded antisense RNAs (Wagner *et al.* 2002). Antisense RNAs regulate the transcription, stability or translation of mRNA encoding proteins critical for replication or stable plasmid inheritance. These antisense RNAs are constitutively expressed, but are metabolically unstable. As a consequence, changes in plasmid concentrations are reflected in changes of anstisense RNAs. The unique example in this field is the regulation of ColE1 replication by RNAI (Eguchi and Tomizawa 1991). The 108-nt RNA interacts with the RNA primer for DNA replication, leading to a decrease in the frequency, with which the RNA primer extends into a DNA primer and thereof lowering plasmid copy number. The use of antisense RNA to control plasmid copy number is widespread in both gram-positive and gram-negative bacteria. Another example is the pT81 plasmid, increased copy number of which results in respective increase of RNAI (~85nt) and RNAII (~150nt) antisense RNA molecules. RNAI and II pair with and hence stabilize transcription termination site upstream of RepC protein required for initiation of replication (Brantl and Wagner 2000). As the plasmid copy number declines, levels of RNAI and RNAII drop, allowing transcription read-through, subsequent enhancement of RepC expression and initiation of a new round of replication.

Different *cis*-encoded antisense RNAs control a so-called plasmid addiction system, which ensures that plasmid containing bacteria survive, whereas the other ones get killed, thus plasmids acting as lethal timers (Gottesman 2002). The **h**ost **k**illing (*hok*) gene of an R1 plasmid encodes a small toxin that damages bacterial membrane leading to cell death. In plasmid-carrying bacterial cells hok mRNA translation is repressed by plasmid – encoded Sok RNA (**S**uppressor **o**f **k**iller). As long as the equilibrium is maintained for Sok prevailence, bacteria survive, but upon loss of plasmid, Sok no longer represses expression of hok and the toxin kills the bacterial host (Gerdes *et al.* 1997) (Fig. 1-4). In other host-plasmid systems antitoxin is a plasmid-encoded unstable protein, degraded by the ATP-dependant cytoplasmic proteases. By and large, complete base pairing between the plasmid-encoded antisense RNAs and their bacterial targets is not required for regulation. Instead the most critical interactions happen within the short single-stranded regions where the first base pairs bring the antisense RNA and target RNA together in what is known as “kissing complex”. Then the base-pairing is extended into a more stable “extended kissing complex”, which exerts the regulation. In addition, antisense RNAs were

found to regulate phage immunity and growth and transposition, by means of RNAs encoded on the antisense strand to their target gene (Wagner and Simons 1994). One of the key distinctions between plasmid and bacterial regulatory RNAs is that plasmid RNAs are dedicated to regulation of only one target.

There are fewer chromosomal cis-encoded antisense ncRNAs in bacterial genome, which are required to repress the expression of toxic proteins or maintain its mRNA transcription low, and most of them fit the toxin-antitoxin scheme. Some other examples include four long-directed-repeat (LDR) sequences region in E.coli, which express both an mRNA encoding a toxic peptide (ldr) and a *cis*-encoded antisense RNA (rdl) (Kawano *et al.* 2002); and two stretches of antisense RNA – 275 nt RyeA/SraC and 100 nt RyeB – the function of which remains elusive. It cannot be excluded that these antisense RNAs also perform independent functions. Thus one ncRNA (GadY/1S183) serves to stabilize transcription of its sense counterpart gadX mRNA and increases its expression. GadY confers increased stabilization by overlapping and pairing with a 3'UTR of gadX (Opdyke *et al.* 2004).

Almost all of the bacterial antisense RNAs were shown to require an Hfq protein, compositionally and structurally similar to eukaryotic Sm protein, involved in splicing (Schumacher *et al.* 2002). Hfq binds to regulatory RNAs and their targets *in vivo* and stimulates base-pairing. (Moller *et al.* 2002; Zhang *et al.* 2002). This chaperone-like function may be necessary to prevent alternative RNA confirmation of the regulator and its target mRNA.

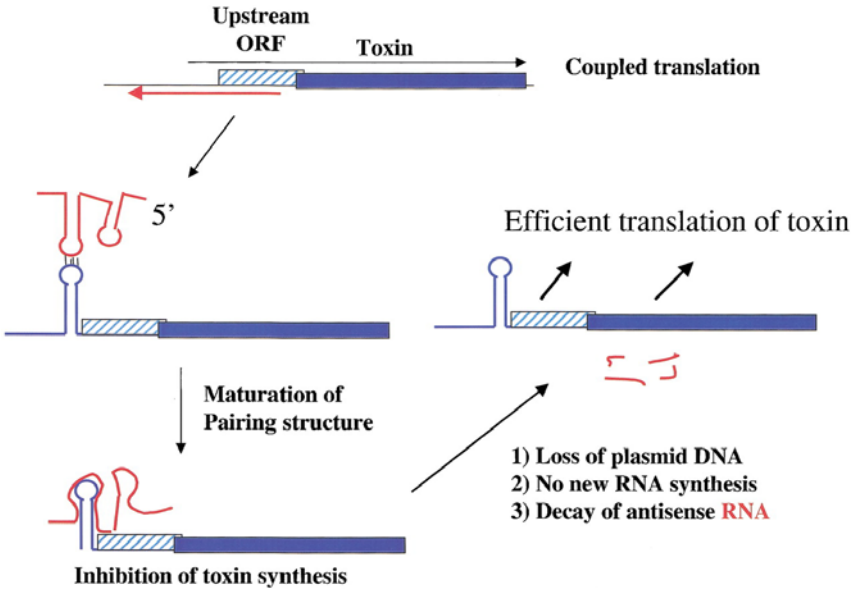


Figure 1-4. Antisense RNA regulating toxin synthesis. In this highly simplified version of the action of the *hok/sok* post-transcriptional killing system of plasmid R1, antisense RNA, encoded on the opposite strand, pairs with regions necessary for translation of an upstream ORF. Translation of the toxin gene requires translation of the upstream ORF. Once new antisense RNA ceases to be made, its decay releases the toxin mRNA for efficient translation. A similar arrangement is found in the regulation of many plasmid replication proteins. (adapted from (Gottesman 2002)).

2.1.2 *Trans*-ncRNAs.

The functions of more than a dozen *trans*-encoded base-pairing RNA regulators have been characterized in *E. coli* (table 1-1). These RNAs destabilize mRNAs and either repress or activate translation.

Table 1-1. Regulatory signal cascades for bacterial regulatory RNAs. (adapted from (Gottesman 2002))

Signals	***Regulator***	***Small RNA***	***Targets***	***Regulatory logic***
oxidative stress	OxyR	OxyS	fhlA (-) rpoS (-) mutagenesis (-)	Repair system priorities
low temperature	None identified	DsrA	rpoS (+) hns (-)	Low temperature requirement for RpoS?
cell surface stress	RcsC/RcsB	RprA	rpoS (+)	Unknown
carbon source preferences	cAMP/CRP	spot 42	galK (-)	Fine-tuning of carbon source metabolism
low iron	Fur (-)	RyhB	sdh (-) ftn (-) fumA (-) others (-)	Intracellular iron use

MicC RNA (109 nt) (Chen *et al.* 2004b) and MicF RNA (93 nt) (Delihas and Forst 2001) repress the translation of the OmpC and OmpF outer membrane porins, respectively, via base pairing with mRNA sequences adjacent to or overlapping the ribosome binding sites. Because MicC and MicF are reciprocally expressed under a variety of environmental conditions, the RNAs contribute to the reciprocal expression of the two porin proteins.

Two RNAs, DsrA RNA (85 nt) and RprARNA(105 nt), are able to release translational block of the *rpoS*mRNA by preventing the formation of an inhibitory secondary structure that normally occludes the ribosome binding site within the long *rpoS* transcript (Lease *et al.* 1998; Majdalani *et al.* 2001; Majdalani *et al.* 2002; Majdalani *et al.* 2005) (Fig. 1-5). Examples of ncRNAs that affect mRNA stability are the RyhB/SraI RNA (90 nt), which is induced under conditions of low iron (Masse and Gottesman 2002), and the SgrS/RyaA RNA (~200 nt), which is induced by elevated phosphosugar levels (Vanderpool and Gottesman 2004). Upon iron starvation, RyhB RNA base pairs with and promotes the degradation of transcripts encoding iron-containing enzymes, thus allowing alternative utilization of the limited iron. Elevated levels of SgrS RNA are associated with decreased levels of the mRNA encoding the major glucose transporter (*ptsG*).

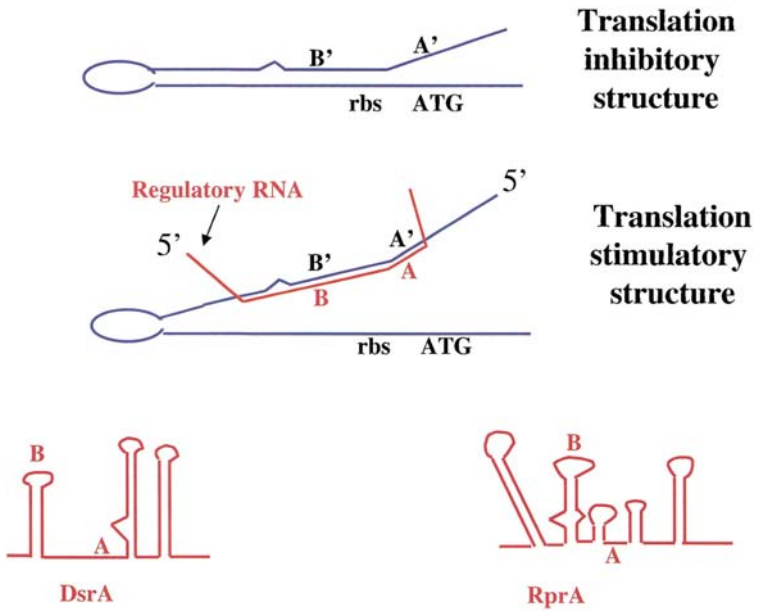


Figure 1-5. Positive regulation by small RNAs. RpoS translation is occluded by a hairpin that includes sequences from the upstream message. rbs - ribosome-binding site; ATG - starting codon of RpoS coding region. At least two small RNAs, DsrA and RprA, are able to increase translation by pairing with the upstream region of the hairpin. The structure of DsrA is based on (Lease and Belfort 2000); the structure of RprA is a computer prediction. A similar example of positive regulation by release of an inhibitory hairpin has been described for RNA III stimulation of -toxin synthesis in *S. aureus*. (adapted from (Gottesman 2002))

Although *trans*-encoded antisense RNAs were first described in *E. coli*, several similarRNAs have recently been discovered in other bacteria. Among these are two ncRNAs (PrrF1 and PrrF2); their expression is induced by low iron in *Pseudomonas aeruginosa*.

In a similar vein, four homologous RNAs (Qrr1, Qrr2, Qrr3 and Qrr4) in *Vibrio harveyi* and *Vibrio cholerae* are required to destabilize the mRNA encoding a key regulator of the quorum-sensing response in which the bacteria monitor their population density (Lenz *et al.* 2004). The reason for multiple homologs is not known, although the presence of multiple RNAs should allow for more nuanced regulation.

Regulation by all of the *trans*-encoded antisense RNAs characterized thus far in *E. coli* requires binding to a shaperon-like common protein, Hfq. Binding to Hfq induces structural changes in RNAs and promotes base pairing between the ncRNA and its mRNA target or in some cases, leads to changes in ncRNA or mRNA accessibility to RNases (Geissmann and Touati 2004; Moller *et al.* 2002; Zhang *et al.* 2002). In addition, binding to Hfq can protect against digestion by the ribonuclease RNase E (Folichon *et al.* 2003; Moll *et al.* 2003). Given that more than one third of all *E. coli* ncRNAs are bound by Hfq and

are suspected to act by base pairing (Zhang *et al.* 2003), *trans*-encoded antisense RNAs appear to be an abundant class of regulatory molecules in bacteria.

2.2 RNA intereference by double-stranded RNA

Eukaryotic cells exhibit several basic responses to dsRNAs.

2.2.1 Interferone response

Since dsRNAs are often associated with viral infection or transposable elements, eukaryotes launch interferon-mediated response, which involves shut down of protein synthesis and induction of interferon. Such general non-targeted response is triggered by introduction of any dsRNa longer than 30 nt. This response is highly sensitive and may be elicited by a single dsRNA molecule (Kumar and Carmichael 1998; Williams 1999) and ultimately culminates in apoptosis.

2.2.2 RNAi and siRNAs

Double-stranded RNA (dsRNA) can induce many different epigenetic gene-silencing processes in eukaryotes, including the degradation of homologous mRNAs known as RNA interference (RNAi) in animals and post-transcriptional gene silencing (PTGS) in plants. Unlike interferon response RNA interference (RNAi) is a highly specific flow of signaling events and molecular interactions, in which dsRNA is processed into small interfereing RNAs (siRNAs), and which has a role in several chromatin and/or genomic DNA modifications.

The understanding of basic mechanism of RNAi is derived mainly from biochemical work in cell extracts from *Drosophila melanogaster* and *Homo sapiens*, complemented by genetic studies in *Arabidopsis thaliana*, *Caenorhabditis elegans* and *Neurospora crassa* (Hannon 2002; Matzke *et al.* 2001a; Matzke *et al.* 2001b; Plasterk 2002; Vance and Vaucheret 2001; Zamore 2002). The dsRNA-induced degradation of homologous RNAs consists of two steps - initiation and effector step (Fig. 1-6). In the initiation step, a long dsRNA is processed into small interfering RNAs (siRNAs) of about 21–23 nt (Zamore *et al.* 2000). This ATP-dependent cleavage is mediated by an RNase-III-like dsRNA-specific ribonuclease, called Dicer in *Drosophila* (Bernstein *et al.* 2001; Zamore *et al.* 2000). Next, in the effector step, the double-stranded siRNAs are incorporated into a multiprotein RNA-induced silencing complex (RISC). RISC subsequently undergoes an ATP-dependent activation step that results in the unwinding of

the double-stranded siRNAs (Nykanen *et al.* 2001). Activated RISC uses a single-stranded siRNA as a guide to identify complementary RNAs (Martinez *et al.* 2002; Nykanen *et al.* 2001), and an endoribonuclease that has yet to be identified, then cleaves the target RNA across from the center of the guide siRNA (Elbashir *et al.* 2001b; Martinez *et al.* 2002). Finally, the cleaved RNA is degraded by exoribonucleases (Hammond *et al.* 2000). siRNAs have been identified in plants, animals, fungi, protozoa and algae, and the apparent products of the *in vivo* endonucleolytic cleavage of mRNAs can be detected in human cells (Holen *et al.* 2002) and tomato (Han and Grierson 2002). On top of that, homologs of some of the key components in the pathway, such as Dicer and the Drosophila RISC protein AGO2, which is a member of the Argonaute gene family (Bohmert *et al.* 1998), have been implicated in RNA-mediated silencing in several eukaryotes (Hannon 2002; Hutvagner and Zamore 2002; Martinez *et al.* 2002; Tabara *et al.* 1999; Vance and Vaucheret 2001; Zamore 2002). In *C. elegans*, for example, RDE-1 (an Argonaute family member) and RDE-4 (a dsRNA-binding protein) are part of a complex with the Dicer ribonuclease and DHR-1 and DHR-2 (putative RNA helicases) (Tabara *et al.* 1999), and are both required to initiate silencing, presumably because they aid the conversion of dsRNA into *si*RNAs (Grishok *et al.* 2000).

Most evidence for subcellular locations of siRNA production and target RNA degradation is consistent with a cytosolic, basic RNAi pathway. In mouse oocytes and Drosophila, RNAi can target cytoplasmic maternal mRNAs (Kennerdell and Carthew 1998; Svoboda *et al.* 2000). RNA viruses, which replicate exclusively in the cytosol, are inhibited by RNA-mediated silencing mechanisms in both plant and mammalian cells (Bitko and Barik 2001).

In human cells, RNAi occurs predominantly in the cytosol (Hutvagner and Zamore 2002; Martinez *et al.* 2002). Injected or fed exogenous dsRNA and viral dsRNA directly enter the cytoplasmic RNAi pathway. In several organisms RNAi can also be induced by inverted repeat transgenes that are transcribed into hairpin dsRNA in the nucleus (Matzke *et al.* 2001a; Vance and Vaucheret 2001). In both plants (Wesley *et al.* 2001) and Drosophila (Kalidas and Smith 2002), inverted repeat transgenes silence more efficiently when the hairpin dsRNA contains an intron and polyadenylation signals that presumably facilitate entry of the dsRNA into the mRNA export pathway (Reed and Hurt 2002).

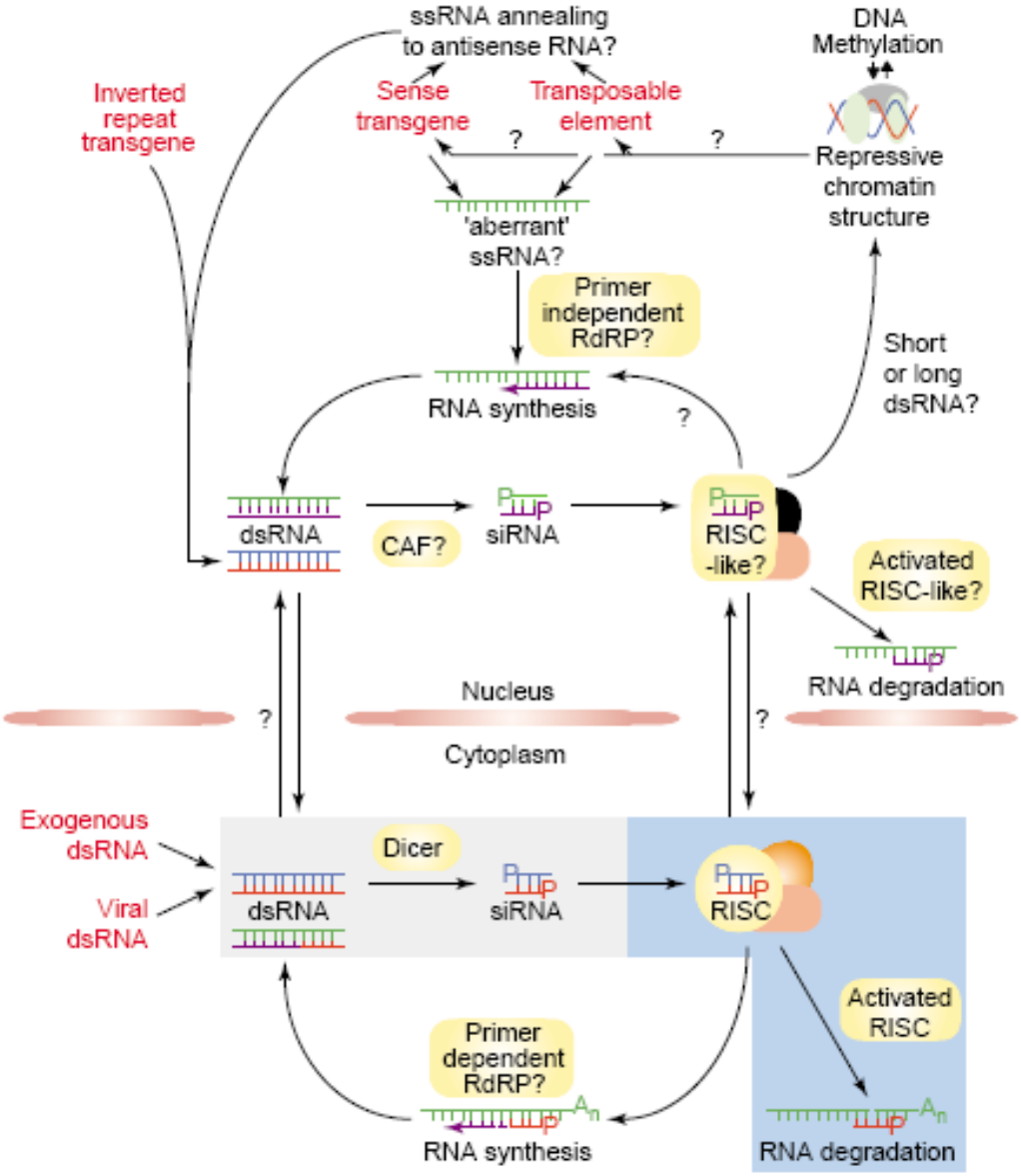


Figure 1-6. Models of molecular pathways involved in double-stranded RNA (dsRNA)-mediated silencing. The basic mechanism, present in most eukaryotes undergoing dsRNA-mediated silencing, is indicated by the gray and blue boxes. The 'triggers' of silencing, either direct sources of dsRNA or transcription units producing single-stranded RNAs that can be presumably converted to dsRNA, are colored red. Green RNA - endogenously transcribed single-stranded RNA; purple RNA - RNA synthesized by a putative RNA-directed RNA polymerase; blue and red RNA - double-stranded RNA introduced exogenously or resulting from viral replication, annealing of complementary ssRNAs and/or hairpin RNA. Proteins or protein complexes are indicated by yellow boxes: CAF, an *Arabidopsis* homolog of Dicer - an RNase-III-like dsRNA-specific ribonuclease; RdRP - an RNA-directed RNA polymerase; and RISC - RNA-induced silencing complex. dsRNA molecules might be delivered differently to the processing Dicer enzymes. Similarly, the RISC and RISC-like complexes might have different components and associated effector proteins depending on their functions. Although a role for dsRNA in directing methylation of homologous DNA sequences has been demonstrated in plants, the molecular machinery involved in this process and the actual nature of the 'guide' RNA have not been resolved. Recent evidence suggests that the RISC complex is equivalent to the miRNP complex in human cells (adapted from (Cerutti 2003)).

Genetic analyses in *C. elegans* suggest that an amplification step might be required for efficient RNA-mediated silencing in several systems. According to such model primary siRNAs (derived from the trigger dsRNA) might prime the synthesis of additional dsRNA, using the target mRNA as a template, in a reaction catalyzed by a putative RNA-directed RNA polymerase (RdRP) (Sijen *et al.* 2001a; Sijen *et al.* 2001b). The newly synthesized dsRNA would then be cleaved by Dicer to generate secondary siRNAs at a sufficient concentration to achieve efficient target mRNA degradation by RISC. In support of this model, the injection of short antisense RNA oligomers into *C. elegans* can trigger silencing of endogenous genes, and this effect is dependent on a functional Dicer (DCR-1) (Tijsterman *et al.* 2002). The results of many experiments in both Drosophila and humans, contradict the amplification hypothesis and argue against an obligatory role for an RdRP in dsRNA-induced RNAi.

In several plant species, dsRNA can direct methylation of homologous DNA sequences (Matzke *et al.* 2001a; Mette *et al.* 2001; Vance and Vaucheret 2001). Methylation of genomic DNA occurs even when silencing is induced by RNA viruses, with sequences homologous to nuclear DNA, that replicate exclusively in the cytoplasm (Waterhouse *et al.* 2001). This suggests that there is communication between the cytoplasm and the nucleus. When the dsRNA has homology to a promoter, it induces transcriptional silencing in association with DNA methylation; while the same homology-dependent mechanism of methylation of a coding sequence leads to PTGS (Pickford and Cogoni 2003).

Connections between the RNAi and the PTGS machinery and chromatin and/or genomic DNA modifications have also been discovered in other organisms. In *C. elegans*, mutations in the putative RNA exonuclease MUT-7 reactivate transgenic arrays that are silenced by a polycomb-dependent, presumably transcriptional, mechanism (Tijsterman *et al.* 2002). Some polycomb group homologs, involved in chromatin repression, are required for RNAi (Dudley *et al.* 2002). Several recent reports have directly implicated the RNAi and PTGS machinery in heterochromatin formation and genome rearrangements (Hall *et al.* 2002). In many eukaryotes, heterochromatin is characterized by a high density of histone H3 methylated at lysine 9 (H3-Lys9). This modification results in the binding to histone H3 of heterochromatin protein 1 (HP1), and presumably other factors, and formation of a transcriptionally repressive chromatin structure. H3-Lys9 methylation also leads to DNA methylation in *Neurospora* and *Arabidopsis* (Lachner and Jenuwein 2002).

In fission yeast, plants and *Drosophila* formation of centromeric heterochromatin is mediated by repeat-associated small RNAs (rasRNA) produced by Dicer from transcripts of centromeric repetitive sequences. They act in association with the RNA-induced transcriptional silencing complex (RITS complex) to direct chromatin modification and transcriptional silencing of centromeric DNA

(Sontheimer 2005; Sontheimer and Carthew 2005). The RISC and RITS complexes are linked not only by Dicer, but also by requirement of Argonaute family proteins, present both in eukaryotes and archaebacteria (Liu *et al.* 2004).

2.2.3 RNA editing

A third response to dsRNA is target RNA editing - deamination of adenosine residues into inosine within a double-stranded homology region. The process is mediated by ADAR enzymes (adenosine deaminases acting on dsRNA) (Bass 2002). Editing modifies the coding potential of the genes, disrupting A to U pairing, as I pairs to C like a purine base. In many instances, ADARs modify intramolecular structure, including inverted repeats imbedded in 3'UTRs (Maas *et al.* 2003). ADARs are also capable of editing larger RNA molecules – a process referred to as hypermutation. Such inosine-containing RNAs are specifically recognized in the nucleus, where they may trigger other signaling processes (DeCerbo and Carmichael 2005; Wang *et al.* 2005b; Zhang and Carmichael 2001).

Components of the RNAi and PTGS machinery are also involved in the processing and function of microRNAs, a class of small RNAs that were originally identified by their role in translational repression in some animals.

2.3 Micro RNAs (miRNA)

2.3.1 A short history of miRNAs

Micro RNAs (miRNAs) are ~21-23nt gene products expressed stage specifically, which control key developmental transitions in worm larvae by acting as antisense translational repressors (Lee and Ambros 2001). First discovered miRNAs were the lin-4 and let-7 in *C.elegans*.

Ambros and co-workers performed genetic screens to characterize the heterochronic gene pathway (the temporal progression of developmental events in C. elegans) and uncovered two transcripts – one small and one even smaller - a 22nt noncoding RNA as the product of the lin-4 gene (Lee *et al.* 1993). Earlier experiments in Sydney Brenner laboratory have shown that loss-of-function mutations in lin-4 disrupted the developmental timing of worms, leading to extra long adults with larval skin, much as did gain-of-function mutations of a protein-coding gene lin-14 (Ruvkun *et al.* 2004; Zamore and Haley 2005). lin-4 RNA repressed the protein levels of lin-14, a gene that functions in the same developmental pathway.

lin-4 RNA could bind with antisense complimentarity, albeit imperfectly, to 3'-UTR sequences found in lin-14 mRNA and repress its translation (Lee *et al.* 1993). For a long time, lin-4 remained the only example of an endogenous 22nt regulatory RNA, in part because back then there was effectively no method to isolate a piece of DNA corresponding to a locus defined by genetics in *C. elegans*. Hence the question stayed open, whether it was just a worm's oddity or a new class of regulatory genes. The situation changed dramatically with the discovery by the Ruvkun laboratory of let-7, a second 22nt RNA that also functioned in the heterochronic gene pathway (Reinhart *et al.* 2000). Similarly to lin-4, let-7 RNA recognized sequences present in the 3'-UTR of its lin-41 mRNA target and repressed LIN-41 protein levels; lin-4 and let-7 RNAs were named small temporal (st) RNAs. A key observation was that let-7 was evolutionary conserved and expressed in all bilaterians tested (Pasquinelli *et al.* 2000). This finding suggested that other 22nt regulatory RNAs might exist in diverse organisms and could be derived, like stRNAs, from double-stranded RNA (dsRNA)-hairpin precursors.

Ruvkun laboratory performed genome database searches using the *let-7* miRNA sequence, which revealed perfect 22 nt matches in the newly emerging *Drosophila* and human genome sequence (Ruvkun *et al.* 2004). The genome regions adjacent to the perfect matches could also fold into bulged and loopy precursors that looked a lot like the probable *lin-4* and *let-7* precursors of *C. elegans,* suggesting that these 22 nt perfect matches were not spurious, but there exists an extensive tiny-RNA world. Importantly, the detection of these homologs demanded and only became possible upon emergence of full genome databases, not the biased "protein world" information of EST databases. Analysis of RNAs from a very satisfying range of animals such as coral, mollusks, annelids, acorn worms, etc., confirmed that *let-7* was conserved across most of animal phylogeny (Pasquinelli *et al.* 2000). The conservation of the *let-7* RNA was the key finding that strongly supported the generality of miRNAs. Bioinformatics searches, which use the characteristic hairpin structure of pri-miRNA as the basis, revealed that there are thousands, of miRNA genes in various genomes, about a third of which are conserved (Grad *et al.* 2003; Lagos-Quintana *et al.* 2001; Lau *et al.* 2001; Lee and Ambros 2001). A few of these genes have now emerged from genetic analysis in *Arabidopsis* and *Drosophila,* but many more have unknown functions (Ruvkun 2001). miRNAs have now been shown to control mRNA abundance in plants, and they could regulate many more RNA steps than translation. In addition, the assignment of the related siRNAs to chromatin silencing in S. *pombe* suggests that miRNAs could act beyond the control of mRNA abundance or translation (Volpe *et al.* 2002).

Meanwhile, the discovery of RNA interference (RNAi) and siRNAs as critical triggers of gene silencing (Fire *et al.* 1998; Hamilton and Baulcombe 1999) suggested that the numerous manifestations of RNA-

mediated gene silencing observed in diverse organisms might have a common mechanism, by which dsRNA (endogenous or experimentally introduced) might give rise to ~22nt RNAs that target and silence homologous mRNA sequences. The Tuschl, Bartel and Ambros laboratories, and soon thereafter the Dreyfuss laboratory, reported the existence in different organisms of 100 endogenous 22nt RNAs. These RNAs are derived from longer hairpin-like precursors and were named microRNAs (Tuschl *et al.* 1995). Ruvkun and colleagues began to explore the action of the RNAi machinery in miRNA maturation and activity. They looked closely at the first RNAi-defective mutants, *rde-1* and *rde-4,* but could not detect any heterochronic defects nor any change in *lin-4* or *let-7* miRNA activity or processing. Soon it was shown that RNAi inactivation of a pair of RDE-1 paralogs and *C. elegans* Dicer disrupt miRNA processing and activity, proving that the RNAi and miRNA pathways were indeed related (Grishok *et al.* 2001).

2.3.2 miRNA biogenesis

In contrast to siRNAs, which derive from dsRNAs hundreds or thousands base pairs long, miRNAs derive from long largely unstructured transcripts (pri-miRNA) containing stem-loop or hairpin structures ~70nt in length (Kim 2005) (Fig. 1-7). The hairpins are cut out of the pri-miRNA by the dsRNA-specific endonuclease Drosha, acting with its dsRNA-binding protein partner DGCR8 in humans or Pasha in flies, to yield a pre-miRNA. Efficient processing of pre-miRNA by Drosha requires presence of extended single-stranded RNA on both 3'- and 5'-ends of hairpin molecule; these motifs could be of different composition while their length is of high importance for processing to take place (Zeng 2006). These findings were confirmed using bioinformatics tools (Kong and Han 2005). Each mature miRNA resides in one of the two sides of the ~30bp stem of the pre-miRNA. In cytoplasm miRNA is excised from the pre-miRNA by another ds-specific endonuclease, Dicer, again acting with a dsRNA-binding protein partner, the tar-binding protein (TRBP) in humans and Loquacious (Loqs) in flies. The pathway in plants varies slightly due to their lack of Drosha homologs; instead, Dicer homologs alone effect several processing steps (Kurihara and Watanabe 2004). When Dicer cleaves the pre-miRNA stem-loop, two complementary short RNA molecules are formed, but only one is integrated into the RISC complex. This strand is known as the guide strand and is selected by the argonaute protein (AGO), the catalytically active RNase in the RISC complex, on the basis of the stability of the 5' end (Preall *et al.* 2006). The remaining strand, known as the anti-guide or passenger strand, is degraded as a RISC complex substrate.

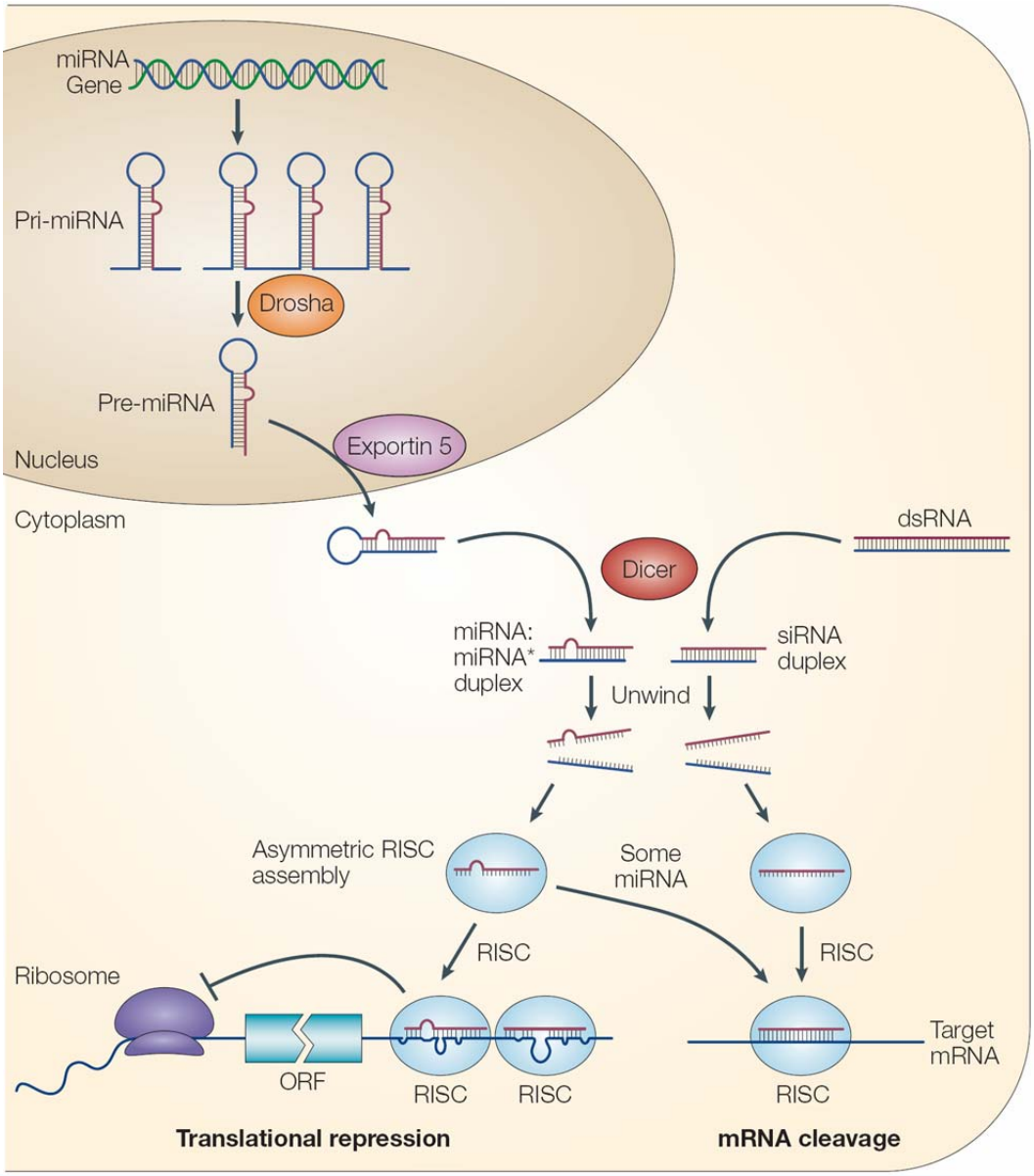


Figure 1-7. miRNA biogenesis and action. miRNA genes are transcribed by RNA Pol II in the nucleus to form large pri-miRNA transcripts, which are capped (7MGpppG) and polyadenylated. pri-miRNA transcripts are processed by the RNase III enzyme Drosha and its co-factor, Pasha, to release the 70-nucleotide pre-miRNA precursor product. (The mature miRNA sequence is shown in red.) RAN–GTP and exportin 5 transport the pre-miRNA into the cytoplasm. Subsequently, another RNase III enzyme, Dicer, processes the pre-miRNA to generate a transient 22-nucleotide miRNA:miRNA duplex. This duplex is then loaded into the miRNA-associated multiprotein RNA-induced silencing complex (miRISC) (light blue), which includes the Argonaute proteins, and the mature single-stranded miRNA (red) is preferentially retained in this complex. The mature miRNA then binds to complementary sites in the mRNA target to negatively regulate gene expression in one of two ways that depend on the degree of complementarity between the miRNA and its target. miRNAs that bind to mRNA targets with imperfect complementarity block target gene expression at the level of protein translation (lower left). However, recent evidence indicates that miRNAs might also affect mRNA stability (not shown). Complementary sites for miRNAs using this mechanism are generally found in the 3' untranslated regions (3' UTRs) of the target mRNA genes. miRNAs that bind to their mRNA targets with perfect (or nearly perfect) complementarity induce target-mRNA cleavage (lower right). miRNAs using this mechanism bind to miRNA complementary sites that are generally found in the coding sequence or open reading frame (ORF) of the mRNA target (adapted from (Zamore and Haley 2005)).

2.3.3 Regulation of gene expression by miRNAs.

miRNAs regulate gene expression in two ways. The annealing of the miRNA to the 3'UTR of target mRNA inhibits protein translation, but sometimes facilitates cleavage of the mRNA (Fig. 1-7).

Cleavage is thought to be the primary mode of action of plant miRNAs. In such cases, the formation of the double-stranded RNA through the binding of the miRNA triggers the degradation of the mRNA transcript through a process similar to RNA interference (RNAi). Even when miRNA is fully complementary to its target, cleavage only occurs when it is bound to the right Argonaute protein (Liu *et al.* 2004; Meister *et al.* 2004). In humans only one of the four Argonaute proteins retains the amino acids required for cleavge catalysis. Argonaute proteins contain two RNA-binding domains: Piwi, which binds to small RNA at its 5'end and PAZ, which binds to the 3'end of a small RNA. The endonuclease that cleaves mRNA target resides in Piwi domain, which is a structural homologue of DNA-dependant RNA endonuclease RNaseH (Song *et al.* 2004). In Drosophila or human cell lysates Ago2 acts as a multiple-turnover enzyme, with each small RNA directing the cleavage of hundreds of target molecules (Haley and Zamore 2004; Hutvagner and Zamore 2002). Upon mRNA cleavage additional proteins may be required to release two pieces of mRNA in an ATP-dependent manner. After release, the 3'end is degraded by a cellular exonuclease XrnI while the 5'fragment is degraded by the exosome, a collection of exonucleases dedicated to 3'-to-5' RNA degradation. A short polyU tail is added to 3' and 5' fragments upon celavage and this as a rule correlates with de-capping (Shen and Goodman 2004).

In other cases, when miRNA only partially pairs with its target, it is believed that the miRNA complex blocks the protein translation machinery or otherwise prevents protein translation without causing the mRNA to be degraded. Binding of a single miRNA is usually insufficient to measurably block translation; instead several miRNAs bind to the same target, providing combinatorial control of gene expression by a set of coordinatedly expressed miRNAs (Bartel 2004). miRNAs were proposed to block translation at the initiation stage (Olsen and Ambros 1999). They might freeze ribosomes in place stalling elongation of a growing protein chain. This idea has been however called into question with the recent discovery that Ago2 bound miRNAs concentrates in "P-bodies", considered sites of mRNA distruction in the cell, whereas free Ago2 remains in the cytosol (Liu *et al.* 2004; Sen and Blau 2005). Moreover, Ago2 associates with decapping enzymes, which remove the 5' 7-methylguanosine from mRNA – a prereqiusite of their degradation in P-bodies. It is tempting to imagine that small RNA molecules guide mRNA confind to degradation to its site – the P-body and that they block translation, by effectively sequestering mRNA in P-bodies free of ribosomes. Fillipowicz and colleagues argue that

translocation to P-bodies is a consequence, rather than a cause of translational repression (Pillai *et al.* 2005).

So far, there is one example of human miRNA that instead of blocking translation, enhances it. Replication of hepatitis C virus (HCV) requires binding of human miR-122 to the 5' non-coding region of the virus (Jopling *et al.* 2005). Thus miR-122 acts as an enhancer of replication for HCV, and only cells expressing miR-122 support efficient HCV replication. It remains obscure, whether this mechanism is unique or represents an undiscovered mode of miRNA action (Zamore and Haley 2005).

miRNAs may also direct transcriptional silencing associated with formation of heterochromatin. It has been first discovered in plants. Heterochromatin is transcriptionally repressed, compact form of chromatin in which the amino terminus of histone H3 is modified by methylation at Lys 9 ("H3K9"). In some organisms, including plants and mammals heterochromatin is hypermethylated. Genetic studies in worms, plants, and *S. pombe* implicate small RNAs as well as canonical components of RNA silencing machinery – Dicer, RdRP and Argonaute in transcriptional silencing (Chan *et al.* 2004; Grishok *et al.* 2005; Robert *et al.* 2005; Volpe *et al.* 2002; Xie *et al.* 2004; Zilberman *et al.* 2003). miRNAs may also target methylation of genomic sites which correspond to targeted mRNAs.

Why are miRNAs so tiny and what predictions about their biology could be made based on their small size? The ~22 nt length is most likely dictated by the enzymology of Dicer (Elbashir *et al.* 2001a). Sometimes, though a 70nt precursor could be functional as well. There are more that 10^{13} different theoretical 22nt sequences and therefore each miRNA has enormous potential in terms of biological specificity. Simultaneously, this relatively short length seems insufficient to include complex structural elements signaling stability, intracellular transport, etc. , in addition to antisense sequence elements.

miRNAs could be regulated by different means depending on their functions in the cell of an animal. Lin-4 and let-7 are developmentally regulated and are expressed at the 2^{nd} and the 4^{th} larval stages of C. elegans, respectively. Temporal regulation of the miRNAs indicates that miRNA coding genes, which seem to come from intergenic regions regulated by autonomous miRNA promoters, respond to a variety of developmental signals. For this type of regulation, key questions to answer would be, what RNA polymerases trancribe these genes, how are their regulatory sequences structured and which transcription factors couple their expression to developmental signals?

In contrast to temporaly regulated miRNAs, some constitutively transcribed miRNAs could be regulated at the level of processing of a mature ~22nt transcript from its 70nt precursor. Some miRNAs identified in C. elegans appear to be processed rather inefficiently (Lee and Ambros 2001). These may represent a

class of miRNAs processed under specific conditions, which implies specific regulation of Dicer and other regulatory proteins.

In some instances miRNas and siRNAs are functionally equivalent (Hutvagner and Zamore 2002; Llave *et al.* 2002b; Rhoades *et al.* 2002; Xie *et al.* 2003).

2.3.4 miRNAs and human diseases.

miRNAs or their machinery might be implicated in several human diseases listed in the table 1-2 below.

Table 1-2. Examples of miRNAs involved in cancer and neurodegenerative diseases in humans.

Type	Disease	Cause	Components involved	Mechanism	Reference
neurodegenerative	SMA - Spinal Muscular Atrophy	Reduced protein levels or loss-of-function mutations of the survival of motor neurons (SMN) gene	SMN complex: Geminin3&4 and miRNPs	Deregulation of miRNA biogenesis or function	(Mourelatos *et al.* 2002)
	FXMR- Fragile X Mental Retardation	Absence or mutations of the fragile X mental retardation protein (FMRP).	FMRP and miRNAs	FMRP represses the translation of specific mRNAs. Identification of Drosophila homologue dFMR1 in RISCs might indicate that miRNAs direct dFMR1 to mRNAs whose translation must be controlled	(Ishizuka *et al.* 2002)
	Early-onset parkinsonism and X-linked mental retardation	unknown	miR-224 gene locus	unknown	(Dostie *et al.* 2003)
cancer	CLL - Chronic Lymphocytic Leukemia & Prostate cancer	deletion localized to chromosome 13q14 (~50% of cases)	miR-15 and miR-16 genes, present within the minimal 30 kb region of loss in CLL	Significant reductions in the levels of miRNAs 15 and 16.	(Calin and Croce 2006)
	Wilms' tumors, primitive neuroectodermal tumors, other types of cancer	multiple	Argonaut gene family (EIF2C1, hAgo3 and hAgo4)	Alterations in the region of chromosome 1p34–35 where Argonaut genes are physically located	(Makeyev *et al.* 2007)
	Adenomatous and cancerous stages of colorectal neoplasia	multiple	miR-143 and miR-145	Reduced steady-state levels of miRNAs	(Michael *et al.* 2003)

A study of mice altered to overexpress c-myc shows that miRNA has an effect on the onset of cancer. Mice that were engineered to produce a surplus of types of miRNA found in lymphoma cells developed

the disease within 50 days and died two weeks later. In contrast, mice without the surplus miRNA lived over 100 days (He *et al.* 2005). Another study found that two types of miRNA inhibit the E2F1 protein, which regulates cell proliferation. miRNA appears to bind to messenger RNA before it can be translated to proteins that switch genes on and off (O'Donnell *et al.* 2005).

By measuring activity among 217 genes encoding miRNA, patterns of gene activity that can distinguish types of cancers can be discerned. miRNA signatures may enable classification of cancer. This will allow doctors to determine the original tissue type which spawned a cancer and to be able to target a treatment course based on the original tissue type. miRNA profiling has already been able to determine whether patients with chronic lymphocytic leukemia had slow growing or aggressive forms of the cancer (Lu *et al.* 2005).

Further finding are anticipated to strengthen that connections between miRNAs and human diseases and reveal more gene networks that they control. Understanding of the regulation of RNA-mediated gene expression will possibly lead to the development of novel therapeutic approaches that will be likely to revolutionize the practice of medicine (McManus and Sharp 2002).

miRNA has been shown to be related to heart disease (Zhao *et al.* 2007). Mice deficient in a muscle-specific miRNA had a high rate of the most common congenital heart disease - ventricular septal defects characterized by the holes between the left and the right ventricles of the heart. Such mice also showed hyperplasia - an increase of the number of cardiac muscle cells that leads to heart enlargement - and abnormalities in cardiac conduction.

2.4 Other ncRNAs

2.4.1 Mechanisms of sense-antisense correlated regulation and their action on target genes.

One of the central questions in the vein of antisense transcriptional regulation is how is it itself regulated?

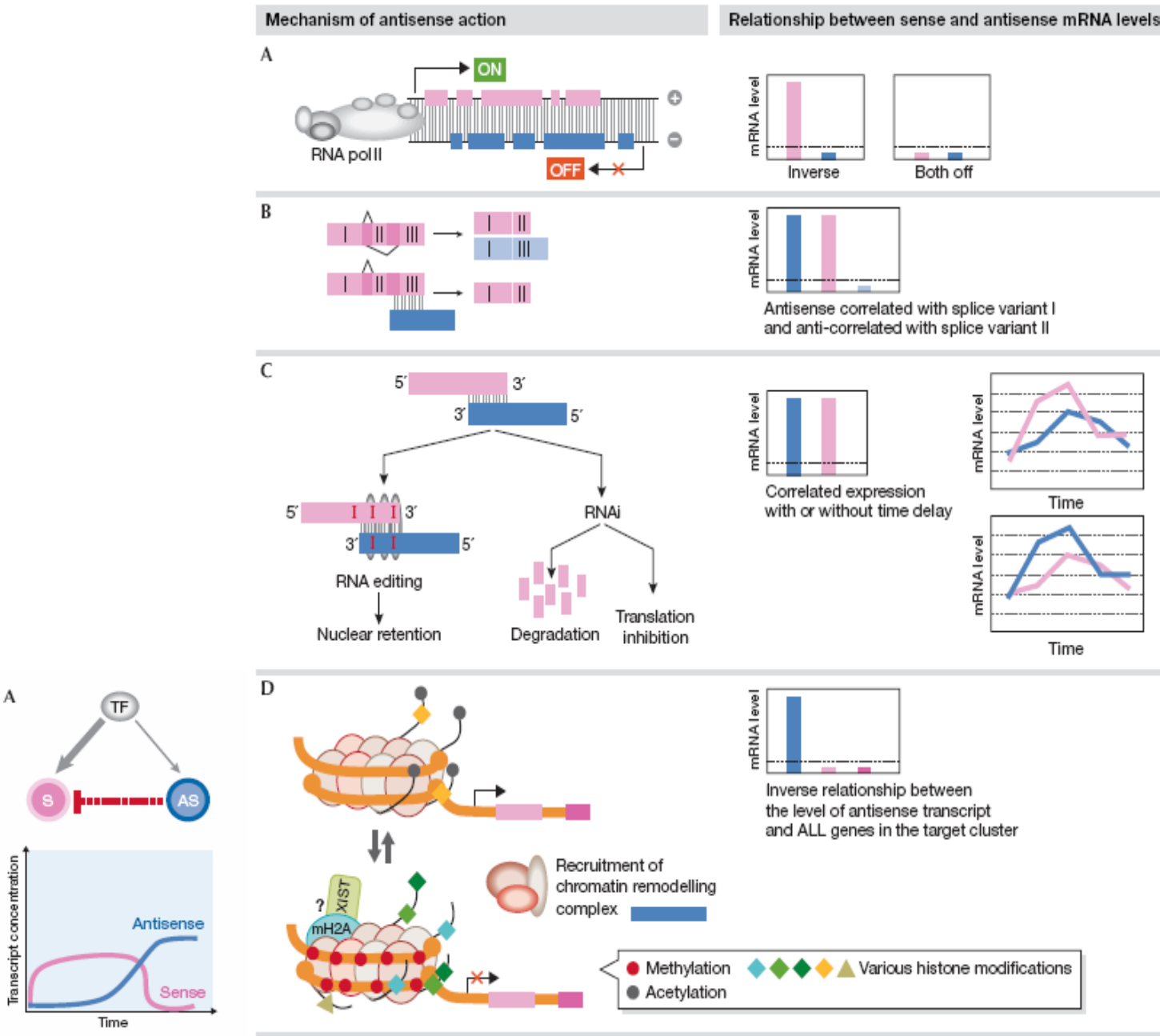
Chen and colleagues found that sense–antisense gene pairs tend to be co-expressed or inversely expressed more frequently than would be expected by chance (Chen *et al.* 2005). Moreover, co-expressed and inversely expressed sense–antisense pairs have striking conservation throughout evolution. Both conservation and coupled sense–antisense expression were more prevalent in tail-to-tail NAT pairs, suggesting that such an orientation is not only the most abundant, but also more likely to have a

regulatory function (Sun *et al.* 2005). The observation of both negative and positive correlations of sense–antisense levels suggests that their mechanisms of action might be diverse. Altogether five general mechanisms have been suggested (Lavorgna *et al.* 2004): *promoter competition*, *transcriptional interference, RNA masking, double-stranded RNA (dsRNA)-dependent mechanisms and chromatin remodelling* (see below). Each mechanism requires different associations between sense and antisense expression patterns, which are coupled to serve different regulatory purposes. Some mechanisms require the concomitant presence of sense and antisense transcripts, whereas others impose their mutual exclusion. Furthermore, because the type of coupling is characteristic of the regulatory mechanism, the relationship between sense and antisense transcription profiles can hint at the mechanism at work as well as the ultimate biological outcome. To illustrate this point, Lapidot and Pilpel have predicted two biological outcomes that might result from a delayed initiation of transcription between the sense and antisense transcripts (or vice versa) (Fig. 1-8). If the sense gene is initially transcribed up to a certain level, then antisense transcription begins and subsequently promotes sense degradation, then the anticipated outcome is a delayed shutdown of the sense gene. Second, if antisense transcription precedes sense transcription, the biological outcome might be the dampening of noise in the level of the sense transcripts; the antisense level sets a threshold and only sense transcripts that exceed it are effectively expressed. Noise dampening was shown to be achieved by microRNAs as well (Hornstein and Shomron 2006). Differences in transcription activation times might be encoded by differential affinities of the sense and antisense promoters to a shared transcription factor, assuming that such a regulator is an activator and that it accumulates with time. Sense and antisense transcripts might be regulated not only at the transcriptional level, but also at the level of mRNA stability. Therefore, differences in mRNA half-lives of the two transcripts might also be predictive of antisense function.

Figure 1-8. Differences in activation times of the sense compared with the antisense transcript. Such differences might be easily encoded in differential affinities to a shared transcription factor, assuming that this transcription factor is an activator and that it accumulates with time. (**A**) A higher affinity of the transcription factor to the sense transcript might result in a delayed shutdown, whereby the transcription factor initially activates transcription of the sense messenger RNA up to a certain level and only then is triggered by antisense transcription. The delayed antisense transcription prevents the sense transcript from exceeding the level it has reached when antisense transcription is switched on. (**B**) A higher affinity of the transcription factor to the antisense transcript. In this case, antisense transcription precedes sense transcription and acts as a buffer for the sense transcript.When the transcription factor accumulates, transcription of sense mRNA begins, but only sense transcripts exceeding the threshold set by the antisense level can be effectively translated. This generates a step-like function in the concentration of the sense transcript. Fluctuations in the amount of sense transcript below the threshold are dampened. (adapted from (Lapidot and Pilpel 2006)).

The relationship between sense and antisense expression patterns can be best described if we scrutinize all four models of antisense regulation (Fig. 1-9).

Figure 1-9. The main mechanisms by which natural antisense transcripts regulate gene expression. (**A**) Transcriptional interference. Two bulky RNA polymerase II complexes on opposite DNA strands might collide with and stall one another. The interference occurs mostly in the elongation step, resulting in either transcription arrest or transcription in one direction (sense or antisense) only. Such a mechanism might occur in cases in which inverse expression is observed. (**B**) RNA masking. A specific case is shown in which the antisense masks a splice site on the sense pre-mRNA sequence. This prevents a given splice variant from being formed and shifts the balance towards splice variants that do not require splicing of the masked region. Such a mechanism could be observed by correlated expression of the antisense and favoured splice variant and an inverse relationship with the repressed variant. (**C**) Double-stranded RNA-dependent mechanisms such as RNA editing and RNA interference require the simultaneous presence of sense and antisense transcripts for duplex formation, and might



therefore account for the observed co-expression of numerous sense–antisense pairs. A delay in expression of sense compared with antisense (or vice versa) is also possible as long as there is a period in which both transcripts are present. (**D**) Chromatin remodelling. Transcription of non-encoding antisense transcripts is involved in monoallelic gene expression, including genomic imprinting, X-inactivation and clonal expression of lymphocyte genes. In these processes, antisense transcripts have been suggested to silence the expression of nearby gene clusters by chromatin remodelling, most likely through the recruitment of histone-modifying enzymes. If such mechanisms are in action, an inverse expression profile of the antisense compared with all genes in the silenced cluster would be expected. (adapted from (Lapidot and Pilpel 2006)).

Transcriptional interference. The presence of an overlapping transcriptional unit might stall sense transcription owing to the collision of two bulky RNA polymerase II complexes on opposite strands. This is most apparent in the transcription elongation step as has been shown for the yeast gene pair GAL10 and GAL7 (Prescott and Proudfoot 2002). Competitive transcriptional interference could be the underlying mechanism when anti-correlated expression levels of sense and antisense are observed. Alternatively, such interference might result in the shutdown of both transcripts (Fig. 1-9A). Conversely, the initiation of transcription on one strand may help activate transcription on the opposite strand, by altering local chromatin structure (Navarro *et al.* 2005) or drawing adjacent promoters into an active transcriptional state (Cook 2003).

Competing for transcription factors or their binding sites. Cis-antisense transcription may regulate expression of overlapping genes by competing for or sharing of transcription factors. This is apparent for SAP with juxtaposed promoter (5'- 5' overlap). It is also possible for 3'- 3' or complete sequence overlap to share common transcription factor binding sites since genomic regions harboring regulatory elements can stretch as much as 1Mb in either direction from transcriptional unit (Tommasi and Pfeifer 1999). Sharing of a common trans-acting factor may lead to co-expression of overlapping transcripts, whereas negative correlation would be expected if distinct transcriptional factors compete for the overlapping binding sites.

RNA masking. Sense–antisense duplex formation might mask *cis* elements residing in either of the transcripts and impede processes that require protein–RNA interactions such as splicing, mRNA transport, polyadenylation, translation and degradation. The best characterized example of this mechanism is the antisense transcript for the thyroid hormone receptor gene *erbA*α, which shifts the balance between two splice variants through the masking of a splice site (Hastings *et al.* 1997). Such a mechanism would result in a correlated expression level of the antisense and the regulated transcript (Fig. 1-9B). In addition antisense transcript can effectively titrate out functioning RNA (Ogawa and Lee 2002).

dsRNA-dependent mechanisms and RNA interference. Antisense transcripts might function through the activation of dsRNA-dependent mechanisms such as RNA editing and RNA interference (RNAi). Such mechanisms require the simultaneous existence of sense and antisense transcripts for duplex formation, and might therefore account for the observed co-expression of numerous sense–antisense pairs (Fig. 1-9C) (Chen *et al.* 2005). RNA editing is thought to constitute part of the nuclear defense strategy against dsRNA. Hyper-editing of long, perfect RNA duplexes can result in their nuclear retention (Zhang and Carmichael 2001) or cytoplasmic degradation (Scadden and Smith 2001a; b). Regulation through RNA editing is therefore not likely to be one of the main mechanisms for antisense action.

RNAi is another component of the defence against dsRNA. As described above RNAi acts within the RNA-induced silencing complex (RISC). The RISC eventually either degrades cognate mRNAs with great specificity or represses their translation (Meister and Tuschl 2004; Mello and Conte 2004). For example, salt tolerance in *Arabidopsis* is regulated by two small interfering RNAs (siRNAs) produced from a pair of tail-to-tail overlapping protein-encoding genes: *P5CDH* (a stress-related gene) and *SRO5*. Salt induces *SRO5* transcription. When both of the genes are transcribed, an RNA duplex is formed and siRNAs are produced that ultimately cleave the P5CDH transcript (Borsani *et al.* 2005). The same mechanism could apply to other eukaryotic *cis*-NAT pairs. Other processes shown to involve dsRNA are the response to iron deficiency in cyanobacteria (Duhring *et al.* 2006) and the maintenance of male fertility in *Drosophila* (Aravin *et al.* 2001). So far, however, there has been no evidence for mammalian antisense transcripts acting through duplex formation.

Antisense involvement in methylation and monoallelic expression. dsRNA can induce the methylation and silencing of corresponding genes. For example, thalassaemia—a form of anaemia—is caused by antisense-induced DNA methylation (and silencing) of the human haemoglobin 2 gene (Tufarelli *et al.* 2003). Monoallelic expression includes X-chromosome inactivation, genomic imprinting and allelic exclusion in B and T lymphocytes. X-chromosome inactivation is a mechanism that balances the expression of X-chromosome-encoded genes in mammalian females. The silencing of one of the X chromosomes is mediated through a large non-encoding RNA (*Xist*), which recruits a histone-modifying protein complex. *Xist* is repressed by its antisense *Tsix*, thus the X chromosome expressing the antisense remains active (Ogawa and Lee 2002). Imprinted genes are genes for which only one allele—maternal or paternal—is actively transcribed. There are about 100 known human and mouse imprinted genes; they are clustered in the genome and often have both DNA methylation and non-encoding antisense transcripts.

Several studies have indicated that imprinting is not mediated through the formation of a sense–antisense RNA duplex (Sleutels *et al.* 2003; Thakur *et al.* 2004), but rather through the modification of chromatin structure or methylation patterns in the vicinity of the imprinted allele. In lymphocytes, immunoglobulins and T-cell receptors undergo clonal selection through which one allele is silenced while the other undergoes recombination. Extensive antisense transcription occurs before and during recombination and is believed to function by inducing an open chromatin structure that is accessible to recombination. In all these cases, non-encoding antisense transcription affects an entire gene cluster, rather than merely the overlapping sense transcript, and exerts its effect by chromatin remodeling, probably through the recruitment of histone-modifying enzymes. Therefore an inverse expression profile would be anticipated for the antisense and all the genes in the silenced cluster (Fig. 1-9D). Despite these well-characterized cases, it is not clear which of these mechanisms might apply to a wider set of antisense RNAs.

2.4.2 ncRNAs that modify protein activity

Not all regulatory RNAs act by base pairing. In recent years several regulatory RNAs that bind to proteins and modify their activities have been characterized.

2.4.2.1 ncRNAs that modulate transcription

A number of different bacterial and eukaryotic ncRNAs bind to and modulate the activities of proteins that impact transcription. Although the precise mechanisms of action are not always understood, it has been postulated that at least a subset of these RNAs act by mimicking nucleic acid interactions normally carried out by the target protein. Several examples of such RNAs are listed in table 1-3 below.

Table 1-3. ncRNA that modulate transcription

Species	ncRNA	Experimental Approach	Function	Mechanism	Reference
E. coli	6S RNA (184nt)	Co-immuno-precipitation & UV cross-linking	Inhibits transcription. Required for optimal long-term survival Increases in response to shortage of nutrients upon entry into stationary phase.	Binds directly to the housekeeping form of RNA pol (σ70-RNA polymerase), repressing σ70-dependent transcription during the stationary phase. 6S RNA secondary structure mimics the conformation of the promoter DNA formed during transcription initiation, acting as a competitive inhibitor for transcription.	(Trotochaud and Wassarman 2004; Wassarman and Storz 2000)

Mus. musculus	B2 RNA (178 nt) expressed by RNA pol III from short interspersed repetitive elements (SINES)	Co-immuno-precipitation and binding; in vivo and in vitro transcription	B2 RNA inhibits RNA pol II	B2 binds to RNA polymerase II upon heat shock, preventing the formation of active preinitiation complexes	(Allen *et al.* 2004; Espinoza *et al.* 2004)
Mammals	7SK RNA (330 nt)	Co-immuno-precipitation	inhibits the function of the transcription elongation factor P-TEFb	Directly binds to P-TEFb	(Nguyen *et al.* 2001; Yang *et al.* 2001)
H. sapiens	U1 snRNA – a core component of the splicesome	Transcription assays	stimulates transcription initiation	binds to the general transcription factor TFIIH	(Kwek *et al.* 2002)
H. sapiens	steroid receptor RNA activator (SRA) (700–850 nt)	screen for cofactors of the steroid hormone receptors; mutational studies	acts as a co-activator to stimulate the transcription of steroid receptor-dependent genes	Variety of domains are required for the co-activation; mechanism not completely understood.	(Lanz *et al.* 1999)
H. sapiens	Neuronal NRSE dsRNA	screen short RNAs from adult hippo-campal neural stem cells; mobility shift assays	by over-expression antagonizes the effects of the NRSF/REST repressor	Binds to the NRSF/REST protein that restricts neuronal gene expression to neurons. has sequence similarity to the NRSF/REST DNA-binding site, hence could compete for NRSF/REST binding to the promoter sequences.	(Kuwabara *et al.* 2004)

2.4.2.2 RNAs that modulate mRNA stability and translation

A few bacterial and eukaryotic RNAs have been found to bind to and modify the activities of proteins that regulate mRNA stability and translation. Some of these regulatory RNAs appear to be acting by mimicking the structures of other nucleic acids, in this case other RNAs.

Bacterial CSRB/RSMY family of RNAs A family of homologous RNA-binding proteins, including CsrA of *E. coli* and RsmA of *Pseudomonas* species, regulate glycogen biosynthesis, flagellar motility, and biofilm formation in bacterial cells (Romeo 1998). These proteins bind to the 5' regions of the target mRNAs, blocking translation initiation and stimulating mRNA decay or, in other cases, stimulating translation and blocking mRNA decay (Dubey *et al.* 2003). A family of RNAs, of which the CsrB (360 nt) and CsrC RNAs (270 nt) of *E. coli* and RsmY (118 nt) and RsmZ RNAs (127 nt) of *Pseudomonas fluorescens* are representative, blocks the actions of the CsrA and RsmA proteins (Valverde *et al.* 2003; Weilbacher *et al.* 2003). Each ncRNA of the CsrB/RsmY family of regulatory RNAs modulates mRNA stability and translation by acting as RNA mimics, sequestering multiple copies of the CsrA and RsmA proteins and blocking their functions.

Dendritic BC1 RNA The FMRP protein is an RNA-binding protein that is highly expressed in brain. Mutations associated with the absence of the FMRP protein or altered FMRP lead to the fragile X syndrome, the most frequent cause of inherited mental retardation. Similar to the bacterial CsrA and RsmA proteins, FMRP represses the translation of target mRNAs. In addition to binding to the target mRNAs, FMRP has been found to bind the BC1 RNA (~150 nt) (Zalfa *et al.* 2003), an RNA transcribed by RNA polymerase III in specific neuronal cells (Tiedge *et al.* 1991). Unlike the inhibitory effect exerted by the CsrB/RsmY family of RNAs, BC1 promotes the interaction between FMRP and its target mRNAs, possibly via base-pairing interactions.

2.4.3 ncRNA implication in different cellular events.

Cell stress

GADD7 and ADAPT33 were cloned as two stress-inducible ncRNA genes under different conditions (Hollander *et al.* 1996; Wang *et al.* 1996). GADD has been suggested to have a tumor suppressor activity. ADAPT33 was identified as a novel component of the apoptotic pathway in eukaryotic cells (Wang *et al.* 2003).

T cell activation

In an attempt to identify new genes involved in regulation and activation of T cells the NTT (non-coding transcript of T cells) and BIC ncRNA were identified. (Haasch *et al.* 2002; Liu *et al.* 1997).

Neural function and differentiation

BC1 and BC200 are expressed in mouse and human nervous system, respectively (DeChiara and Brosius 1987; Martignetti and Brosius 1993). BC1 is a non-coding transcript specifically targeted to dendritic domains in neurons (Tiedge *et al.* 1991). BC1 KO mice show changed behavioral patterns and lower survival rate. In addition, deregulation of BC1 expression was associated with synaptic dysfunction phenotype in Fragile X syndrome (Zalfa *et al.* 2003).

NRSE ncRNA (Neuron Restrictive Silencer Element) is able to change neural stem cell fate through interacting with NRSF/REST transcriptional machinery, resulting in the transition from neural stem cells to differentiated neuronal cells (Kuwabara *et al.* 2004).

2.4.4 ncRNA implication in disease

2.4.4.1 ncRNAs implicated in cancer

Recent transcriptome analysis comparing tumor cells to normal ones has provided strong evidence that defects in ncRNAs might occur in tumors. Several ncRNAs and cancers associated with their deregulation are listed in table 1-4.

Implication of an extensive number of ncRNAs in cancer biology indicates that non-protein-coding genes maybe as important for tumorigenesis as previously identified and studied in detail protein-coding ones. This revelation opens the door for new therapeutic approaches.

Table 1-4. Examples of ncRNAs implicated in cancer.

Non-coding transcript	Organism	Size (nt)	Genome map	Deregulation	Type of cancer
BC200	Human	200	2p16	Over-expression	Breast, cervix, esophagus, lung, ovary, parotid, and tongue
MALAT-1 (Metastasis Associated in Lung Adeno-carcinoma Trancript)	Human	8000	11q13	Over-expression	NSCLC (Non Small Cell Lung Carcinoma)
H19	Human	2700	11p15.5	Loss of imprinting, over-expression	Liver and breast
miR-143	Human	22	5	Down-regulation	Colon
miR-145	Human	22	5	Down-regulation	Colon
miR-155/BIC	Human	22	21q21	Over-expression	Burkitt and B cell lymphomas
miR-15a	Human	22	13q14	Deletion and/or down-regulation	B-CLL (B-cell Chronic lymphocytic leukaemia)
miR-16a	Human	22	13q14	Deletion and/or down-regulation	B-CLL
let-7	Human	21–25	22	Down-regulation	Lung adenocarcinoma
BIC	Avian / human	1000–2600/800–1700	Unknown/21q21	Over-expression	Lymphomas and leukemia

2.4.4.2 ncRNAs in nervous system and neurological diseases.

ncRNAs are implicated in quite a number of neurological disorders (table 1-5).

RNA editing in humans plays an important role in CNS functioning. RNA editing exhibits precise CNS regional specificity and essential regulatory roles during neuronal maturation (Bernard *et al.* 1999; Kohr

et al. 1998). RNA editing can also affect multiple sites on the same RNA with diverse functional outcomes catalysed by different ADARs (Valente and Nishikura 2005). ADAR mutants exhibit complex behavioural defects in *C. elegans*, *Drosophila* and mice (Reenan 2001; Tonkin *et al.* 2002). Moreover, abnormalities in RNA editing have been implicated in a spectrum of nervous system disorders including Alzheimer's and Huntington's diseases, amyotrophic lateral sclerosis, epilepsy, schizophrenia, depression, suicidal ideation, autosomal dominant episodic ataxia typeI and Prader-Willi and Angelman syndromes (reviewed in (Valente and Nishikura 2005)).

Non-coding RNAs are important for imprinting in brain. Imprinted genes have essential roles in both neural development and adult CNS functioning, and alterations in their expression profiles are associated with a spectrum of complex neurodevelopmental and neuropsychiatric disorders (Costa 2005; Davies *et al.* 2007; Davies *et al.* 2005). These allele-selective genes exhibit preferential and exquisite cell-specific patterns of expression within the brain, and are frequently processed from larger transcriptional units encompassing multiple tandemly repeated snoRNAs and miRNAs (Lewis and Reik 2006; Seitz *et al.* 2004; Sleutels *et al.* 2002). These imprinted loci also generate a complex spectrum of spliced and unspliced larger ncRNAs of presently unknown function (Davies *et al.* 2005; Furuno *et al.* 2006; O'Neill 2005). Additional ncRNAs associated with imprinted loci include the production of antisense RNAs to reciprocally imprinted neighbouring protein-coding genes (Davies *et al.* 2005; Sleutels *et al.* 2002). The seminal role of imprinted genes in regulating distinct brain signaling systems and in mediating brain–behaviour relationships is illustrated by the spectrum of neurological diseases associated with parent of origin effects and caused by disruptions in imprinted loci: autism, schizophrenia, attention deficit hyperactivity disorder, bipolar disorder and Tourette's syndrome (Davies *et al.* 2005).

Table 1-5. Examples of ncRNAs implicated in neurological diseases.

Non-coding transcript	**Organism**	**Size (nt)**	**Genome map**	**Deregulation/function**	**Disease**	Reference
Prion associated RNAs	Human	Unknown	Unknown	Prion PrP(c) to PrP (Sc) conversion?	Prion Pathologies	(Jeong *et al.* 2005)
IPW	Human	2200	15q11–q13	Deletion? Deregulated expression?	PWS (Prader-Willi Syndrome)	(Valente and Nishikura 2005)
ZNF127AS	Human	6166	15q11–q13	Deletion? Deregulated expression?	PWS	(Valente and Nishikura 2005)
UBE3A/SNURF-SNRPN	Human	Several small ncRNAs	15q11.2	Over-expression	AS (Angelman Syndrome)	(Valente and Nishikura 2005)
H19	Human	2700	11p15.5	Micro-deletion	BWS (Beckwith-	(Niemitz *et al.* 2004; Sparago

					Wiedemann Syndrome)	*et al.* 2004)
LIT1	Human	Unknown	11p15.5	Micro-deletion	BWS	(Niemitz *et al.* 2004; Sparago *et al.* 2004)
SCA8	Human	32,359	13q21	Over-expression causes progressive retina neurodegeneration in flies. SCA8 antis-sense target is a brain-specific transcript of actin protein KLHL1.	SCA (Spinocerebella ataxia)	(Koob *et al.* 1999; Nemes *et al.* 2000)
DISC2	Human	15,002	1q42.1	Translocation detected in family linkage studies	Schizophrenia and bipolar affective disorders	(Blackwood *et al.* 2001; Millar *et al.* 2000)
BC200	Human	200	2p16	Down-regulation	Alzheimer's	(Valente and Nishikura 2007)

3. Systematic approaches for identifying ncRNAs

Most of the ncRNAs recognized today were discovered either in genetic screens or even by accident. This is primarily due to their lack of defined sequence features, which are easy to identify by manual sequence analysis, such as start and stop codons of protein-coding genes. ncRNA genes also are missed in genetic studies because of their small size and resistance to frameshift and nonsense mutations.

3.1 Computational Approaches

Computational approaches have been quite successful in finding families of ncRNAs with well-defined sequence elements or characteristics, such as the C/D box family of snoRNAs (Lowe and Eddy 1999; Omer *et al.* 2003). However, few regulatory RNA families contain such defined elements. Approaches based on straight sequence conservation between related species in intergenic regions, alone and in combination with algorithms based on other criteria listed below, have been very successful in predicting ncRNA genes in many pro- and eukaryotic species. Other mentioned algorithms are based on predictions of RNA structure conservation in intergenic regions, the presence of binding sites for specific DNA-binding proteins as well as promoter and terminator sequences in the intergenic regions, finally extracting features of known ncRNAs using a machine learning approach to search for other ncRNAs (Storz *et al.* 2005).

The computational approaches have led to the identification of many ncRNAs in bacteria, yeast, *C. elegans*, and *A. thaliana*, but they are limited because they have focused on the intergenic regions and thus have missed ncRNAs encoded within protein-coding regions of the genome, such as expected for

cis-encoded antisense RNAs. Most of the computational approaches also rely heavily on sequence conservation and therefore overlook ncRNAs that are species-specific or are less well conserved.

3.2 Experimental Detection of ncRNAs

Various experimental strategies, altogether termed 'Experimental RNomics' (Huttenhofer *et al.* 2002; Huttenhofer and Vogel 2006), were employed to identify novel ncRNAs in genomes of model organisms. They comprise: (i) RNA sequencing (enzymatically or chemically) as the most traditional method to reveal novel ncRNA species; (ii) parallel cloning of many ncRNA by generating specialized cDNA libraries; (iii) the use of microarrays to predict ncRNAs that are expressed under a given experimental condition; (iv) 'genomic SELEX' and its potential application to select ncRNA candidates from the sequence space represented by the genome of an organism of interest (Fig. 1-10). They all are all valid for certain classes of ncRNAs and all carry certain disadvantages.

RNA sequencing approach was successful in the detection of very abundant RNAs, however, it has a number of drawbacks. Firstly, the low abundance ncRNAs molecules are significantly overlooked; secondly they are often masked by the highly abundant mRNAs in the same size range. This could be resolved by 2D gel electrophoresis. Thirdly, chemical or enzymatic sequencing does not always result in unambiguous modification and cleavage of nucleotides, because RNases are not strictly specific for a distinct base but possess residual cleavage activity for other bases, thus obscuring the readout of obtained sequence data. Finally, due to the resolution capacity of polyacrylamide gels, sequencing is limited to RNAs sized—at the most—a couple of hundred nucleotides.

The advantage of direct RNA sequencing, as compared with sequencing cDNA clones generated from ncRNAs is that reverse transcription step is not involved and hence secondary/tertiary structures of ncRNAs do not interfere with cDNA synthesis.

In the past, numerous studies have been performed to identify ncRNAs in the genomes of model organisms by *constructing specialized cDNA libraries*. In the mouse analysis of 5000 cDNA clones derived from size-selected RNAs (50–500 nt), identified 201 candidates for ncRNAs, about half of which belongs to the class of snoRNAs (Huttenhofer *et al.* 2001). This study was followed by using a similar approach for *Arabidopsis thaliana* (Marker *et al.* 2002), *Drosophila melanogaster* (Yuan *et al.* 2003), the two archaeal species *Archaeoglobus fulgidus* and *Sulfolobus solfataricus* (Tang *et al.* 2002) and the eubacteria *E.coli* (Vogel *et al.* 2003) *and Aquifex aeolicus* (Willkomm *et al.* 2005).

The downfalls of this method are that it might not always be possible to reverse transcribe an ncRNA into cDNA because of its structure or base or backbone modifications. Thus, a cDNA library neither reflects all the pallet of ncRNAs in a cell, nor does it necessarily reflect—by number of individual cDNA clones—the abundance of the respective ncRNA. The rationale behind this is that less structured/modified ncRNAs are more easily reverse transcribed than others and will be overrepresented within a cDNA library; similarly, smaller ncRNAs will be more abundant than longer ones, since they are more likely to be fully reverse transcribed.

In addition, by the very nature of a cDNA expression library, only those ncRNA species will be detected, which are transcribed from a genome. This might depend, however, on a specific developmental state of the organism or on expression in a certain tissue.

Microarrays have become the preferred method to monitor the levels of many transcripts in parallel and often at the whole-genome level. Microarrays were mostly used for mRNA expression profiling, hence their main caveat for the discovery of ncRNAs, was—and in may cases still is—the design of the commercially available microarrays (discussed below). Since tailored for mRNA profiling, most of these arrays carry probes only for coding regions, thus transcripts from non-coding genome regions were simply missed. Development of new high-density array technology extended the opportunities for ncRNA detection, leading to a flood of various ncRNA species from the whole-genome transcriptome studies in different organisms (Bentwich *et al.* 2005; Hiley *et al.* 2005; Selinger *et al.* 2000; Tjaden *et al.* 2002). The major issue for microarray experiments is sample preparation, which may cause noise in the data. However, because sample preparation consists of several steps, there are many possibilities to prepare it and moreover the nature of hybridization material can vary, array technology bears a great improvement potential, also in the sensitivity of ncRNAs detection.

The techniques discussed so far allow identification of ncRNAs from the pool of expressed cellular RNAs after co-purification with proteins, i.e. by cloning, direct sequencing or microarray analysis. Many ncRNAs form ribonucleo-protein particles (RNPs) at various time points in their life cycle. Such RNA-binding proteins may help an ncRNA fold into its active conformation, shield it from nucleases prior to exerting its function or promote its annealing with target RNAs up to guiding a protein to its proper target. Other ncRNAs interact with proteins to directly regulate their activity. Given that many such proteins bind their RNA ligands in a nanomolar range, it should also be possible to select RNA ligands from the pool of ncRNAs that an organism can possibly express even without isolating their in vivo transcripts. This approach, termed *genomic SELEX* (Singer *et al.* 1997), is based on the in vitro generation of RNA species that are derived from a library of an organism's entire genomic DNA.

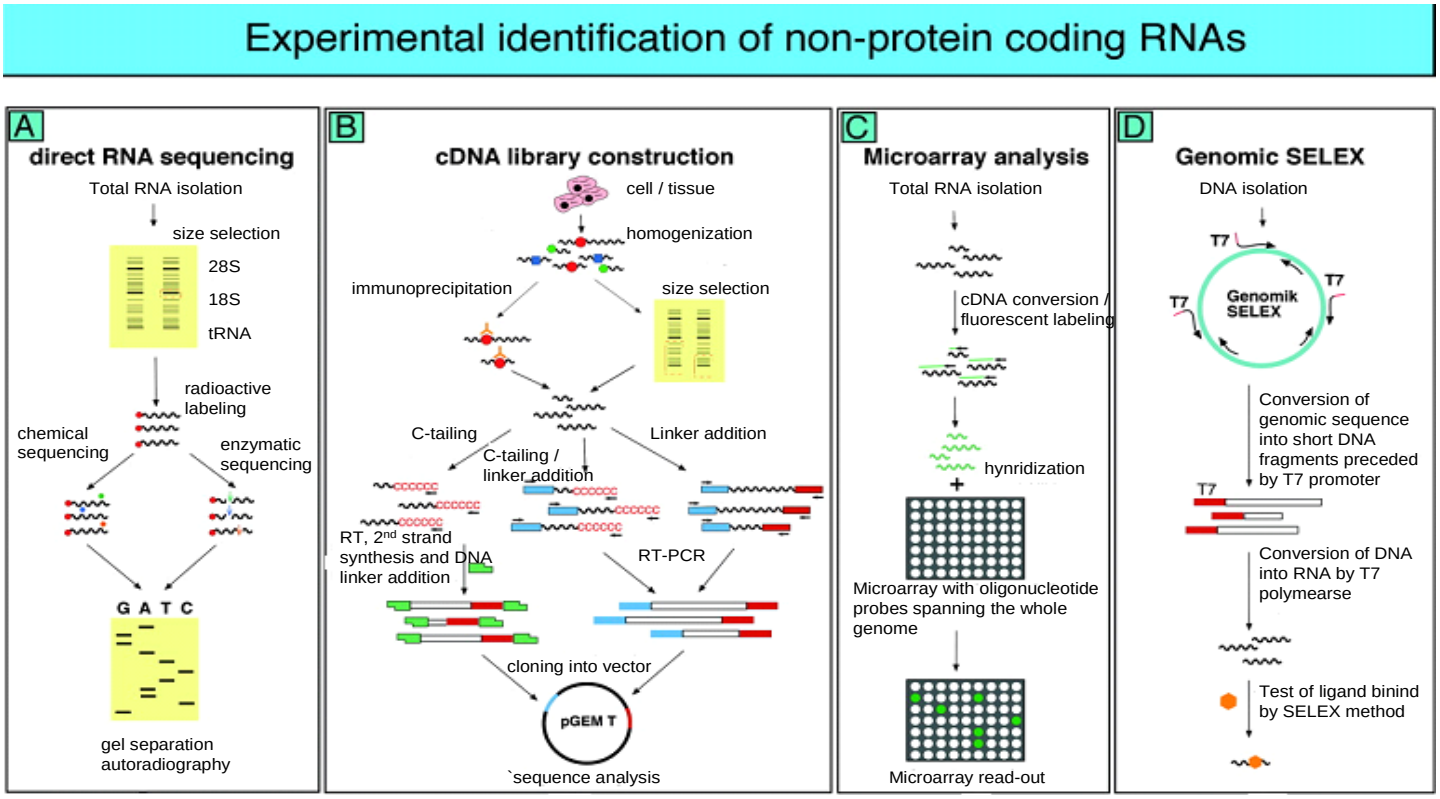


Figure 1-10. Four experimental approaches (A–D) to identify candidates for ncRNAs. (A) Identification of ncRNAs by chemical or enzymatic sequencing of extracted abundant RNAs. (B) Identification of ncRNAs by cDNA cloning and sequencing; three different methods are indicated to reverse transcribe ncRNAs, usually lacking poly(A) tails, into cDNAs (e.g. by C-tailing, C-tailing and linker addition, or linker addition, only, followed by RT–RCR). (C) Identification of ncRNAs by micro-array analysis.DNAoligonucleotide covering the sequence space of an entire genome are spotted onto glass slides, to which fluorescently labelled samples, derived from cellular RNA, is hybridized. (D) Identification of ncRNAs by genomic SELEX. By random priming, the sequence of a genome is converted into short PCR fragments containing a T7 promotor at their 50 ends. Subsequently, in vitro transcription by means of T7 RNA polymerase converts this genomic sequence of an organism into RNA fragments, which can then be assayed for function, such as binding to a specific protein or small chemical ligand, by SELEX. (adapted from (Huttenhofer *et al.* 2002)).

The generated RNA pool will undergo successive rounds of association with a given RNA-binding protein, partitioning and re-amplification. As a result, RNA sequences that are stringently bound by the protein partner will be enriched. Once the sequence of the bound RNAs is determined, this information can be used to search for matches in the genome, and so predicted genomic regions could then be tested for the expression of unknown ncRNAs. Genomic SELEX would clearly have its strength in finding ncRNAs that are overlooked by methods that require an ncRNA gene to be expressed at a certain level. With their small genome sizes, prokaryotes should be particularly amenable to this type of approach. As a further advantage of genomic SELEX, the tight association of an ncRNA with a given protein that is a prerequisite for its successful selection could also point to a biological role of this ncRNA, e.g. its function as an antagonist or cofactor of the protein's activity. Genomic SELEX generates RNA species from all regions of a genome und thus is not dependent on isolating RNAs from all these different states.

On top, fishing for the ncRNA of interest through an RNA-binding protein, the target RNA, complementary to the ncRNA, could be co-isolated as well.

The methods presented above offer a rich tool-box to search for and identify ncRNAs at both large and small scale in virtually any genome and are fruitfully employed in functional genomics. Microarray expression studies as well as the analysis of libraries of mRNAs and expressed sequence tags (EST) have allowed the detection of antisense transcripts and of transcripts outside the known transcription units, some of which are likely to be ncRNAs (Lavorgna *et al.* 2004). Despite the many screens that have been carried out in recent years, the total number of ncRNAs is not known for any organism. The number of predicted ncRNAs for any given organism varies widely, ranging from a few 10s to 100s in bacteria and from several 100s to 1000s in mammalian cells.

III. High-density tiling arrays

DNA microarrays have become ubiquitous and the most convenient platform in genomic research serving as the tool for the large-scale analysis of gene expression (Bertone *et al.* 2006). The design of DNA microarrays has been tailored to the measurement of mRNA transcript levels from annotated genes, represented either by PCR products comprising entire cDNA sequences (Schena *et al.* 1995), or by short oligonucleotides complementary to internal regions of spliced messages (Lipshutz *et al.* 1999). Microarrays of this design provided a genome-wide portrait of transcriptional activity in a given physiological context and allowed the simultaneous interrogation of thousands of nucleotide sequences.

High-density oligonucleotide microarrays have become a powerful and versatile tool, allowing parallel hybridization assays to be carried out at a miniaturized scale (Hu *et al.* 2001; Kapranov *et al.* 2003; Lockhart and Winzeler 2000; Saha *et al.* 2002; Steinmetz and Davis 2004). They have augmented bioinformatics approaches like *ab initio* gene predictions, homology studies, analysis of motifs, which prevailed in systematic gene prediction and annotation during the last 10 years (Griffin *et al.* 2002; Jansen and Gerstein 2000; Kellis *et al.* 2003; McCutcheon and Eddy 2003; Winzeler *et al.* 1998; Wu *et al.* 2002).

For the high-density oligonucleotide arrays oligos are synthesized directly on the surface of glass wafers using the principles of solid-phase nucleic acid synthesis chemistry and photolithography. The photolithographic resolution of the mask determines the number of oligonucleotide synthesis areas (features). Commercially available arrays measure 1.28x1.28 cm^2 containing ~10^7 copies of each selected probes in features ranging from 11 to 24 micron slide length. The major applications of

microarrays could be broadly categorized into RNA expression profiling, genotyping and resequencing. First arrays used for expression profiling exploited a so-called "surrogate" annotation-based strategy, which places interrogating probes based on genomic annotations and considers the behavior of the probes as a surrogate for the behavior of the entire gene (Kapranov *et al.* 2003). Expression profiling with microarrays offered high efficiency in obtaining global and quantitative information about the current functional state of genes in a cell or an organism (Young 2000), however this method could not fully unravel the complexity of the transcriptome and its regulatory circuits because it was mostly based on cDNAs or ESTs and relies on prior genome annotation (DeRisi *et al.* 1997; Oliver 2002; Shoemaker *et al.* 2001). For the same reason and due to technical limitations, identification of genes encoded by small ORFs has been limited (Kessler *et al.* 2003).

To alleviate these shortcomings a method with much higher resolution and sensitivity was needed, which should employ an unbiased strategy of interrogation, placing probes either at every base or at some fixed separated distance over the entire genome, independent of annotations. Hence, "tiling arrays" were developed. "Tiling" microarrays represent a complete non-repetitive tile path over a chromosome or locus, irrespective of any genes that may be annotated in that region (Fig. 1-11). This unbiased representation of genomic DNA has enabled the discovery of many novel transcribed sequences (Bertone *et al.* 2004; Kampa *et al.* 2004; Kapranov *et al.* 2002; Rinn *et al.* 2003; Yamada *et al.* 2003), as well as the global identification of transcription factor binding sites (Cawley *et al.* 2004; Euskirchen 2004; Horak *et al.* 2002; Horak and Snyder 2002a; b; Iyer *et al.* 2001; Martone *et al.* 2003; Ren *et al.* 2000).

There are a lot of important factors to consider when designing the tiling array. Firstly, with the increase of an organisms' complexity, the proportion of repetitive DNA sequence elements in the genome increases as well. Accommodating repetitive genomic sequences on the genome would contribute to cross-hybridization and superfluous representation of certain cDNA sequences. Secondly, the placement and subdivision of the remaining non-repetitive DNA is an issue. The serial placement of oligonucleotides along segments of non-repetitive genomic DNA can either be contiguous, covering all of the available sequence, or discontiguous, where gaps of a predetermined size range are allowed between adjacent probes. This depends on the type of experiment for which the microarray is intended, and what kind of biological information the array is capable of measuring given a particular experimental sample. A third factor concerns thermodynamic properties of oligonucleotide probes based on their predicted hybridization affinities (SantaLucia 1998).

Refinement of the tiling array technology together with completion of genome sequencing projects taking place for the last 5 years has challenged contemporary researchers with an unprecedented

endeavor of having almost complete genomic and transcriptome datasets for a given conditions in an organism of interest. The new complexity requires new comprehensive post-genomic strategies: advanced studies in regulatory mechanisms and application of new bioinformatic methods in an integrative biology perspective. This can be accomplished in studies with model organisms under controlled conditions, sometimes calling for the use of "surrogate" model species. On of these excellent models is yeast Saccharomyces cerevisiae which has long been at the vanguard of genomic research (Oliver 2002).

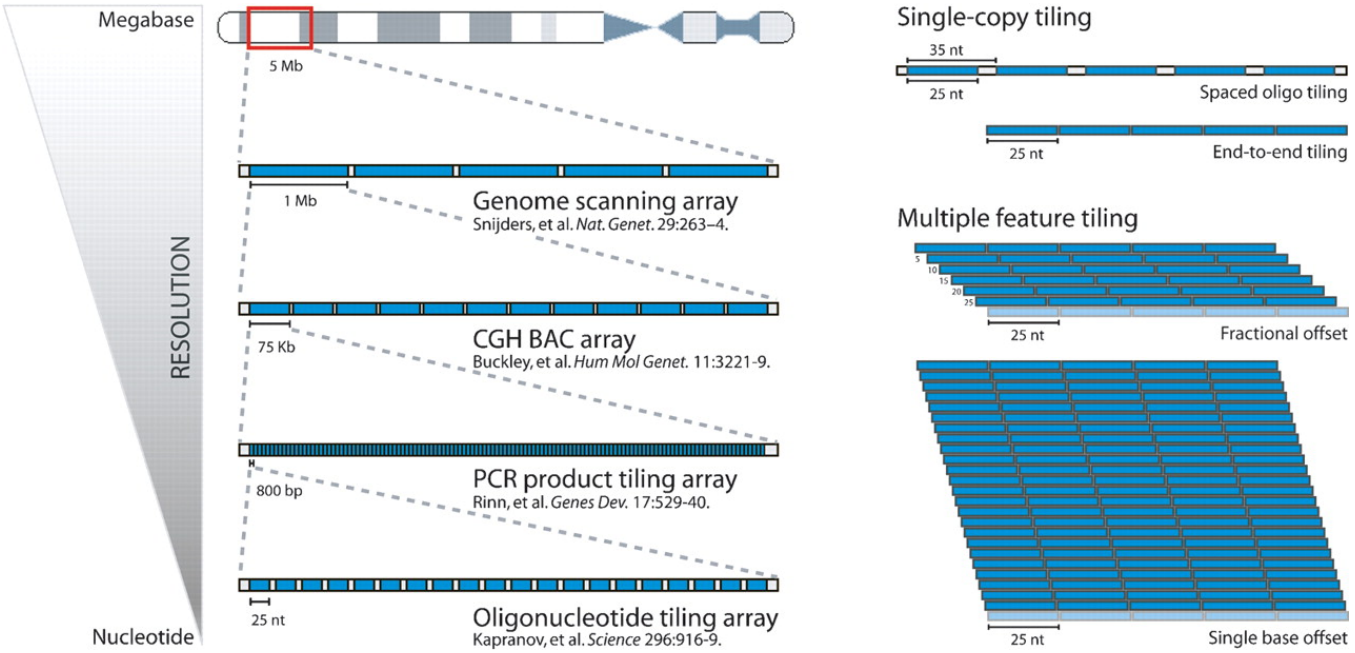


Figure 1-11. Evolution of genomic tiling arrays. (*Left*) Representing large spans of genomic DNA with bacterial artificial chromosome (BAC) clones facilitates global experimentation using relatively few array features, at the expense of low-tiling resolution. Higher-resolution designs using PCR products or oligonucleotides allow precise mapping of transcripts and regulatory elements, but require labor-intensive or technologically sophisticated approaches to implement. (*Upper right*) Linear feature tiling with gapped and end-to-end oligonucleotide placement. (*Lower right*) Overlapping tiles using fractional offset (e.g., one 25-mer probe placed every 5 nt) and single-base offset placement. The latter strategy provides a finer-resolution tiling of the genomic sequence, and can give a more precise indication of where hybridizing sequences are located on the chromosome. (adapted from (Bertone *et al.* 2006)).

IV. Yeast as a model system for functional genomic studies.

The optimal characteristics of yeast as a well-defined system for comprehensive studies under controlled conditions makes it a perfect "touchstone" model to be used in integrative, 'systems biology' studies to get new insights into the mechanisms of regulation, responsible for specific phenotypes under particular environmental conditions, to be applied to more complex organisms (e.g. plants, human) (Castrillo and Oliver 2004). The relatively small and well-characterized genome of *Saccharomyces cerevisiae* (Goffeau

1996), its simple methods of cultivation under controlled conditions, well characterized genetics and facile techniques of genetic manipulation. Many basic cellular mechanisms and biochemical processes were first elucidated in yeast, and a wide knowledge of the genetics, biochemistry and physiology of this yeast is presently available (ROSE and HARRISON 1987-1995).

The knowledge of the sequence, combined with the existence of a comprehensive collection of gene deletion mutants (http://www.unifrankfurt.de/fb15/mikro/euroscarf/complete.html), and new high-throughput technologies for global analyses at a genome-wide scale are rapidly extending the range of applications. At the genomic level these include gene disruption and construction of a collection of deletion mutants for functional profiling and characterization (Giaever *et al.* 1999; Ross-Macdonald *et al.* 1999; Winzeler *et al.* 1999a), new methods of gene trans-complementation between human cells and yeasts, for the elucidation of the function of human genes (Osborn and Miller 2007; Simon and Yen 2003). At the transcriptome level, yeast microarrays are used extensively (David *et al.* 2006; Hayes *et al.* 2002; Lashkari *et al.* 1997; Spellman *et al.* 1998; Wodicka *et al.* 1997). Transcriptional responses and patterns of expression of yeast under carefully controlled conditions (e.g. chemostat culture) are progressively being studied (Boer *et al.* 2003; Hayes *et al.* 2002). At the proteome level - the first whole-proteome microarray has been developed in yeast (Michaud *et al.* 2003; Zhu *et al.* 2003) along with new strategies for the preparation of protein arrays (Washburn *et al.* 2003). The first studies on subcellular localization of proteins on a proteome-wide scale (Ghaemmaghami *et al.* 2003; Huh *et al.* 2003), phosphoproteome analysis (Ficarro *et al.* 2002) and protein-protein interaction maps (Gavin *et al.* 2002; Ho *et al.* 2002; Ito *et al.* 2001; Uetz and Hughes 2000) were reported in yeast as well. Also, a new strategy for the investigation of enzymatic activities associated to specific metabolic pathways was applied in yeast (Chen *et al.* 2003), and the role of posttranslational effects as an overlooked dimension in proteomics was scrutinized (Pratt *et al.* 2002). In parallel new methods for the analysis of yeast metabolites (Castrillo *et al.* 2003), strategies to ascribe function to unknown genes (Raamsdonk *et al.* 2001) and classification of yeast mutants using metabolic footprinting (Allen *et al.* 2003) emerged at the yeast metabolome level. Some of the best examples may be found at the metabolic control and bioinformatic level, such as the development of new machine learning methods for the analysis of transcriptome, proteome and metabolome data and for the study of regulatory networks (Fiehn and Weckwerth 2003). Many of these advanced resources were first conceived for use in yeast (Cornell *et al.* 2003; Garrels *et al.* 1997; Payne and Garrels 1997).

Yeast has been used as a model organism to study the cell cycle (discussed below) and checkpoints (Gould and Nurse 1989; Spellman *et al.* 1998; Weinert and Hartwell 1989); cell polarity (Chang and

Peter 2003); mechanisms of evolution and speciation (Delneri *et al.* 2003); ageing and extension of lifespan (Howitz *et al.* 2003); mechanisms of infection and propagation of prions (Bach *et al.* 2003; Fernandez-Bellot and Cullin 2001; Kryndushkin *et al.* 2003; Sherman and Muchowski 2003), and as a model to gain insight into the molecular pathology of neurodegenerative diseases (Outeiro and Lindquist 2003).

The integrative knowledge, obtained in yeast, can be related to information from other organisms, towards the objective of a better understanding of the cell biology of more complex systems (e.g. fly, plants, mouse, and human).

1. Non-coding RNAs in yeast

Yeast presents a good model to correlate proteome and transcriptome (Griffin *et al.* 2002). The first tiling array design was developed for yeast and resulted in an unprecedented high-resolution survey of yeast complete transcriptome during growth in rich media conditions (David *et al.* 2006). In parallel, another group set for the global delineation of yeast transcriptome by sequencing cDNA libraries from cells exponentially growing on minimal media and meiotic cells (Miura *et al.* 2006). Both libraries were generated by a vector capping method which allowed the accurate mapping of transcription start sites (TSSs). The two analyses proved that even such small and well characterized genome has transcriptional complexity far beyond current annotation, by revealing novel transcripts in intergenic regions and transcripts derived from antisense strands of the annotated ORFs. Transcripts distinct from annotated genes and those on antisense channel didn't overlap completely between these two studies as they were performed under different growth conditions.

In parallel a different set of yeast ncRNAs was identified in a global screen for non-coding RNAs with tiling arrays in the cells where an essential pathway required for tRNA processing and maturation was modulated (Samanta *et al.* 2006). And yet another group revealed a set of conserved ncRNAs bound to Lhp1p proteim – a yeast homologue of La (Inada and Guthrie 2004). Lhp1p is involved in processing of newly synthesized RNA pol III transcripts and is known to bind infrastructural non-coding RNA precursors thereby facilitating their maturation by stabilizing them from digestion.

The extent of antisense transcription made it clear that ncRNAs play essential tasks in yeast. *S.cerevisiae* lack all the vestiges of protein machinery required for siRNA or miRNA processing, which makes them uniquely suited for ascertaining ncRNA-related pathways acting by means of transcriptional interference.

To exemplify this mechanism, several cases of regulation via transcriptional interference by ncRNA were discovered in yeast to this end.

SRG1 (SER3 regulating gene 1) is an intergenic transcript upstream of SER3 gene – a phosphoglycerate dehydrogenase that catalyses a step in serine biosynthesis. SRG1 is transcribed in *cis*-from the same strand and regulated by a fully functional promoter. Its transcription across the SER3 promoter region blocks the binding of transcription factors required for SER3 activation, thus keeping SER3 repressed (Martens *et al.* 2004; Martens *et al.* 2005).

Entry into meiosis is a key developmental decision in *S. cerevisiae*. It was thought to be under the control of proteins, not RNA. In diploid cells haploid functions are repressed by the a1/α2 protein heterodimer. Entry of the diploid into meiosis requires the function of a number of genes including the **IME genes (Initiator of Meiosis)**. One of the IME genes required by MAT a/α diploids to initiate meiosis is **IME4** (Initiator of Meiosis 4), a putative RNA methyltransferase (Clancy *et al.* 2002). It has been shown that IME4 transcription and subsequent entry into meiosis is controlled by antisense transcription of the IME4 gene itself. Sense and antisense transcription is cell-type specific: haploids produce IME4 antisense RNA, whereas MAT a/α diploids produce IME4 sense RNA. a1/α2 heterodimer represses the transcription of antisense in a/α diploids to allow sense IME4 expression from a weaker sense promoter. Conversely, haploid cells should constitutively express antisense IME4 as they are devoid of a1/α2 heterodimers. These transcripts inhibit one another in *cis*, but not in *trans*. These data are consistent with transcription interference model, in which high levels of transcription from the strand with the stronger promoter effectively reduces incoming transcription from the complementary strand. The strength of the promoter driving either sense or antisense transcription from the IME4 locus is cell-type-specific. Consequently, IME4 sense transcription determines a cell-type capable of meiosis, whereas IME4 antisense transcription determines a cell-type incapable of meiosis (Hongay *et al.* 2006).
Several attempts to develop silencing approaches in *S.cerevisiae* have led to conflicting results (Bonoli *et al.* 2006). Law and Devenish pioneered the use of asRNAs in yeast by cloning a 5' region of ADE1 ORF in inverse orientation into an expression vector (Law and Devenish 1988). Despite detectable sufficient amounts of ADE1 asRNA, no silencing was observed. The expression of asRNA complementary to the 5'UTR and the whole coding sequence of MIG1 also has not led to any silencing of the sense (Olsson *et al.* 1997). On the contrary, essential yeast gene YBR136w can be conditionally suppressed by expression of its antisense construct complementary to the intragenic short region of 235bp (Nasr *et al.* 1995).

Bonoli and colleagues have undertaken a more successful silencing approach in yeast and concluded that (i) targeting the 5'UTR of a chosen mRNA by means of antisense RNA expressed *in vivo* can lead to effective gene silencing in *S.cerevisiae*, (ii) regulation of transcription of antisense constructs is crucial for the conditional silencing of a target gene and can be affected by the sequence context at the junction(s) of the expression vector with the antisense construct, (iii) it appears difficult in budding yeast to attain silencing levels above 50% (Bonoli *et al.* 2006).

Although transcriptional interference appears to be the emphasized mechanism of antisense regulation in *S.cerevisiae*, two other studies show that epigenetic and chromatin modulation mechanisms could be involved.

For **PHO5** gene it has been shown that non-coding transcription plays a role in activation, but not repression. Histone eviction from the PHO5 promoter during activation occurs with normal kinetics even in the absence of the PHO5 TATA box, showing that transcription of the gene itself is not required for promoter remodeling. Mutations that impair transcript elongation by RNAPII affect the kinetics of histone eviction from the PHO5 promoter. Most dramatically, inactivation of RNAPII itself abolishes eviction completely. Under repressing conditions, an approximately 2.4-kb non-coding exosome-degraded transcript was detected that originates near the PHO5 termination site and is transcribed in the antisense direction. Abrogation of this transcript delays chromatin remodeling and subsequent RNAPII recruitment to PHO5 upon activation. It was proposed that noncoding transcription through positioned nucleosomes can enhance chromatin plasticity so that chromatin remodeling and activation of traversed genes occur in a timely manner (Uhler *et al.* 2007).

Genome-wide studies in *S.cerevisiae* reveal that the transcriptome includes numerous antisense RNAs as well as intergenic transcripts regulated by the exosome component Rrp6. Upon the loss of Rrp6 function, two **PHO84 antisense** transcripts are stabilized, and PHO84 gene transcription is repressed (Camblong *et al.* 2007). The same phenotype is observed in wild-type cells during chronological aging. In yeast cells loss of Rrp6 function is paralleled by the recruitment of Hda1 histone deacetylase to PHO84 and neighboring genes. However, histone deacetylation is restricted to PHO84, suggesting that Hda1 activity depends on antisense RNA. The knockdown of antisense production prevents PHO84 gene repression, even in the absence of Rrp6. Hence, the stabilization of antisense transcripts results in PHO84 gene repression via a mechanism distinct from transcription interference, and the modulation of Rrp6 function contributes to gene regulation by inducing RNA-dependent epigenetic modifications.

The studies described above emphasize that yeast operate an antisense RNA regulatory system at different levels for various processes in the cell. Despite the abundance of newly identified ncRNAs

many of them remain orphan. The burning question is: what are the functions of all of these RNA transcripts? Or, if they are not functional, why does the cell devote its resources to producing them? Thus, next to novel methods to identify them in model organisms, also novel high-throughput approaches are needed to tackle the biological roles of ncRNAs and their function in different genomic and physiological contexts.

A major challenge for delineating antisense regulation in eukaryotes is to understand the integration of transcriptional and post-transcriptional regulatory mechanisms. Such insight is essential for elucidation of the principles of genetic regulation within the complex genomes. However, a well-characterized system is needed to perform this kind of study. Mitosis is such a system in yeast.

V. Cell cycle in Yeast

1. Basics of yeast cell cycle

The cell cycle entails the orchestration of virtually all cellular processes: metabolism, protein synthesis, secretion, DNA replication, organelle biogenesis, cytoskeletal dynamics and chromosome segregation (Tyers 2004). For more than two decades research groups all over the globe have been assembling the detailed puzzle of tightly controlled sequence of events in the eukaryotic cell cycle. The early scrupulous work to parse and then put together various signaling strings and the regulatory molecules therein, whose changes in activity were responsible for driving the intricate mitotic mechanism, was mainly done by biochemical studies in *Xenopus laevis* and genetic analyses in the budding yeast, which both served as excellent models to manipulate with facile for deciphering the order of key chromosome cycle transitions. Many of the critical regulatory players in the cell cycle landscape were uncovered, most notably the cyclin dependent kinases (CDKs).

Sachcharomyces cerevisiae possesses at least four CDKs: Cdc28p (Nasmyth 1993), Pho85p (Cross 1995), Kin28p (Cismowski 1995) and Srb10p (Green and Johnson 2004), of which only Cdc28p has a clear role in regulating the cell cycle; the others are all involved in regulating transcription. Cdc28 (or Cdk1) performs the tasks carried out by cdc2 in *Schizosaccharomyces pombe* and those that are shared between Cdk4, Cdk2 and Cdk1 (Cdc2) in mammalian cells. Its functions are performed by varieties of the kinase that differ mainly, if not solely, in the cyclin subunit associated with it, which also endures transition from one stage of the cycle to the next (Fig. 1-12 a, b).

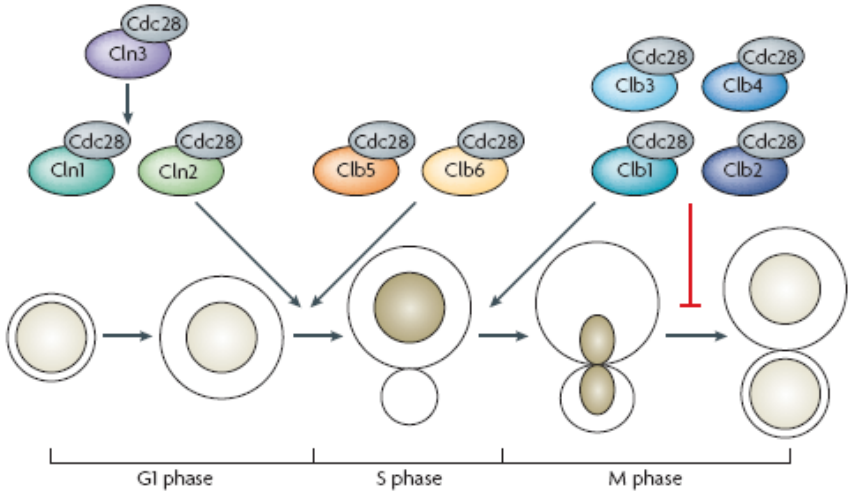
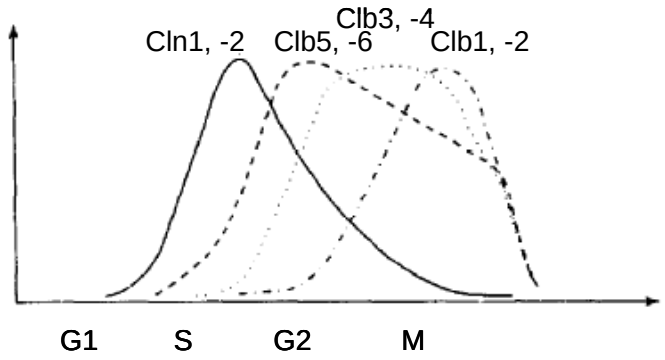


Figure 1-12. (a) Cyclins in the budding yeast cell cycle. Budding yeast cyclins activate a single cyclin-dependent kinase (Cdc28). The G1-phase cyclins (Cln1, Cln2 and Cln3) promote bud emergence, spindle pole body duplication (not shown) and activation of the B-type cyclins. The S-phase cyclins (Clb5, Clb6) advance DNA replication (shaded nucleus), and the M-phase cyclins (Clb1, Clb2, Clb3 and Clb4) promote spindle formation and the initiation of mitosis. Mitotic cyclins inhibit mitotic exit and cell division. Following cytokinesis, a mother and daughter cell are generated. (adapted from (Bloom and Cross 2007a)).

(b) Activation of different cuclins during progression through cell cycle. (adapted from (Nasmyth 1996)).



The main paradigm of the regulation of mitosis was that different CDK subtypes catalyze different cell cycle transitions; however, it is not only the state of activity of Cdc28- associated to a particular cyclin subunit that determines the position in the cell cycle, rather, a complex network consisting of (i) cyclin-dependent kinases (CDKs) and their associated cyclins, (ii) kinases and phosphatases that regulate CDK activity, and (iii) stoichiometric inhibitors that sequester cyclin-CDK dimers controls the progress through the division cycle (Novak *et al.* 1998). Stage-specific transcriptional programs in mitosis are coupled to each other through an intricate interplay of transcription factors, CDK-cyclins activity and ubiquitin-dependent proteolysis of stage-specific cyclins (reviewed in (Tyers and Jorgensen 2000)).

Three G1 cyclins (Cln1-3) activate Cdc28 in G1 phase and Cln1- and 2-Cdc28 complexes initiate entry into mitosis. Cln3p-Cdc28 activates transcription of G1/S transcription factors Swi4, Swi6 and Mbp1 in late G1; Cln1 and Cln2 turn off proteolysis of B-type cyclins (Clb1-Clb6), responsible for triggering later mitotic stages, turn on proteolysis of a cyclin-B-Cdk1-specific inhibitor Sic1 (Dirick *et al.* 1995; Schwob *et al.* 1994), turn off the ability of haploid cells to respond to mating pheromones, and trigger the

polarization of the cytoskeleton needed for bud formation. Clb5 and Clb6 trigger DNA replication, Clb3 and Clb4 trigger the formation and elongation of mitotic spindles, and Clb1-Cdc28 and Clb2-Cdc28 activate the expression of Cdc20 (an activator of anaphase-promoting complex/cyclosome (APC/C), which is required for metaphase/anaphase transition, and directs ubiquitination of mitotic cyclins), Swi5 (a transcription factor that activates genes expressed at the M/G1 boundary and in G1 phase), and Ace2 (a transcription factor that activates expression of early G1-specific genes, localizes to daughter cell nuclei after cytokinesis and delays G1 progression in daughters) and trigger nuclear division and isometric bud growth during G2 (Lew *et al.* 1993; Lew and Reed 1993). Different cyclins determine either the location of Cdc28 or its association with other proteins and/or its substrate specificity (Nasmyth 1996). Multiple cyclins are advantageous because they allow for flexible control of the cell cycle. On top different cyclins are independently regulated transcriptionally and post-transciptionally providing regulatory flexibility at the level of input. Precisely timed proteolytic degradation of cyclins and other cell-cycle regulators by the ubiquitin system turns on CDK activity in G1, and turns it off at the end of mitosis. As part of this regulatory circuit, the cyclins themselves form crucial elements of the G1/S and G2/M transitional programs, often referred to as CLN2 and CLB2 clusters, respectively.

To understand the basic logic of the cell cycle groups of Novak, Tyson and Cross and, independently, Kim Nasmyth have envisioned that the cell cycle in budding yeast is an alternation between two self-maintaining stable steady states – Start and Finish (G1 and S/G2/M, respectively). The Start transition carries a cell from G1 to S/G2/M, and the Finish transition from M back to G1 (Nasmyth 1996; Tyson and Novak 2001; Tyson *et al.* 1995). The two self-maintaining steady states arise primarily from the mutual antagonism between B-type cyclins (Clb1-6, in association with Cdc28) and the G1 stabilizers (Cdh1 – an activator of the anaphase-promoting complex/cyclosome (APC/C), which directs ubiquitination of cyclins resulting in mitotic exit, Sic1 and Cdc6). Cdh1/APC degrades the Clbs, whereas Sic1 and Cdc6, referred to together as the CDK inhibitors (CKIs), stochiometrically inhibit Cdc28/Clb complexes. Clb-kinases, on the other hand, can inactivate Cdh1 and destabilize CKIs. Since Clb-kinases and the G1 stabilizers mutually inhibit each other, these two classes of proteins cannot coexist. In the G1 state, Clb-kinase activities are low because Clb synthesis processes are turned OFF, their degradation by APC/Cdh1 is ON, and their inhibitors, the CKIs, are abundant. The reverse is true in the S/G2/M phase (Fig. 1-13).

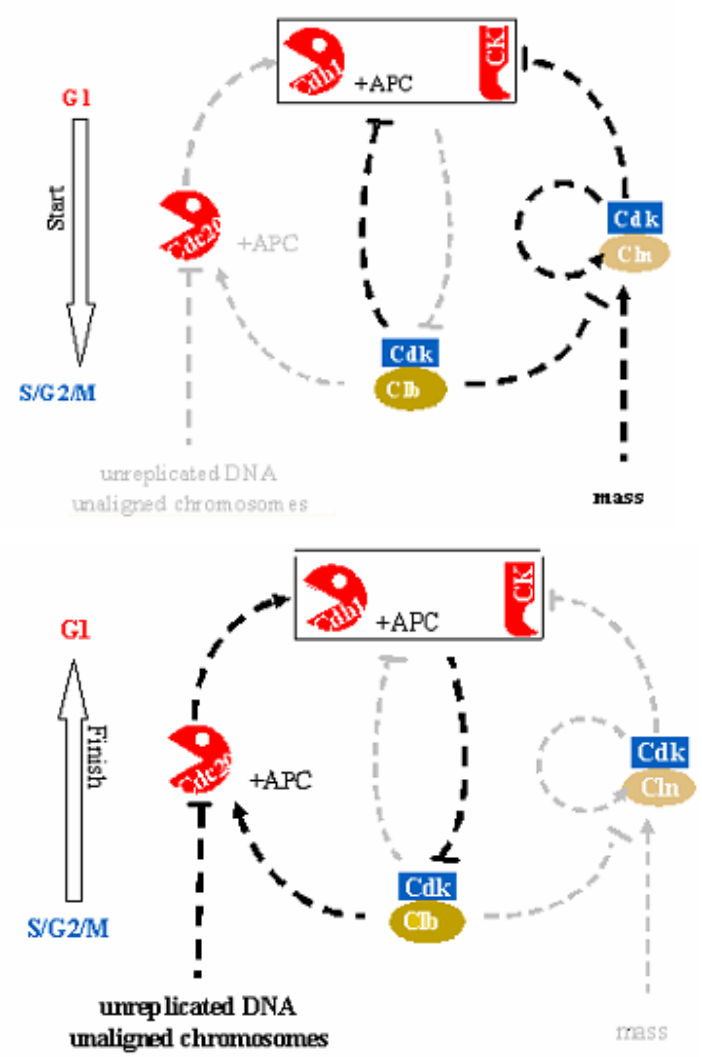


Figure 1-13. Cell cycle transitions.
(a) The Start transition is facilitated by Cln-kinases (Cln1-3/Cdc28 complexes) that can phosphorylate and inactivate CKI and Cdh1, but are not themselves opposed by CKI and Cdh1. This transition is driven by cell growth. When the small daughter cell has grown to a critical size and Cln-kinase activities have reached a critical level, CKI and Cdh1 are inactivated, Clb-kinase activities increase, a bud emerges, DNA replication commences and spindle pole is duplicated. (The mother cell executes Start soon after birth because it has already attained the critical size.) The rising activity of Clb-kinases turns off Cln synthesis, causing Cln-kinase activities to drop in preparation for the Finish transition.

(b) The Finish transition is facilitated by Cdc20, which is activated indirectly by Clb-kinases. When the spindle assembly checkpoint is lifted, Cdc20 is activated, sister chromatids are separated, and Clbs are partially degraded. Cdc20 also activates the Cdc14 phosphatase, which reverses the inhibitory effects of Clb-kinases on Cdh1 and CKIs, allowing the latter two to overpower the Clb-kinases and extinguish their activities. As Clb-kinase activities drop after Finish, Cdc20 activity also disappears, preparing the cell for the subsequent Start transition

Passage through Start of the cell cycle is coupled to growth and has several requirements to fulfill, namely growth to a critical size, nutrient sufficiency, attainment of a critical translation rate and for haploids, absence of mating pheromone (Jorgensen *et al.* 2004). The critical size threshold maintains uniform cell size through many generations and under minimal supply of nutrients forces cells to accumulate enough stores of energy to complete mitosis. The Start transition is facilitated by Cln-kinases (Cln1-3/Cdc28 complexes) that can phosphorylate and inactivate CKI and Cdh1, but are not themselves opposed by CKI and Cdh1. When the small daughter cell has grown to a critical size and Cln-kinase activities have reached a critical level, CKI and Cdh1 are inactivated, Clb-kinase activities increase, a bud emerges, DNA replication commences and spindle pole is duplicated. The rising activity of Clb-kinases turns off Cln synthesis, causing Cln-kinase activities to drop in preparation for the Finish transition.

Late G1 expression depends on SBF and MBF transcription factor complexes. SBF and MBF are composed of two structurally related DNA binding proteins Swi4 and Mbp1, which form a complex with

a common transcription regulatory subunit Swi6 to drive the expression of a suit of ~200 genes. These proteins are bound to late G1 promoters throughout G1 in an inactive state due to the presence of the inhibitor Whi5. Cln3-Cdc28 complex, active throughout the whole cell cycle, phosphorylates Whi5, which facilitates its dissociation from promoters and induces its nuclear export and activates transcription of late G1 genes, including *CLN1* and *CLN2*.

Among the late G1 transcripts there is a repressor called Yox1, which forms a negative feedback loop by repressing the transcription of Cln3 and Swi4 as well as some other transcripts which peak at the M/G1 border (Fig. 1-14). An additional level of cyclin-specific regulation of SBF and MBF is provided by Clb6–Cdc28, which phosphorylates Swi6 to promote its nuclear export.

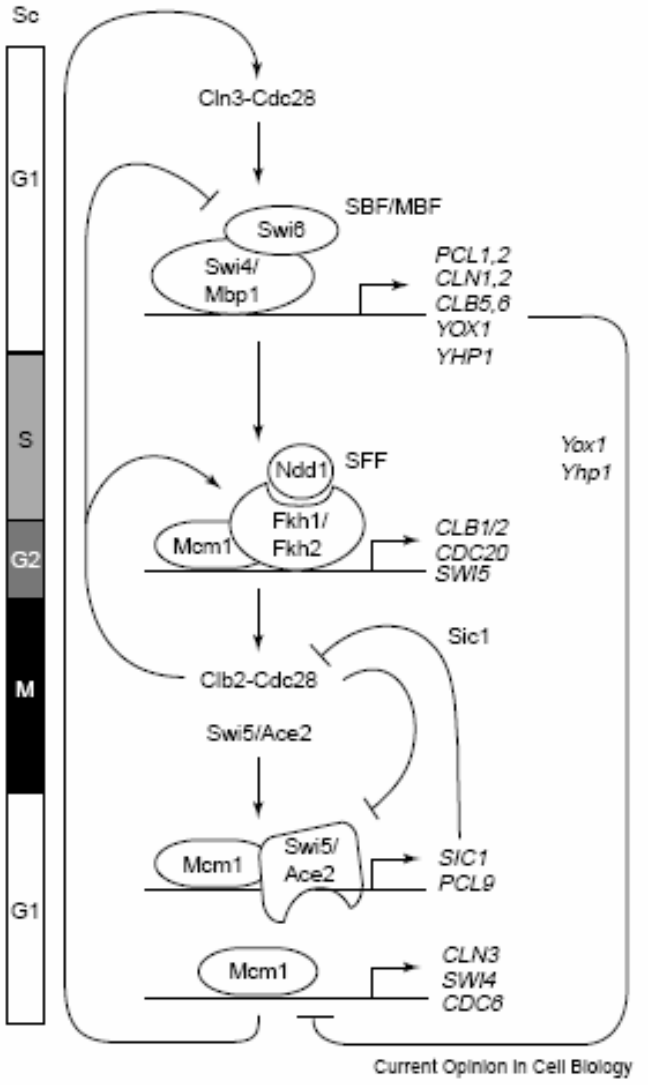


Figure 1-14. Expression of transcription factors, which determine the flow of events in mitotic pahse progression. In the closed transcriptional circuit of budding yeast transcription actors that drive one cluster are expressed as part of the previous cluster. (adapted from (Tyers 2004))

The Finish transition is facilitated by Cdc20, which is activated indirectly by Clb-kinases. When DNA synthesis is complete and chromosomes are aligned on the metaphase plate, Cdc20 is activated, sister chromatids are separated, and Clbs are partially degraded. Cdc20 also initiates the activation of the phosphatase Cdc14, which reverses the inhibitory effects of Clb-kinases on Cdh1 and CKIs, allowing the latter two to overpower the Clb-kinases and extinguish their activities. As Clb-kinase activities drop after Finish, Cdc20 activity also disappears, preparing the cell for the subsequent Start transition. Later in the cell cycle, one of the Cln3-Cdc28 activated late G1 genes - Clb2 can specifically inactivate SBF-mediated gene expression. Clb2 also positively regulates its own expression by a feed forward loop (Bloom and Cross 2007a; b). The transcription factor Mcm1 recruits the forkhead

transcription factor Fkh2 and the co-activator Ndd1 to regulate the expression of CLB1 and CLB2 and the next wave of transciption factors – Swi5 and Ace2. Clb2–Cdc28 phosphorylates Ndd1, which is important for its recruitment to the *CLB2* gene promoter, and phosphorylates Fkh2, which enhances the interaction of Fkh2 with Ndd1 (Fig. 1-14, 1-15b). Transcriptional control of *CLB2* is highly specific because Clb2 is probably specialized in Ndd1 phosphorylation and because *CLB2* is also a specific cyclin target of this circuit. Swi5 and Ace2 in turn activate genes involved in cell separation and in distinguishing mother from daughter cells. At about the same time Mcm1 is released from repression by two repressors Yox1 (activated in late G1) and Yhp1 at another set of M/G1-specific promoters, which induce the transcription of genes required to set up pre-replication comlexes for DNA-synthesis (Mcm2-7 and cdc6) and to restart the cell cycle - Cln3 and Swi4.

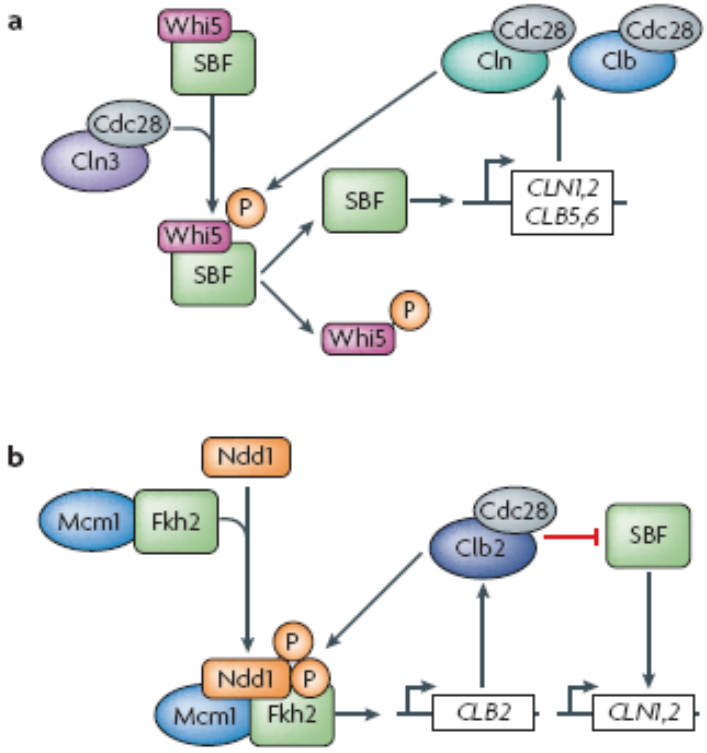


Figure 1-15. Transcriptional regulation of cyclins.
(a) The transcription repressor Whi5 inhibits the activity of the SBF transcription factor. Phosphorylation of Whi5 by Cln3–Cdc28 induces the nuclear export of Whi5 and activates SBF, which induces the transcription of the genes that encode Cln1, Cln2, Clb5 and Clb6. Cln–Cdc28 and Clb–Cdc28 phosphorylate Whi5, which might provide a positive-feedback loop.
(b) The co-activator Ndd1 recruits minichromosome maintenance-1 (Mcm1) and the transcription factor Fkh2 for the activation of the gene that encodes Clb2. Phosphorylation of Ndd1 and Fkh2 by Clb2–Cdc28 promotes Ndd1-dependent recruitment of Mcm1–Fkh2 to the promoter of CLB2. Clb2–Cdc28 also phosphorylates and inhibits SBF to repress the transcription of G1-phase cyclins. P, phosphate. (adapted (Bloom and Cross 2007a))

The completion of mitotic exit requires the release and activation of the Cdc14 protein-phosphatase, which is kept inactive in the nucleolus during most of the cell cycle. Activation of Cdc14 is controlled by two regulatory networks called FEAR (Cdc Fourteen Early Anaphase Release) and MEN (Mitotic Exit Network). It has been shown recently that the anaphase promoting protease (separase) is essential for Cdc14 activation, thereby it makes mitotic exit dependent on execution of anaphase (Toth *et al.* 2007).

The activation of Cdc20, which causes the activation of separase and a decrease of Cdc28 kinase activity provides an initial trigger for the activation of the MEN-Cdc14 positive feedback loop, which in turn, flips the second irreversible Cdk–APC^{Cdh1} switch on the APC^{Cdh1} side. MEN is initiated and controlled by Tem1, a small GTPase. Tem1 is regulated by a putative guanine exchange factor, Lte1, but the function and regulation of Lte1 remains poorly understood.

2. Levels of the cell cycle control

In the course of mitotic cell cycle the genome is replicated and identical copies of it are passed to two daughter cells. To maintain ploidy from one generation to the next, unidirectional progression through the phases of the cell cycle (G1 → S → G2 → M → G1) is essential. To keep cells from regressing, the transitions of the mitotic cell cycle are irreversible processes. Although the transitions are triggered by transient signals, the cell does not revert to an earlier state when the signal disappears (Novak *et al.* 2007). Several mechanisms and levels of control ensure unidirectional irreversible transition through the cell cycle.

2.1 Regulation of cyclins

2.1.1 Transcription peaks

Regulation of specificity of cyclins' action occurs at several distinct points, which include transcriptional regulation, degradation of cyclins, direct inhibition of cylin-cdc28 complexes by stage-specific inhibitors and intracellular localization.

As described above, a crucial mechanism for cyclin specificity is the differential regulation of G- and B-type cyclins at the level of transcription during the cell cycle. Transcription of the *CLN3* gene is detectable throughout the cell cycle, but peaks in late M–early G1, whereas the transcription of the *CLN1* and *CLN2* genes peaks during G1–S. Transcription of the *CLB5* and *CLB6* genes also peak at G1–S, followed by the transcription of the *CLB3* and *CLB4* genes and then the transcription of the *CLB1* and *CLB2* genes. *CLN1* and *CLN2* expression has been primarily attributed to SBF transcription factor, and the transcriptional activation of the *CLB5* and *CLB6* genes is primarily attributed to MBF. Cln3–Cdc28 activity is important for SBF-mediated gene activation, and this might reflect some intrinsic specialization of Cln3 relative to another G1 cyclin, Cln2 (Levine *et al.* 1996).

The phosphorylation of Whi5 by Cln3–Cdc28 early in the cell cycle probably reflects, at least in part, the fact that Cln3 is the only cyclin that is expressed at this time. Cln1-, Cln2- and Clb5-directed Cdc28 activity can also phosphorylate Whi5, which has the potential to provide a positive-feedback loop for their expression (Fig. 1-15a).

2.1.2 Differential degradation of stage-specific cyclins.

Cyclins are degraded by ubiquitin-mediated proteolysis. The sensitivity of various cyclins to different ubiquitin ligases constitutes an important mechanism for cyclin specificity in controlling the cell-cycle engine. The ubiquitylation and degradation of Cln1 and Cln2 is mediated by an SCF complex (a multisubunit ligase, consisting of Skp1, a member of the cullin family (Cdc53) and a RING-finger containing protein (Roc1=Rbx1) that contains the F-box protein Grr1 (SCFGrr1), which recognize specific substrates confined to degradation and deliver them to the E2 ubiquitin-conjugated enzyme. Clb6 is the only B-type cyclin that has been shown to be degraded in an SCF-dependent manner in yeast by an SCF complex that contains the Cdc4 F-box protein (SCF-Cdc4). Its degradation happens prior to that of Clb5.

The other B-type cyclins are ubiquitylated by the APC (Peters 2006). During metaphase, APC is bound to Cdc20 and targets Clb5 and mitotic B-type cyclins for degradation. After its anaphase-promoting functions have been carried out, Cdc20 itself is ubiquitylated and targeted for degradation by the APC/C, which limits its functional window mainly to the duration of anaphase. Later in mitosis, APC, bound to the adaptor protein Cdh1, completes the degradation of mitotic B-type cyclins, including the main mitotic cyclin, Clb2. Cdh1, the alternative cofactor, is expressed constitutively, but it is negatively regulated by phosphorylation — specifically by Cln–Cdk1 and Clb5–Cdk1 complexes (Burton and Solomon 2001). So APC/Cdh1 takes over at the end of mitosis and persists during G1 when Cdk activities are low.

APC/Cdc20 and APC/Cdh1 are differentially regulated by Cdc28-mediated phosphorylation. Several subunits of the APC/C are phosphorylated by Clb–Cdc28 complexes *in vivo* and can be phosphorylated by Clb2–Cdc28 specifically *in vitro*, which activates APC/Cdc20 complex. By contrast, Cln–Cdc28 and Clb–Cdc28 complexes phosphorylate Cdh1 to inactivate APC/Cdh1 (Bloom and Cross 2007a). Differential degradation of cyclins by APC/Cdc20 and APC/Cdh1 has clear functional consequences. For example, the immunity of Clb5 to Cdh1-mediated proteolysis might allow Clb5–Cdc28 to inactivate Cdh1 at G1–S, thereby allowing the subsequent accumulation of Clb2 (Yeong *et al.* 2001).

2.1.3 Inhibition of cyclin–Cdk complexes.

Specific cyclin– Cdk complexes are negatively regulated by binding to stoichiometric inhibitors. The best-characterized Cdk inhibitor that functions in the context of sharpening a cell-cycle phase transition is the Sic1 protein of *Saccharomyces cerevisiae* – a potent inhibitor of B-type cyclins (Schwob *et al.* 1994). Synthesized at the end of the cell cycle, at which point its Cdk inhibitory activity initially potentiates mitotic exit, Sic1 is stable until the G1–S transition. The primary function of Sic1 is the inhibition of the Clb5–Cdc28 complex, which is responsible for initiation of DNA replication. Clb5–Cdk1 complexes accumulate progressively during the latter part of G1 but are maintained in an inactive pool to prevent the premature initiation of DNA replication. At the G1–S boundary, the entire pool of Sic1 is subjected to concerted and complete ubiquitin-mediated proteolysis. Cln2–Cdc28 and Clb–Cdc28 can phosphorylate Sic1, which allows Sic1 to be recognized by the SCF-Cdc4 ubiquitin ligase (Fig. 1-16, 1-17). This confers simultaneous activation to the entire resident pool of Clb5–Cdk1 and presumably also robust and decisive initiation of DNA replication (Reed 2003).

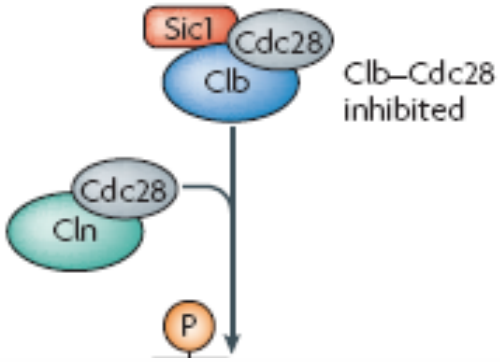


Figure 1-16. Sic1 inhibits the activity of Clb–Cdc28 complexes. Cln–Cdc28 phosphorylates Sic1, which promotes SCFCdc4-mediated ubiquitylation and subsequent degradation of Sic1, allowing for Clb–Cdc28 activation and S-phase entry. Clb–Cdc28 complexes also phosphorylate Sic1 to induce its proteolysis. P, phosphate. (adapted from (Bloom and Cross 2007a))

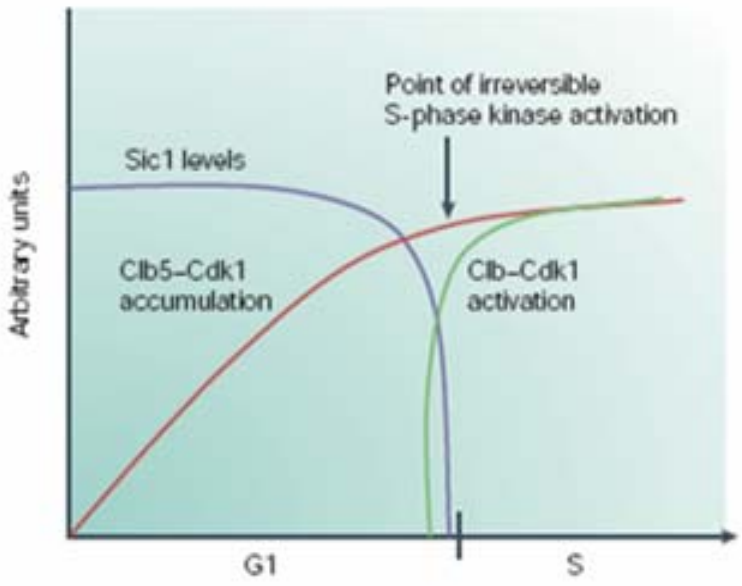


Figure 1-17. Creating an irreversible cell-cycle transition.

The gradual kinetics of Clb5–Cdk1 accumulation are shown alongside the rapid kinetics of Clb5–Cdk1 activation in the budding yeast, *Saccharomyces cerevisiae*. The ability to accumulate inactive Clb5–Cdk1 molecules is mediated by the Cdk inhibitor Sic1. The concerted destruction of Sic1 — by G1-cyclin–Cdk-complex-mediated phosphorylation and Cdc4-mediated ubiquitylation — and concomitant activation of Clb5–Cdk1, promotes an irreversible transition to S phase. G1, Gap phase 1; S, DNA synthesis.
(adapted from Reed 2003)

In contrast to Sic1 inactivation of Cln2-Cdc28, Swe1 has been shown to phosphorylate and inactivate Clb2–Cdc28, but not Cln2–Cdc28 (Booher *et al.* 1993). Genetic evidence indicates that Swe1 inhibits different Clb–Cdc28 complexes to varying degrees, with Clb2–Cdc28 being strongly inhibited, Clb3 and Clb4–Cdc28 being moderately inhibited, and Clb5 and Clb6–Cdc28 being unaffected when these cyclins are expressed during S phase. Cdc28 reciprocally regulates Swe1. Mitotic Clb–Cdc28 complexes phosphorylate Swe1, priming it for degradation by the APC. In addition, Cdc28 activity has recently been shown to both positively and negatively regulate the association of Swe1 with Cdc28. Clb2–Cdc28 phosphorylates Swe1 *in vivo*, and removal of the Cdc28 phosphorylation sites in Swe1 (Swe1–18A) causes premature entry into mitosis (Harvey *et al.* 2005). When the tyrosine phosphatase Mih1 removes the inhibitory phosphorylation on Cdc28, Swe1 becomes hyperphosphorylated and dissociates from the Clb2–Cdc28 complex to allow for the full activation of Clb2–Cdc28.

2.1.4 Cyclins localization

The localization of individual cyclins to different subcellular compartments provides a means for cyclin-specific targeting. The G1 cyclin Cln3 is primarily nuclear, whereas Cln2 is primarily cytoplasmic and can localize to sites of polarized growth. This localization pattern contributes to the abilities of Cln2 and Cln3 to regulate different substrates. The localization of Cln2 is regulated by Cdc28- mediated phosphorylation. When the Cdc28 consensus phosphorylation sites in Cln2 are mutated, Cln2 is exclusively nuclear, which indicates that phosphorylation can conceal an NLS or expose a nuclear export signal.

Clb1–4 are localized primarily to the nucleus, with portions associated with the mitotic spindle and spindle pole bodies (SPBs) (Bailly *et al.* 2003). Clb2 is the only mitotic cyclin that is also distributed to the bud neck. Localization of Clb2 to the bud neck is independent of its binding to Cdc28, but is dependent on its hydrophobic patch. Deletion of the gene that encodes the bud-neck protein Bud3

eliminates bud-neck-localized Clb2 and delays cytokinesis, which indicates that Bud3 is important for targeting Clb2 to the bud neck, and this might positively regulate cell division (Bailly *et al.* 2003).

2.2 Cell cycle checkpoints

The events of DNA replication, chromosome segregation, and mitosis define a fundamental periodicity in the eukaryotic cell cycle. Precise coordination of the unidirectional transitions between these stages is critical to cell integrity and survival. Loss of appropriate cell cycle regulation leads to genomic instability and is believed to play a role in the etiology of both hereditary and spontaneous cancers (Cho *et al.* 1998).

Checkpoints are an additional system to ensure smooth error-free passage through the cell cycle. Checkpoints block mitosis until structural requirements are met. They include morphogenesis checkpoint, the DNA-damage response, the Mad2-dependent spindle-integrity-sensing pathway and the Bud2-dependent nuclear-position-sensing pathway.

2.2.1 Morphogenesis checkpoint

A morphogenesis checkpoint in budding yeast delays cell cycle progression in response to perturbations of cell polarity that prevent bud formation (Lew and Reed 1995a; b). Environmental stresses, such as increases in temperature or osmolarity, cause a temporary disruption of cytoskeletal polarity and delay bud formation (Lillie and Brown 1994). During this delay, cell cycle progression is halted, preventing the accumulation of binucleate cells. The cell cycle delay depends upon the tyrosine kinase Swe1p, which phosphorylates and inhibits the cyclin-dependent kinase Cdc28p (Sia *et al.* 1996). This lowers the activity of G2 cyclin-Cdc28p complexes, preventing nuclear division. It was shown that a Swe1p-dependent cell cycle delay was triggered by direct perturbations of the actin cytoskeleton, even when polarity establishment functions remained intact and moreover, in cells that had already formed a bud (McMillan *et al.* 1998). This suggested that the checkpoint directly monitors actin organization, rather than (or in addition to) polarity establishment or bud formation. However, the ability to respond to such perturbations by delaying cell cycle progression was restricted to a narrow window of the cell cycle, delimited by the periodic accumulation of the checkpoint effector, Swe1p. An alternative hypothesis is

that instead of monitoring completion of a cell cycle event, the checkpoint continuously monitors the status of the actin cytoskeleton throughout mitotic progression.

2.2.2 DNA damage and Spindle-assembly checkpoints

The DNA and the spindle assembly checkpoints are surveillance pathways in charge of the proper transmission of genetic material. They check for (i) the complete and accurate replication of nuclear DNA, (ii) the absence of DNA lesions, and (iii) the equal repartition of the sister chromatids in the daughter cells. The DNA checkpoints are activated in cases of DNA damage or replication defects. They do not seem to be sensitive to incomplete DNA replication per se but rather to pathological DNA structures resulting from stalled replication forks or DNA lesions.In *Saccharomyces cerevisiae*, their main components include the kinases Mec1, Tel1, Rad53, Dun1, and Chk1, along with adaptors (Rad9 and Mrc1) mediating their interactions (Clemenson and Marsolier-Kergoat 2006).

This spindle-assembly checkpoint checks for the accurate segregation of the chromatids by monitoring either the attachment of microtubules to kinetochores (performed by Mad and Bud) or the tension that is exerted at kinetochores upon bipolar attachment. The spindle assembly checkpoint is thus sensitive to defects altering all aspects of the spindle function.

The DNA checkpoint can be activated at any point of the cell cycle depending on the nature of DNA lesions and on the phase when DNA damage occurs, whereas the activation of the spindle assembly checkpoint is restricted to G2/mitosis (G2/M). The two checkpoints have in common the ability to block both the metaphase/anaphase transition and the exit from mitosis. The progression from metaphase to anaphase is triggered by the degradation of the securin Pds1, which depends on the anaphase promoting complex (APC) (Nasmyth 2001). Upon checkpoint activation, Pds1 is stabilized and sequesters the separin Esp1 into an inactive complex, thus precluding the release of the cohesin Mcd1/Scc1 from the chromosomes and sister chromatid separation (Clemenson and Marsolier-Kergoat 2006). The stabilization of Pds1 also concurs to inhibit mitotic exit.

The spindle-position checkpoint delays the activation of the mitotic exit network (MEN) until the spindle is properly aligned along the mother–bud axis. The MEN ultimately causes the activation of the phosphatase Cdc14. Cdc14 promotes mitotic exit by activating the APC/C-Cdh1 for degradation of mitotic B-type cyclins, and by stabilizing Sic1, for the inhibition of Clb–Cdc28 complexes (Bloom and Cross 2007a). Components of the MEN are restricted to the daughter SPB or the bud cortex, which prevents the initiation of this signaling cascade until the nucleus has entered the bud. MEN includes the

phosphatase Cdc14; the kinases Cdc5, Cdc15, Dbf2, and Dbf20; the GTPase Tem1; and the two-component GTPase-activating protein Bub2/Bfa1. The Bub2/Bfa1 complex inhibits Tem1, whose activation promotes APC-dependent destruction of B-type cyclins, activation of the Cdk1 inhibitor Sic1, and mitotic exit. *BFA1* and *BUB2* are required to prevent mitotic exit after activation of the DNA or the spindle assembly checkpoints.

3. Cell cycle goes global.

Wealth of data has been accumulated over decades of cell cycle studies on the particular steps and stages and their regulation at place. The availability of new functional genomics tools called researchers to make a big step forward in integrating these knowledge obtained through studies of distinct signaling pathways. Genome-wide cell cycle regulation analysis should have buttressed the detailed picture of known events and revealed new connections which remained obscure hitherto. Moreover, all eukaryotic cells experience important physiological changes during the cell cycle, and diverse biological events depend on the maintenance of this periodicity. Since mitosis imposes global regulatory events on the cell at each step to ensure proper maintenance and transfer of genetic information over to next generations, genes other than cyclins, CDK or their inhibitors were expected to be periodically co-regulated coherently with the cell cycle control genes and phases of their peak expression. Many of these regulatory emanations from the cell cycle to other cellular machineries remained obscure.

Earlier experiments, using traditional methods, identified 104 cell cycle regulated messages, and it was estimated that some 250 cell cycle–regulated genes might exist (Price *et al.* 1991). In 1981 Hereford and coworkers discovered that yeast histone mRNAs oscillate in abundance during the cell division cycle (Hereford *et al.* 1981).

Two almost simultaneous efforts were undertaken by Cho and colleagues (Cho *et al.* 1998) and Spellman (Spellman *et al.* 1998) to comprehensively identify all cell cycle-regulated genes in budding yeast.

One of the key mechanisms of gene regulation takes place at the level of mRNA transcription. In these studies, high-density oligonucleotide expression arrays were used to quantitate mRNA transcript levels in synchronized yeast cells at regular intervals during the cell cycle. Different synchronization methods were used, namely temperature-sensitive (ts) Cdc28 mutant alleles to arrest cells in G1, cdc15 ts mutant, which arrested cells in G2 and alpha-factor pheromone induced arrest at G1.

By assembling periodic genes into clusters of phase-dependent gene expression, using previously identified “phase landmarks” transcripts, characteristic of each stage of mitosis, Cho and colleagues

identified 416 genes that demonstrated consistent periodic changes in transcript level. This is approximately 7% of all yeast genes and this number agreed well with previous estimates of the number of genes in *S. cerevisiae* that display cell cycle–dependent transcription (Koch and Nasmyth 1994). The largest observed change in induction, 25-fold, was observed for *CLN1* and *RNR1* genes. 134 of the 416 cell cycle–regulated transcripts peaked in late G1, while only 56 transcripts peaked during M phase. Transcripts that peaked in late G1 displayed particularly sharp rates of accumulation and decay, while transcripts that peaked in S generally displayed a less dramatic induction pattern. More than half of the transcripts that peaked in late G1, including the *CLN1* and *CLN2* cyclins, displayed a minor peak in G2, which could indicate that a transcript is affected by more than one cell cycle–dependent regulatory sequence. An additional 33 of the 416 identified genes were induced in two different cell cycle phases, but did not display a predominant peak. This includes the cyclin-dependent kinase gene *CDC28*, which peaks twice, in G1 and G2. Having examined the upstream regulatory elements of the periodic genes, the investigators concluded that of all genes with an early cell-cycle box (ECB) element within 500 bp of the start codon, 67% displayed periodic transcription in early G1 phase. As expected at least two clustres of periodicity were formed – that coincident with the CLN1 expressiona and the CLB cluster.

Spellman et al. complemented this work by an additional time-series of periodic transcription and significantly improved the analysis. These data were analyzed by deriving a numerical score based on a Fourier algorithm (testing periodicity) and by a correlation function that identified genes whose RNA levels were similar to the RNA levels of genes alredy known to be regulated by the cell cycle (landmark principle) (Spellman *et al.* 1998). They identified ~800 cell cycle regulated genes, which constitutes 10% of all protein-coding genes in the genome.

Clusters of co-regulated genes were established using the clustering algorithm, which sorted through all the data to find the pairs of genes that behave most similarly in each experiment and then progressively adds other genes to the initial pairs to form clusters of apparently coregulated genes (Eisen *et al.* 1998). These clusters provided a foundation for understanding the transcriptional mechanisms of cell cycle regulation. The major functions of the identified cell cycle regulated genes were cell cycle control, DNA replication, DNA repair, budding, glycosylation, nuclear division and mitosis, structure of the cytoskeleton, and mating. Additionally promoter elements 700 bp immediately upstream of the start codon of each of the 800 genes in periodic list were examined and the majority of them matched well to known cell cycle transcription factor binding sites, including SCB and MCB as well as four extensions and modifications of MCM1, SFF, extended SWI5, SCB variant, and degenerate MCB.

One of the largest discrepancies between the two genome-wide analyses concerned genes that might peak twice per cell cycle. Fourier transform algorithm is designed to find genes with single expression peaks; it significantly penalizes more than one peak. Thus Spellmam et al. identified only 10 genes as cell cycle regulated that according to Cho *et al.* showed more than one peak but had no single prominent peak in expression (Cho *et al.* 1998).

This leaves open the possibility that there may be other genes with more than one peak per cell cycle. A number of these genes, whose expression was affected by induction of Cln3p and Clb2p and which harbor relevant upstream DNA motifs remains to be elucidated with regard to their cell cycle behavior (Spellman *et al.* 1998).

Detailed microarray analysis described above has revealed that the expression levels of approximately 800 genes vary in a periodic fashion during the yeast cell cycle (Cho *et al.* 1998; Spellman *et al.* 1998), but little was known about the regulation of most of these genes. The set of genes controlled by MBF and SBF has been identified by using a genome-wide binding method, confirming that these factors are largely bound to genes expressed in late G1 and revealing how sets of functionally related genes are regulated during this time (Iyer *et al.* 2001). A more complete understanding of cell cycle regulation is constrained, however, by our limited knowledge of the transcriptional regulatory network that controls the clock.

Transcription factors have been identified that have roles in regulating transcription of a small set of yeast genes whose expression is cell-cycle dependent; these include Mbp1, Swi4, Swi6, Mcm1, Fkh1, Fkh2, Ndd1, Swi5, and Ace2 (Breeden 2000; Koch and Nasmyth 1994; Kumar and Carmichael 1998; Mendenhall and Hodge 1998). Based on these studies, the following model has emerged. MBF (a complex of Mbp1 and Swi6) and SBF (a complex of Swi4 and Swi6) control late G1 genes. Mcm1, together with Fkh1 or Fkh2, recruits the Ndd1 protein in late G2, and thus controls the transcription of G2/M genes. Mcm1 is also involved in the transcription of some M/G1 genes. Swi5 and Ace2 regulate genes at the end of M and early G1. It is not yet clear whether this model, developed using a small set of genes, will extrapolate to regulation of all cell cycle genes.

Identification of the genes regulated by all nine transcription factors in living cells was essential for further understanding how the cell cycle is regulated at the transcriptional level. Genome-wide location analysis was used to determine how the yeast cell cycle expression program is regulated by each of the nine known cell cycle transcriptional activators (Lee *et al.* 2002; Simon *et al.* 2001). Simon et al. reported the genomic targets of all nine known cell cycle transcription activators in living yeast cells. The results reveal a fundamental feature of cell cycle regulation in living cells: cell cycle transcriptional

activators that function during one stage of the cell cycle contribute to the regulation of transcriptional activators that function during the next stage, forming a fully connected regulatory circuit. In addition, transcription factors that regulate the cyclin genes during each phase of the cell cycle also regulate genes involved in transitioning on to the next stage (Fig. 1-18). For example, the G1/S activators SBF and MBF control transcription of G1/S cyclin genes, but also regulate expression of the G2/M cyclin Clb2, which subsequently inhibits further expression of the G1/S cyclins Cln1 and Cln2 and promotes entry into mitosis. Thus, the cell cycle transcriptional regulatory network has evolved so that some transcriptional regulators contribute to the control of both stage entry and exit. The identification of sets of genes that are bound by each of these regulators reveals how coordinate regulation of a wide variety of stage-specific cell cycle functions is regulated. For example, the G1/S activators regulate genes involved in cell budding, DNA replication and repair, and chromosome maintenance. The G2/M activators bind genes that regulate the transition through mitosis. The late M factors regulate genes involved in cytokinesis and prereplication complex formation (Fig. 1-19).

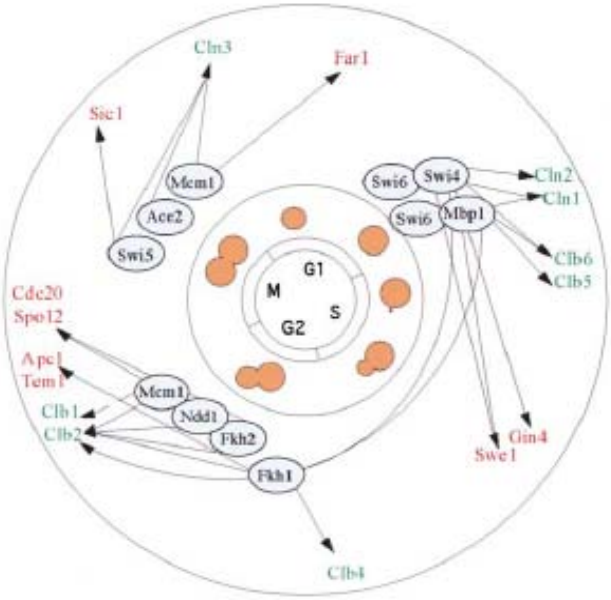


Figure 1-18. Model for transcriptional regulation of cyclin and cyclin/CDK regulators. Each group of transcription factors regulates key cell cycle regulators that are needed for progression through the cell cycle. (adapted from (Simon *et al.* 2001)).

A more comprehensive picture of cell cycle regulation emerges when existing knowledge of cell cycle regulatory mechanisms is combined with the new information on the transcriptional regulatory network. For instance, the emphasis on START regulation at the G1/S boundary is evident from the regulatory events involving Swi4 (Fig. 1-20). The Swi4 regulator becomes functionally active at START, via a mechanism that is dependent on Cln3-Cdc28, when the cell reaches a critical size (Dirick *et al.* 1995). The *SWI4* promoter is bound by Swi4 itself, froming a positive feedback loop, exists to ensure that

adequate levels of Swi4, and thus SBF, are present prior to commitment to start. To get to the next step Ndd1 is required to initiate the G2/M transcriptional program. It is a limiting component of the activator complex, as Mcm1 and Fkh2 are bound to promoters throughout the cell cycle, and activation depends on proper recruitment of Ndd1. The observation that the G1/S regulators SBF and MBF both regulate *NDD1* suggests how appropriate levels of Ndd1 are reached. The Mcm1/Fkh2/Ndd1 complex regulates *SWI5* and *ACE2*, whose products become functional only in late anaphase after relocalization to the nucleus in a mechanism that is dependent on low Clb-Cdc28 activity (Nasmyth 1999; Shirayama *et al.* 1999). Later in the cell cycle, the Swi5, Ace2, and Mcm1 factors all bind to the *CLN3* promoter, thus assuring adequate levels of the Cln3 cyclin at START. Thus cells that have entered the cell cycle at START may progress through an entire cycle because of the design of the connected transcriptional regulatory network, and perhaps then arrest in G1 because of the requirement for adequate levels of Cln3/Cdc28 (Simon *et al.* 2001).

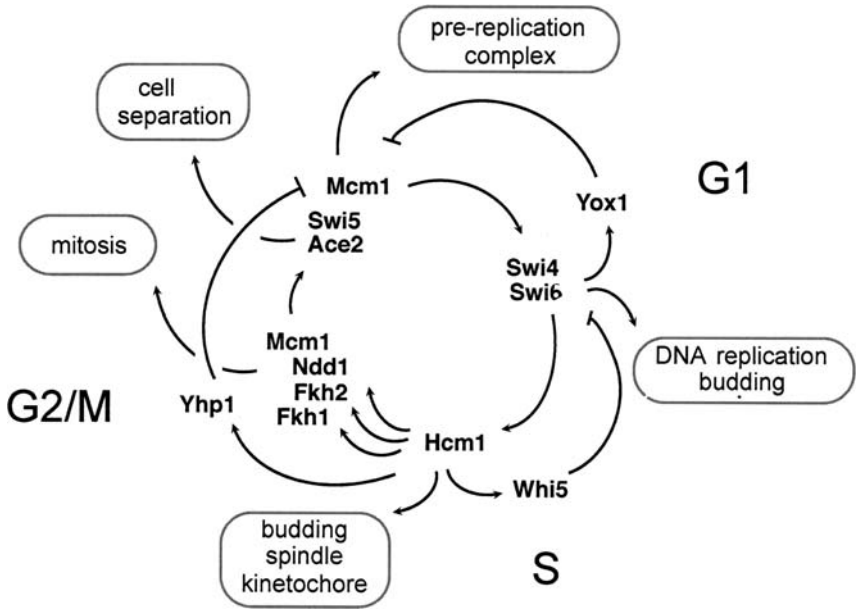


Figure 1-19. Trancription factor circuitry in the cell cycle. Hcm1 functions as an S-phase-specific transcriptional activator. *HCM1* is activated by the Swi4/Swi6 complex. It then transcriptionally activates *WHI5*, which represses the subsequent round of Swi4/swi6 targets until late G1. Similarly, activation of *YHP1* maintains the repression of M/G1 transcription. At the same time, activation of *FKH1, FKH2*, and *NDD1* induces the next wave of G2/M transcription. This model connects the known cell cycle regulatory TFs to each other in a continuous cycle. However, there are hundreds of transcripts that are not targets of these factors that must be accounted for before we have a comprehensive picture of the cell cycle-regulated transcription that underlies the cell cycle. (adapted from (Pramila *et al.* 2006))

The results also reveal how these cell cycle-specific transcriptional regulators control key temporal features of the cell cycle and how coordinate control of genes with shared stage-specific functions is accomplished. This information can ultimately help to construct a map of the transcriptional and posttranscriptional regulatory networks that control the complex and highly regulated processes that occur throughout the cell cycle. It is important to point out that part of the tight regulation of cell cycle is also due to combinatorial control by multiple regulatory elements (Lee *et al.* 2002).

To better define the transcriptional circuitry determined by other groups Pramila et al. collected new refined microarray data with higher resolution across the cell cycle of budding yeast (Pramila *et al.* 2006). Performing the combined analysis with other datasets by improved computational method (de Lichtenberg *et al.* 2005b) the group has identified new periodic transcripts and additionally a late S-phase specific promoter element (Pramila *et al.* 2006).

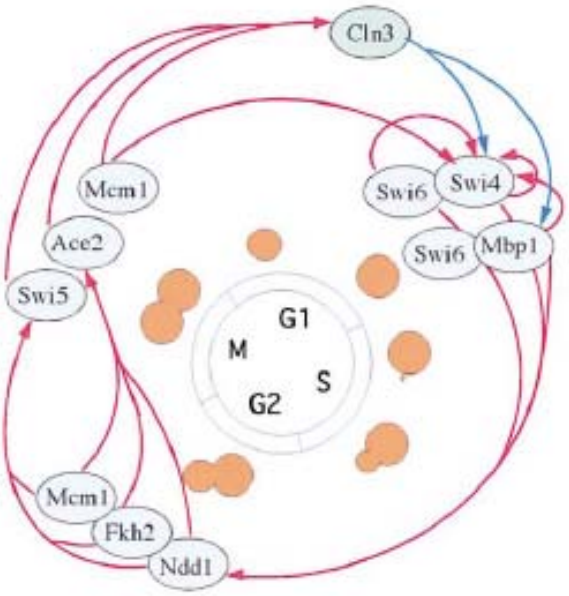


Figure 1-20. Model for the closed regulatory circuit produced by cell cycle transcriptional regulators based on genome-wide binding data. The genome-wide location data indicate that each group of transcriptional activators regulates activators acting in the next cell cycle stage. The red arrows represent binding of a transcription factor to the promoter of another regulatory factor. The blue arrows represent posttranscriptional regulation. (adapted from (Simon *et al.* 2001)).

One other outcomes of the global scrutiny of periodic transcription through the cell cycle is the realization that gene products required for a certain stage in the cell cycle are often peak just at time appropriate for their function (de Lichtenberg *et al.* 2005a). Unlike bacterial cell cycle, in yeast just-in-time synthesis of entire complexes is rarely observed. The only exception is nucleosome, all subunits of which are expressed in S phase. The general design principle is that only some subunits of the complex are periodically regulated to ensure temporal activation of a complex in an appropriate phase of the cell cyle. Just-in-time assembly has an advantage over just-in-time synthesis in that only a few components of a complex need to be tightly regulated. Thus just-in-time transcription may not only reinforce the order of events during the cell cycle, but may also replenish the pools of proteins, which are inactivated after having performed their function in mitosis (de Lichtenberg *et al.* 2005a).

The accelerating pace of development of functional genomics tools called for meta-level integration of genome-wide data available about yeast cell cycle. Since the major goal of the cell cycle research is to determine the complexity of the underlying system by identifying all genes and proteins regulated in it,

proteomic approach has complemented transcriptome studies. Given that overlap between different transcriptome studies is relatively small, mainly due to discrepancies in the methods of follow-up computational analyses and differencies attributed by synchronization methods, de Lichtenberg and colleagues investigated protein features that could be used to distinguish highly regulated cell cycle periodic proteins (de Lichtenberg *et al.* 2003). They have provided an overview of proteome dynamics and demonstrated that co-expressed periodic genes encode proteins, which share combination of features such as phosphorylation, glycosylation, subcellular location and instability/degradation). Following a new periodicity analysis of a combined set of publicly available microarray data (de Lichtenberg *et al.* 2005b), a training set was selected, comprising 97 proteins, which displayed very significant periodicity during cell cycle, as well as 556 non-periodic genes. Both sets displayed large diversity of biochemical features. Neural networks were trained to distinguish cell cycle proteins from non cell cycle proteins based on combination of these feature (de Lichtenberg *et al.* 2003). The strength of the feature-based approach is its independence of experimental errors and biases. Hence it should add new touches to the cell cycle periodicity palette.

The pace of development and application of functional genomic reagents has accelerated over the past few years (Bader *et al.* 2003), and has now reached the point where virtually every cell cycle study has a genomic slant of one form or another. Recent functional genomic and proteomic approaches have yielded new insights into almost all aspects of cell cycle control, including transcriptional circuits, DNA replication, sister chromatid separation and regulation by environmental signals. Most notably, systematic analysis has begun to reveal meta-level connections between previously distinct subprocesses. As the interconnections between these huge datasets are beyond intuition, mathematical representation and automated analysis of functional genomic are also developing rapidly (Tyers 2004). The combination of these approaches may culminate in a watershed in our understanding in the cell cycle field, including its myriad connections to development and disease.

Aim of the thesis

There is abundant transcription from eukaryotic genomes unaccounted for by protein coding genes. Experimental and computational studies have confirmed that non-coding RNAs play essential roles in functionality of different pro- and eukaryotic organisms. They exert their function in different physiological contexts, interspersing with the proteins to govern important cellular processes.

The aim of this thesis is to fulfill a high-resolution genome-wide survey of transcription in a well annotated genome, which will ultimately help relate transcriptional complexity to function. By quantifying RNA expression on both strands of the complete genome of *Saccharomyces cerevisiae* using a high-density oligonucleotide tiling array, this study aims to identify the boundary, structure, and level of coding and noncoding transcripts.

Given the very high resolution of experimental data obtained with tiling arrays it is well worth to comprehensively characterize periodic transcription during the cell cycle in order to better define transcriptional circuitry, which shifts dramatically during cells' transition through cell cycle phases. Moreover, it is of great interest to delineate the regulatory roles of antisense messages in the well characterized, but nonetheless complex functional mechanism of mitotic division.

Combined with computational methods to relate these data and map the complex interactions of transcriptional regulators, these tiling array experiments will provide insight towards a more comprehensive understanding of fundamental molecular and cellular processes and the role of non-coding RNA transcripts therein.

.

2. RESULTS

I. A high-resolution map of transcription in the yeast genome

1. Introduction

Proteins constitute most structural and functional components of cells. The assumption has been that protein-encoding genes are also the main controllers of cellular processes. Recent evidence challenges this assumption, suggesting a wide-spread involvement of noncoding RNA in regulation, including through the activity of untranslated regions of mRNAs (Wilkie *et al.* 2003), antisense transcripts (Katayama *et al.* 2005; Storz *et al.* 2005), and isolated non-coding RNAs such as microRNA that control transcript levels or their translation (Mattick 2004).

High-resolution transcriptome analysis in higher eukaryotes using tiling arrays has improved ORF annotations and exon-intron predictions and discovered many new transcripts of currently unknown function (Bertone *et al.* 2004; Cheng *et al.* 2005; Yamada *et al.* 2003). However, these studies have encountered challenges, due to noise, limited resolution, lack of strand-specific signal, and drawbacks in the analysis methods (Royce *et al.* 2005). Sequencing of cloned cDNAs has also revealed a high level of transcriptional complexity, including the presence of many new transcripts, alternative promoter usage, splicing, and polyadenylation, as well as the presence of many sense–antisense transcript pairs (Carninci *et al.* 2005; Katayama *et al.* 2005). However, because of the cost and labor of large-scale sequencing, this approach has been limited. Therefore, there was a need to develop high-throughput, precise, and high-resolution technology to map the full transcriptional activity. Yeast is a simple and relatively small eukaryotic genome that provides opportunities to rapidly characterize novel findings.

Lars Steinmetz and Ron Davis developed an oligonucleotide array for *Saccharomyces cerevisiae* that contains 6.5 million probes and interrogates both strands of the full genomic sequence with 25-mer probes tiled at an average of eight nucleotide intervals on each strand (17 nucleotides overlap) and a four nucleotides offset of the tile between strands. This design enables a 4-nt resolution for hybridization of double stranded targets and an 8-nt resolution for strand-specific targets.

We profiled transcription during exponential growth in rich media (YPD), the standard laboratory growth condition, to generate a comprehensive map of transcription. Our study has been published (David *et al.* 2006) and is largely reproduced here below. My contribution to this work comprises optimization of total RNA and PolyA-RNA isolation procedures, optimization of RT reaction and priming methods for

polyA-RNA and total RNA fraction as well as optimization of hybridization conditions for all types of samples, bioinformatics analysis on the initial stages of the project, performance of all experimental work to generate complete polyA-RNA data set, DNA control hybridizations and involvement in the manuscript preparation.

2. Microarray Experiments and Analysis.

We hybridized first-strand cDNA synthesized using random primers from polyadenylated [poly(A)] and total RNA. To calibrate the sequence-specific probe effect (Hekstra *et al.* 2003; Naef and Magnasco 2003; Wu and Irizarry 2005), we background-corrected and adjusted (Huber *et al.* 2002) the signal of each probe by sequence-specific parameters, estimated from a calibration set of genomic DNA hybridizations (Fig. 2-1). This method allowed us to quantitatively compare the signal from probe to probe on the array.

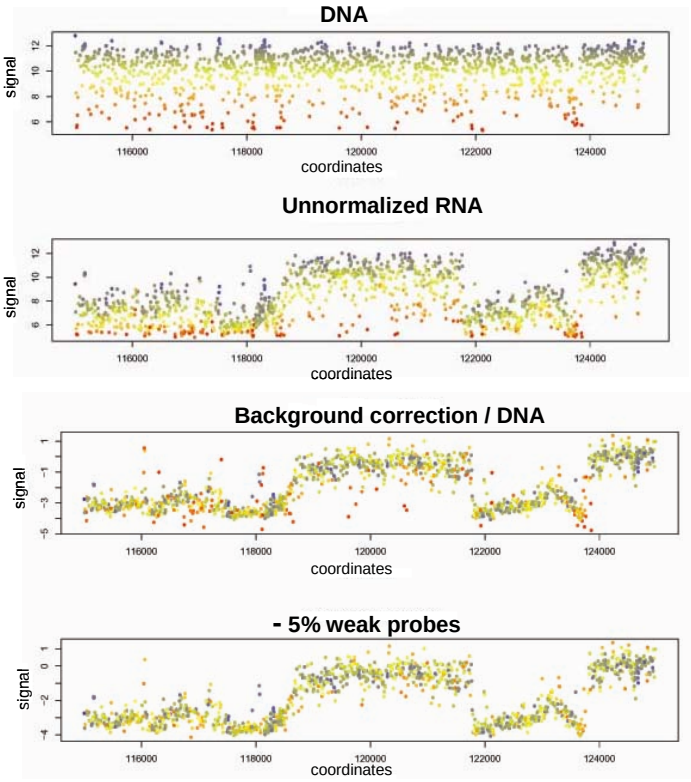


Figure 2-1. Probe normalization using genomic DNA hybridization signals. The plots of a genomic region show a comparison between signals from genomic DNA, unnormalized data for poly(A) RNA and successively more effective normalization methods. The colors correspond to the DNA signals (from dark red = weak, to dark green = strong). The normalization methods are as follows: Method 1: divide RNA-signal by DNA-signal then take logarithm (base 2). (Not shown in the figure.) Method 2: background subtraction of the RNA-signal, divide by DNA-signal, then variance stabilizing normalization (vsn, glog base 2). Method 3: in addition to method 2, drop the 5% weakest probes in the DNA hybridization. Method 3 was chosen because it yielded the highest gain in signal to noise.

3. Normalization and quality control.

Normalization is an essential procedure to compare the data from multiple one-color arrays. It removes sources of variation of non-biological origins between arrays, such that observed differential expression

is solely due to biological variation between individual samples. Among the normalization approaches developed in the recent years the most widely applied for tiling arrays are so called complete data methods, which make use of data from all arrays in the experiment to form the normalization relation among which quantile normalization method performs best (Bolstad *et al.* 2003).

The fluorescence intensity values obtained from oligonucleotide microarray hybridization do not directly correspond to interpretable physical units. The same abundance of a target transcript can result in systematically different values when measured with different oligonucleotide probes. This is due to a variety of reasons, among them the different thermodynamic properties of different polynucleotide sequences and biases in labeling efficiency (Huber *et al.* 2006).

We have applied the Variance Stabilization and Normalization (VSN) method, which builds upon the fact that the variance of microarray data depends on the signal intensity and that a transformation can be found after which the variance is approximately constant. Vsn assumes that less than half of the genes on the arrays is differentially transcribed across the experiment. An advantage of *vsn*-transformation over log-transformation is that *vsn* works also on values that are negative after background subtraction (Huber *et al.* 2002).

3.1 Estimation of array performance

To address the question of how much of the genome is transcribed, we analyzed the coding regions of 5,654 ORFs that were annotated as verified or uncharacterized genes in the *Saccharomyces* Genome Database (SGD, www.yeastgenome.org) and represented by unique probes on the array. Significant expression above background was detected for 90% of the yeast genes (Binomial test, false discovery rate = 0.001). The transcripts not detected by the tiling array fall into cellular processes not expected during exponential phase in rich media, such as mating, sporulation, meiosis, vitamin metabolism (GO categories enrichment). In addition, analyzing 11,412,997 bp of unique genomic sequence, we detected expression above background on either strand for 85%. Comparing this to existing annotation, which covers ~75% of the genome, shows that 16% of the transcribed base pairs had not been annotated before.

4. Segmentation and detection of new transcripts

To obtain an unbiased map of the position, abundance, and architecture of transcripts, the hybridization signals were examined along their chromosomal position for each strand (Fig. 2-2).

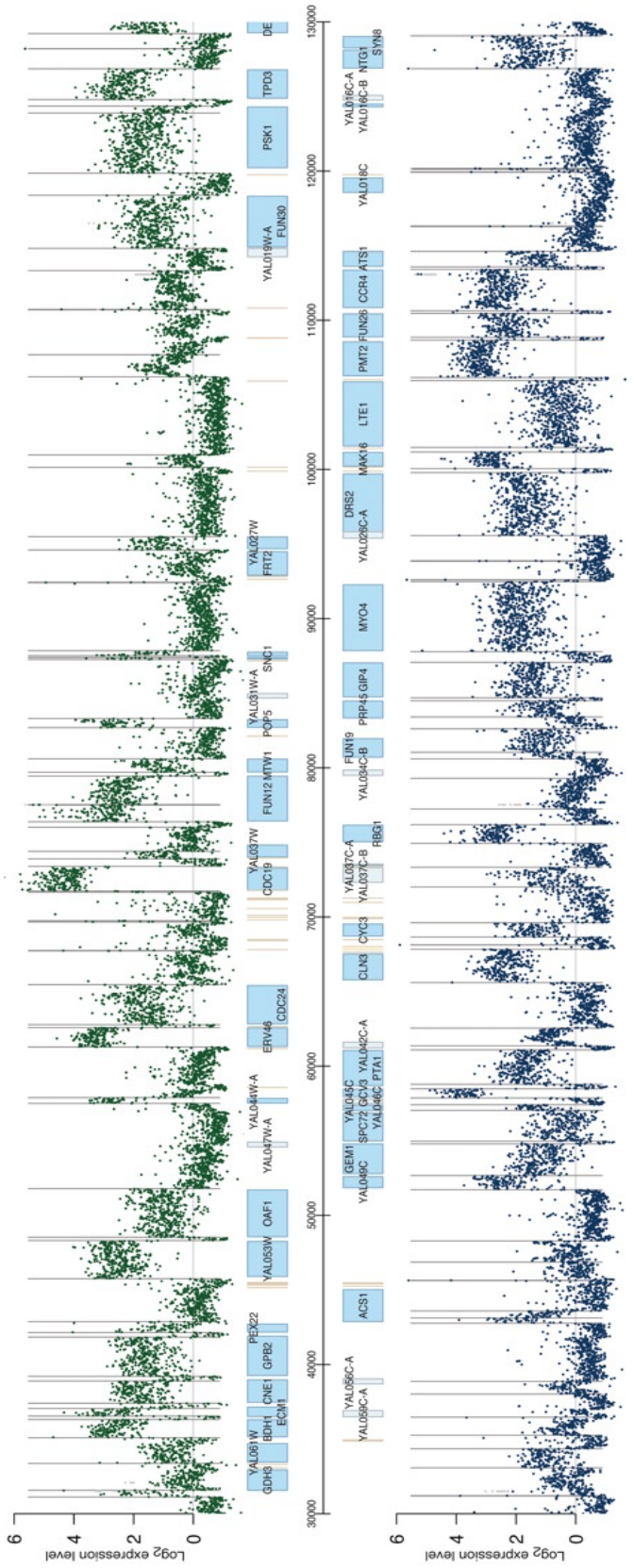


Figure 2-2. Visualization of yeast tiling array intensities along 100 kb of chromosome 1, corresponding to 1% of the genome. The plot shows the normalized hybridization intensities (y axis) along genomic coordinates (x axis in bp). Each dot corresponds to a probe, Watson strand in green and Crick strand in blue. Probes with more than one perfect match in the genome are colored gray. Annotated ORFs are shown as blue boxes, dubious ORFs are shown as light blue boxes, and transcription factor binding sites are shown as gray bars. Vertical lines are segment boundaries. The background threshold (y = 0) is shown as a horizontal line.

The profiles were partitioned into segments of constant hybridization intensity, separated by change points demarcating transcript boundaries. We used a change point detection algorithm that determines the global maximum of the log-likelihood of a piece-wise constant model by dynamic programming (Gentleman 2005; Picard *et al.* 2005).

Compared to running-window approaches, it finds more accurate estimates of change point locations and depends on fewer user-defined parameters. Segments were determined separately for poly(A) and total RNA. Segments from poly(A) and total RNA were remarkably concordant, and many noncoding RNA (ncRNA) were also found in the poly(A) data. Overall, the poly(A) RNA hybridization data were cleaner and therefore were the focus of our analysis. The automated segmentation algorithm provides an unbiased global analysis, but the data complexity invites additional manual curation. Profiles for all genomic regions are provided in a database that is searchable by gene symbol or chromosomal coordinate (www.ebi.ac.uk/huber-srv/queryGene). With this approach we mapped transcription in regions of genome absent of prior annotation, exons in spliced genes, transcripts opposite an annotated feature, untranslated regions of an ORF and discovered unusual transcription architecture, such as overlapping transcripts for neighboring genes and uneven transcription levels for different segments of a coding region (Fig. 2-3).

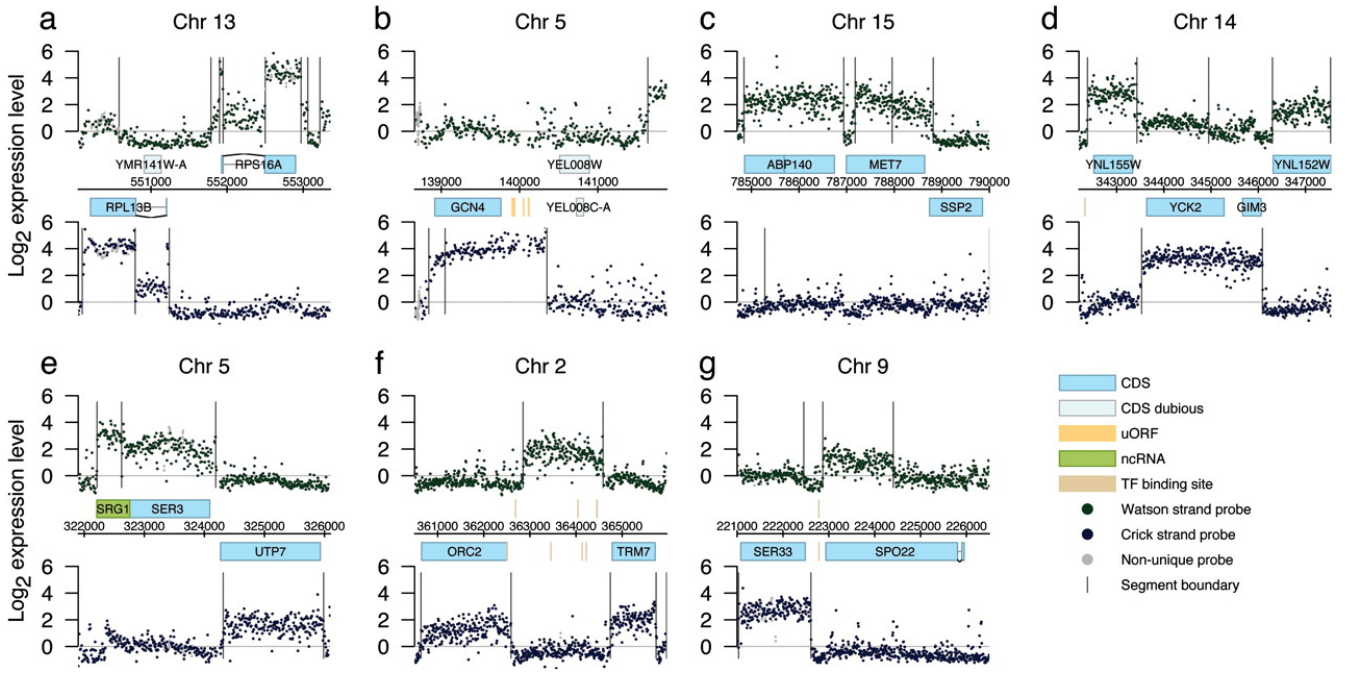


Figure 2-3. Examples of transcriptional architecture. (*a*) Detection of spliced transcripts. (*b*) Long 5' UTR of *GCN4* including its cotranscribed upstream ORFs. (*c*) Complex transcript architecture of *MET7*. (*d*) Overlapping transcripts of two ORFs. (*e*) Adjacent transcripts of *SER3* and the noncoding *SRG1*. (*f*) Nonannotated isolated transcript. (*g*) Transcript antisense to *SPO22*. CDS refers to coding sequence; uORF, upstream ORF; ncRNA, noncoding RNA; TF, transcription factor. Plot layout as in Fig. 2-2.

5. Mapping untranslated regions (UTRs) of protein-coding genes.

To map UTRs, we compared ORF boundaries with segment boundaries. We automatically determined UTR lengths for verified or uncharacterized nuclear-encoded genes whose annotated coding sequence was fully contained within a single segment. A total of 2,223 ORFs were selected to map UTRs with a stringent filter, which required a sharp signal decrease on both transcript boundaries and a presence of exactly one ORF in the segment. We proceeded with analysis of the 2,044 poly(A)-determined UTRs because the poly(A) hybridization data were cleaner and yielded most of the UTR determinations. For many remaining genes that did not pass the confidence filter, the UTRs can be mapped by closer inspection.

We found that 3' UTRs were significantly longer than the 5' UTRs, with a median of 91 vs. 68 nt (Fig. 2-3a). Longer 3' UTRs are consistent with them containing posttranscriptional regulatory regions that influence mRNA stability, localization, and translation (Kuersten and Goodwin 2003), and with findings from other species (Mignone *et al.* 2002). The mean sum of 3' and 5' UTR lengths was 211 nt and similar to a mean of 256 nt found by a gel-mobility assay (Hurowitz and Brown 2003). We computed 662 3' UTRs from ESTs (Graber *et al.* 1999) and compared them to 435 ORFs that had UTRs in our data set. The Pearson correlation coefficient between the UTR length estimates was 0.63. A contribution to the differences is that in the EST data the longest transcript was chosen, whereas the array measures the average transcript abundance at each probe position.

We compared UTR lengths with transcript levels and coding sequence (ORF) lengths. Although transcript level was generally lower for genes with long coding sequences, neither transcript levels nor ORF lengths were significantly associated with UTR lengths. We also compared length distributions of UTRs for different functional and localization categories (GO annotations) and detected significant correlations (Fig. 2-4). The longest 3' UTRs were found for transcripts of proteins that are targeted to the mitochondrial electron transport chain, the plasma membrane, and the cell wall. These longer 3' UTRs may contain mRNA localization signals, as has been well demonstrated for mitochondrial targeted proteins (Gerber *et al.* 2004; Marc *et al.* 2002). Genes involved in phosphorylation, transporter activity, ion transport, and specific stages of the mitotic cell cycle had both ends longer. Genes involved in RNA processing, rRNA metabolism, and ribosome biogenesis had both ends shorter. Therefore, genes with longer UTRs seem to fall into categories that require regulation, whereas genes with short UTRs seem to fall into categories with a reduced need for posttranscriptional regulation, such as housekeeping genes.

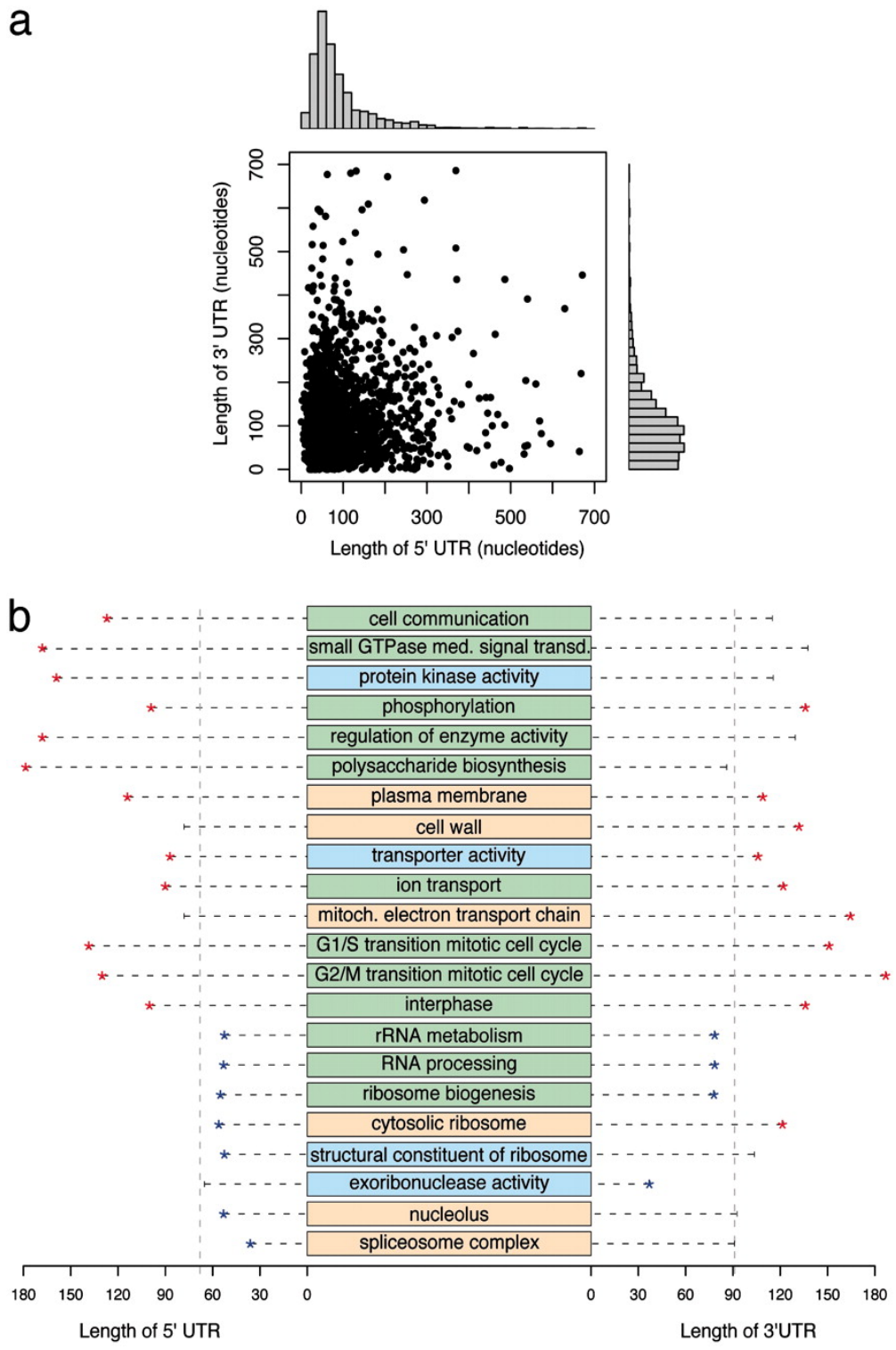


Figure 2-4. Length of UTRs and functional categories with exceptional UTR length. Analyses were based on 2,044 genes from poly(A) samples. **(*a*)** Scatterplot and histogram of 3' vs. 5' UTR lengths. **(*b*)** Association between UTR length, cellular localization, and biological process. Length distributions between genes inside and outside of GO categories were compared, and selected significant categories are shown (orange, cellular component; green, biological process; blue, molecular function). For each category, a horizontal line shows the 5' and 3' median UTR lengths measured in nucleotides (*x* axis). The median over all genes is shown by a vertical dashed line. Significant medians are indicated by asterisks, red longer, blue shorter (two-sided Wilcoxon test, P 0.002).

6. Complex transcriptional architecture

Many expressed segments flanked other expressed segments with different signal levels, thus making up complex transcriptional architectures. In many cases, different parts of the same gene are expressed at different levels: 921 ORFs from the poly(A) RNA sample were divided into at least two expressed segments, one covering >50% of the feature and others <50%. Such complex architectures could be due to alternative transcription initiation, termination, or alternative splicing, as has been described in mouse (Carninci *et al.* 2005) and for several human genes (Kapranov *et al.* 2005), as well as multiple overlapping transcripts with different lengths, alternative RNA splicing, RNA secondary structures or regional transcript instability. In yeast, it has been suggested that up to 20% of mRNAs have alternative 3' ends (Graber *et al.* 2002). Additional unusual architecture was found in poly(A) RNA data, where one segment included complete or partial annotations of more than one gene. This indicates that promoters of one gene could be contained within the transcribed region of another gene. This observation provides the basis for further investigations of regulatory mechanisms of closely spaced and possibly overlapping transcripts.

Complex hybridization patterns on the array could also be caused by RNA decay or variation introduced by reverse transcription, because the array captures the sum of cDNA molecules present at the time of hybridization. The explanation of our observations by such mechanisms will require a case-by-case analysis.

To illustrate a few cases, for *CPB1* and *RNA14*, our observed architecture matches previous results describing alternative 3' ends in response to carbon source regulation (Sparks and Dieckmann 1998). For *GCN4*, lower hybridization signal was observed at the 3' end (Fig. 2-3b). *GCN4* is not translated during nutrient-rich growth because of the translation of the upstream ORFs encoded in the same transcript (Hinnebusch 2005). 3' end degradation due to a lack of translation could explain the lower 3' end signal. In support, this decrease was not seen in an oligo(dT) reverse-transcribed sample, where no priming would occur on degraded poly(A) transcripts. At the 5' end, the segment boundary matches the previously determined position to within nine bases. For *MET7*, the annotated gene was segmented into three regions (Fig. 2-3c), suggesting a misannotation of the translation start site. A later transcription start site is supported by the multiple sequence alignment of yeast species in SGD, which shows that conservation of *MET7* starts at a later methionine (M55), whose position agrees with the transcription start site detected by the array. Also, the level difference between the central and the 3' segment was not seen in a poly(A) sample that was reverse transcribed by using oligo(dT) primers consistent with early

transcript termination or RNA decay. Altogether, we tested 27 regions from 10 genes by quantitative real-time PCR (including *MET7*). For seven genes, the PCR results matched the architectures in the array data.

7. Neighboring Transcription

Additional unusual architecture was found for adjacent ORFs not separated by an unexpressed region. Such architectures can result from more than one ORF being encoded from a single transcript, like the upstream ORFs in *GCN4*, or from distinct transcripts not separated by untranscribed intergenic regions. We found the ORFs of *GIM3* and *YCK2* within one segment resembling a bicistronic transcript (Fig. 2-3d). The PhastCons multiple alignments (Siepel *et al.* 2005) of the intergenic region with other yeast species shows high sequence conservation, but includes frame-shifting gaps, which suggests that the two ORFs are not translated as one. By reverse-transcription PCR across the gap between the ORFs, a product was obtained supporting either bicistronic or overlapping transcripts. Operon-like transcription was reported for few eukaryotic species and mostly for *Caenorhabditis elegans* (Blumenthal and Gleason 2003). A bicistronic transcript had been reported previously in yeast for *YMR181C* and *RGM1* (He *et al.* 2003), and we observed different transcript levels for the ORFs, but no separation by an untranscribed region.

Figure 2-3e shows two other adjacent transcripts, *SRG1* and *SER3*, expressed at different levels and not separated by an intergenic region. It had been proposed that *SRG1*, an upstream noncoding RNA, represses the expression of *SER3* in rich media, by reducing the binding of *SER3* transcription factors (Martens *et al.* 2004). In contrast, we find that *SER3* is expressed significantly above background, suggesting that even though *SRG1* is expressed at a higher level, its transcription does not prevent *SER3* from being transcribed. There are many cases of adjacent genes not separated by unexpressed, intergenic regions in our data set, and this suggests that transcription over active promoters of adjacent genes is common in yeast. Some further examples are *QCR6*, *PHO8*, *RIB3*, *HCH1*, *UBI4*, *SEC53*, *RPS26A*, and *ADE13*.

8. Detection of unannotated transcripts

Many segments with signal above background did not overlap existing annotation (1943 in polyA and 1688 in total RNA). They fall into two classes: nonannotated isolated segments if there was no prior annotation on either strand (Fig. 2-3f), and nonannotated antisense segments if there was an annotation on

the opposite strand (Fig. 2-3g). Many are not independent transcription units: some represent UTRs of genes with complex transcriptional architecture; others are part of unannotated transcripts that are divided into multiple segments. The identification of antisense transcripts requires caution because reverse transcription can generate double stranded cDNA from secondary mispriming (Johnson *et al.* 2005; Kampa *et al.* 2004). Considering these concerns, we applied a filter requiring segments to be at least 48 bp long, be flanked by segments with reduced hybridization signal on both sides, and have higher expression signal than seen on the opposite strand for at least part of their length. In these filtered categories, we obtained 427 nonannotated segments from poly(A) and 357 from total RNA hybridizations. These segments divide approximately equally into the isolated and antisense categories (Fig. 2-5a). Antisense segments and segments overlapping annotation (>/=50%) had similar length distributions and tended to be longer than those of the isolated categories (Fig. 2-5b). Isolated segments showed similar levels of expression as annotated segments, whereas antisense segments had lower expression levels (Fig. 2-5c).

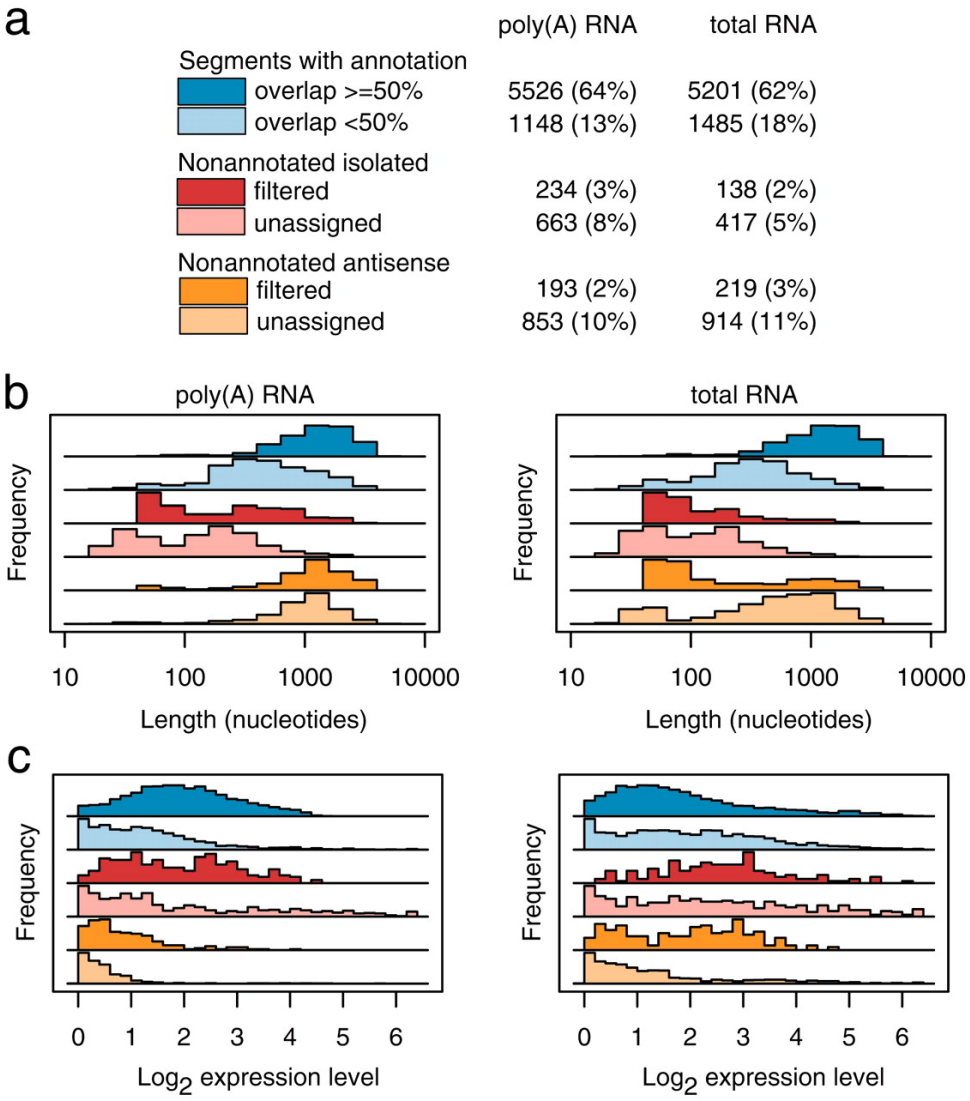


Figure 2-5. Categories of expressed segments, their length, and their expression levels. (*a*) Number and percentage of the expressed segments detected from the poly(A) RNA and total RNA hybridizations. Categories ">= 50%" and "<50%" consist of segments that overlap more, or less, than half of an annotated feature, respectively. The "nonannotated isolated" category consists of segments that have no overlap with annotated features on either strand, whereas the "nonannotated antisense" category consists of those that overlap with features on the opposite strand. The "filtered" categories consist of the high confidence segments that passed our filter, and the "unassigned" categories consist of the remaining segments. Length **(*b*)** and transcript level **(*c*)** distributions for segments from the above categories are given.

9. Detection of novel transcripts

We verified the array identification of 126 nonannotated isolated transcripts by RT-PCR. All were expressed in both total and poly(A) RNA reverse transcribed by using random and oligo(dT) primers, respectively. For 10 of them, a quantitative real-time PCR analysis showed their levels to be similar to expressed ORFs in both sample types.

The 1.7-kb transcript between *ORC2* and *TRM7* (Fig. 2-3f) is an example of a nonannotated isolated segment that is highly conserved across other yeast species in the PhastCons multiple alignment (Siepel *et al.* 2005). We analyzed evolutionary conservation of novel isolated transcripts by multiple alignment of *S.cerevisiae* genome with *S.paradoxus*, *S.mikitae* and *S.bayanus*. Assessment of 125 nonannotated isolated segments showed that only 48 had a multiple alignment >50 nt across four yeast species (Kellis *et al.* 2003). To assess protein-coding potential, we tested for dissimilarity in evolutionary rates among first, second, and third codon positions in all reading frames. There was no protein coding signature for the 48 nonannotated segments (median likelihood-ratio statistic of 1.06, compared to 1.14 for undetected unannotated segments and 162.0 for verified genes). Conservation of DNA sequence or protein coding ability is nevertheless neither a necessary nor a sufficient attribute of transcript function.

We generated knockouts for 47 nonannotated isolated segments and tested for growth defects in rich media conditions. A growth defect was identified for two knockouts: one on chromosome 6, positions 54813–55221, the other on chromosome 7, positions 622039–622295. On chromosome 6, the deleted segment contained annotated transcription factor binding sites upstream to *ACT1*, an essential gene, which likely accounts for the observed inviability. On chromosome 7, the deletion does not overlap any annotation, and strains with deletions of the neighboring ORFs (*YGR066C*, *YGR067C*) did not have a growth defect. This segment does not appear to be evolutionarily conserved or to contain a long ORF. The proportion of growth defects found within the 47 knockouts is much lower than the ~ 40% found for knockouts of protein-coding genes (Winzeler *et al.* 1999b).

10. Function of antisense transcripts.

Regulation of transcript-translation by antisense sequences was reported in prokaryotes (Heidrich and Brantl 2003) and higher eukaryotes (Calderon and Lavergne 2005). We identified antisense transcripts opposite to 1,555 genes, of which 402 were in the filtered set from both poly(A) and total RNA samples . The extent of antisense transcription suggests that it may have a previously undiscovered function in

yeast. *S. cerevisiae* lack the protein machinery for post-transcriptional mRNA regulation by degradation of double stranded RNAs like miRNAs.

The antisense transcripts are not caused by read-through from ORFs on the opposite strand, but appear as independent transcription units. For example, antisense transcription was found opposite *SPO22*, a meiosis-specific protein induced early in meiosis (Fig. 2-3g). Upstream of this antisense transcript, there is a binding site for *CBF1*. *CBF1* is involved in regulation of DNA replication and chromosome cycle and is important for growth in rich media, suggesting that the antisense expression may be negatively correlated with the expression of *SPO22*. Many genes with antisense transcripts had products that localize to the cell cortex and cell wall, and that function in the meiotic cell cycle and in transcriptional regulation (table 2-1). Some of these categories included genes not active during growth in rich media, like meiosis. Others included genes that are active during growth in rich media, but which may need posttranscriptional regulation. Further correlations were found between UTRs and their opposite antisense segments: More antisense transcripts overlapped the 3' UTRs than the 5' UTRs; also, UTRs that had overlapping antisense transcripts were longer than UTRs that did not (table 2-2).

Table 2-1. Selected GO categories found overrepresented among the 355 genes opposite filtered nonannotated antisense segments

GO term	Ng	Nobs	Nexp	Odds ratio	P
Cell wall	95	17	5.1	3.3	2 x 10–6
M phase of meiotic cell cycle	127	21	6.8	3.1	9 x 10–7
Transcriptional activator activity	33	9	1.8	5.1	5 x 10–6
Transcriptional repressor activity	23	6	1.2	4.8	1 x 10–4
Monosaccharide transporter activity	19	5	1	4.9	3 x 10–4

*N*g, number of genes in the genome annotated to this category; *N*obs, number of genes observed in this category that were opposite an antisense segment; *N*exp, number of genes expected if genes opposite antisense segments are randomly distributed over GO categories; *P,* hypergeometric test *P* value.

Table 2-2. Association of UTR lengths with presence of antisense transcript, and the 3'/5' bias in position of antisense transcripts

	Antisense		Control, no antisense
	Filtered	All	
Number of 3' overlaps	145	783	NA
Number of 5' overlaps	94	355	NA
Median length of 3' UTR	111 (49)	104 (408)	87.5 (1,588)
(no. of genes)	$P = 0.05$	$P = 0.00001$	
Median length of 5' UTR	89 (28)	82 (142)	67 (1,588)
(no. of genes)	$P = 0.08$	$P = 0.003$	

Overlap was measured with respect to the start and stop codons. Significance was calculated by comparing the length distributions of UTRs with antisense to controls where UTRs had no antisense partner by using the two-sided Wilcoxon test. NA, not applicable

Most ncRNAs previously reported as novel have since been annotated in SGD, and hence do not overlap with our expressed, nonannotated segments (McCutcheon and Eddy 2003; Olivas *et al.* 1997). We compared our data to transcriptome surveys, carried out by using serial analysis of gene expression (SAGE) (Velculescu *et al.* 1997) and ESTs (Graber *et al.* 1999). Thirteen percent of the nonannotated isolated and 42% of the nonannotated antisense transcripts were represented by SAGE tags. For the EST data, these numbers were 1% and 6%, respectively. Analysis of SAGE tags on microarrays described a number of novel transcripts in a mutant strain defective in the RNA degradation pathway (Wyers *et al.* 2005); however, the eight primary examples were not found expressed in our study of wild-type yeast.

II. Profiling periodic transcription of the cell cycle – regulated genes.

1. Introduction

Mitotic cell cycle is a well characterized system in yeast and therefore is an attractive model for the study of the genome-wide RNA-mediated regulation of gene activity in yeast. It is associated with important

physiological changes in the cell and diverse biological events depend on this periodicity (Cho *et al.* 1998). To ensure the proper functioning of the mechanisms that maintain order during cell division about 800 genes of diverse GO categories are coordinately regulated in a periodic manner coincident with the cell cycle (Spellman *et al.* 1998). In particular, mRNA fluctuations during cell-cycle were observed for genes involved in DNA replication, budding, glycosylation, nuclear division, control of mRNA transcription (Oehlen and Cross 1994; Wittenberg *et al.* 1990), responsiveness to external stimuli (Oehlen *et al.* 1996; Zanolari and Riezman 1991) and subcellular localization of proteins (Scully *et al.* 1997). Cell cycle-regulatory proteins are required for normal DNA repair (Weinert 1997), meiosis (Jang *et al.* 1995; Verlhac *et al.* 1996) and multicellular development (Dong *et al.* 1997; Thomas *et al.* 1997). There is evidence for existence of local chromosomal regulation of genes induced at the same stage of the cell cycle (Cho *et al.* 1998; Cohen *et al.* 2000).

Much of the literature has focused on the posttranscriptional mechanisms that control the basic timing of the cell cycle. However, there is also clear evidence that *trans*-acting factors play a critical role in the regulation of the abundance of many cell cycle– regulated transcripts. Several genome-wide studies have been done to catalogue cell cycle-regulated genes (Cho *et al.* 1998; de Lichtenberg *et al.* 2005b; Pramila *et al.* 2006; Spellman *et al.* 1998). Identification of all of the genes in a genome that are coordinately regulated during the cell cycle provided a basis for comparison of gene activity over time and made it possible to search statistically for regulatory sequences. Given a high resolution of our technique, profiling genome-wide periodic expression with the tiling arrays allows taking a step forward to prove the existence and perhaps decipher the mechanism of RNA-mediated regulation of transcription. These results also provide an opportunity to identify related genes in the human genome that may be involved in cell cycle period-specific roles.

As a result of the present genome-wide characterization of periodic transcripts novel cell cycle-regulated sequences as well as periodic antisense transcripts were identified. I aimed to find more regulatory sequences, including promoter regions, previously overlooked by less comprehensive methods.

2. Transcriptional analysis of the mitotic cell cycle.

I used tiling arrays to undertake a detailed genome-wide study to characterize transcription during cell cycle. I quantified mRNA transcript levels at regular intervals in yeast cells undergoing mitosis after being synchronized.

3. Synchronization methods

Two independent methods – temperature sensitive cdc28 mutant – based synchronization and alpha-factor growth arrest, were used to obtain synchronized culture. This excluded confounding effects on transcription caused by heat shock or activation of the mating cascade, respectively (Cooper and Shedden 2003; Shedden and Cooper 2002).

3.1 Synchronization of temperature – sensitive cdc28 mutant strain

Various cyclin-dependent kinases are activated throughout the cell cycle to regulate progression through its stages. Their temperature-sensitive mutant alleles are classically used for synchronization experiments. A good degree of synchrony has been observed for temperature–sensitive mutant cdc28-13 allele in the W303 background strain (Cho *et al.* 1998), which has been therefore chosen for our studies. Arrest occurs at 37-38 ^{0}C and the culture is released by decreasing the temperature back to 25 ^{0}C. Synchronization can only be maintained for two to three cell cycles due to the asymmetry of the budding (Laabs *et al.* 2003) and natural variation of the doubling time of the budding cell (Bean *et al.* 2006).
I performed several pilot synchronization experiments to optimize the temperature and the duration of the arrest as well as to ensure smooth re-entry into mitosis at an appropriate cell density, which allows for continuous growth in exponential mode.
After re-initiation of cell division following the arrest, samples were collected every 5 minutes for 215 minutes (equals to ~3 cell cycles).

3.2 Synchronization by alpha-factor arrest

α-Factor is a mating pheromone that is secreted by haploid *S. cerevisiae* cells of the α mating type. It facilitates mating by binding to a seven-transmembrane domain receptor (Ste2) of haploid a cells and by inducing a cascade of events involving a heterotrimeric G protein activated MAPK signaling pathway. MAPK cascade entails block of cell division in G1 and induces mating-specific gene expression, production of agglutinin on the cell surface (Hagiya *et al.* 1977; Yamaguchi *et al.* 1994), and formation of projections which appear to act as copulation tubes (Lipke *et al.* 1976; Tkacz and MacKay 1979), known as “shmoos” (Fig. 2-6). This pheromone-mediated arrest is a useful tool in cell cycle studies because it produces a tight arrest, from which cells have evolved to recover rapidly and efficiently.

There are two general mechanisms of recovery from alpha-factor mediated arrest. At a low cell density recovery is achieved by desensitization of the G-protein coupled receptor (GPCR) (Catt *et al.* 1979). At a higher cell density MATa cells inactivate α factor by cleaving it with a cell-associated protease BAR1 (Ciejek and Thorner 1979; Finkelstein and Strausberg 1979; Maness and Edelman 1978; Moore 1983).
A fairly stable arrest can be obtained with *barl* cells with much less α-factor. However these protease deficient strains release less synchronously from the arrest than do wild type cells, suggesting that degradation of α-factor is critical for rapid recovery. Moreover bar1 cells exposed to α-factor for an extended period of time develop growth debility. Hence, if *barl* cells are used, protease should be added during the recovery phase. Generally a crude preparation of the Bar1 protease is generated from **"a"** cells and used to degrade the residual α-factor. An alternative is to use a preconditioned medium, that is, medium into which the Bar1 protease has been secreted by growing cells (Breeden 1997).

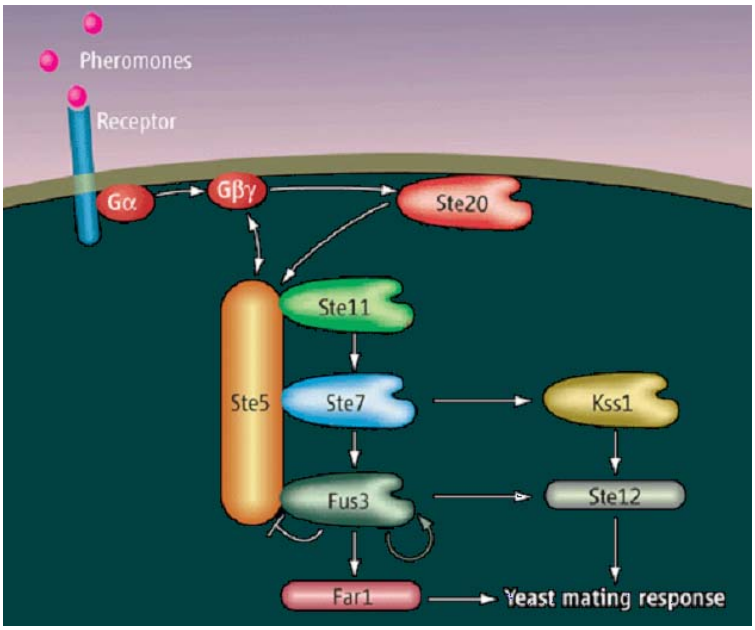


Figure 2-6. Pheromone response pathway. In response to ligand-induced activation of a pheromone receptor, activated G protein subunits in the membrane recruit and oligomerize Ste5, which in conjunction with the PAK-like kinase Ste20, activates the MAPK module. Once activated, the MAPKs Fus3 and Kss1 phosphorylate a variety of proteins that effect the pheromone response, including the transcription factor Ste12 and the polarization factor and cyclin-dependent kinase inhibitor Far1. Not all signaling components or substrates are shown.

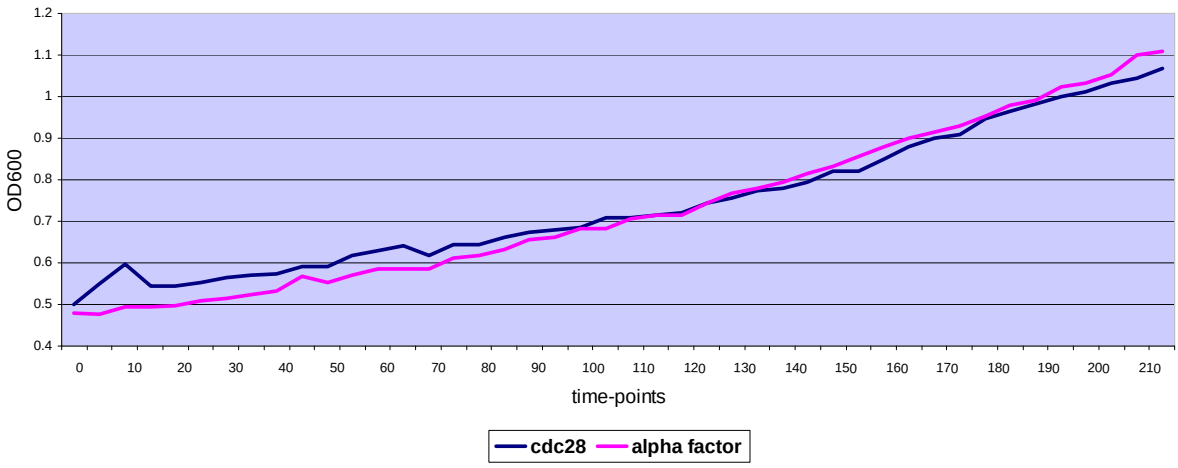
I have performed a series of pilot experiments to determine the optimal concentration of α-factor and the best conditions of release, which would ensure smooth re-initiation of cell division.

After 2 hours of pheromone-induced arrest in G1, *bar1* strain DBY2487 was recovered in fresh preconditioned medium to facilitate initiation of mitosis and samples were collected every 5 minutes for 220 minutes (equals to ~3.5 cell cycles).

As an additional quality control step cell densities were measured for each sample of both experiments to ensure exponential growth of the whole culture throughout the time-courses (Fig. 2-7a). Cells exhibit more than 98% synchrony upon release from arrest as determined by morphology of budded cells and nuclear position (Fig. 2-7b).

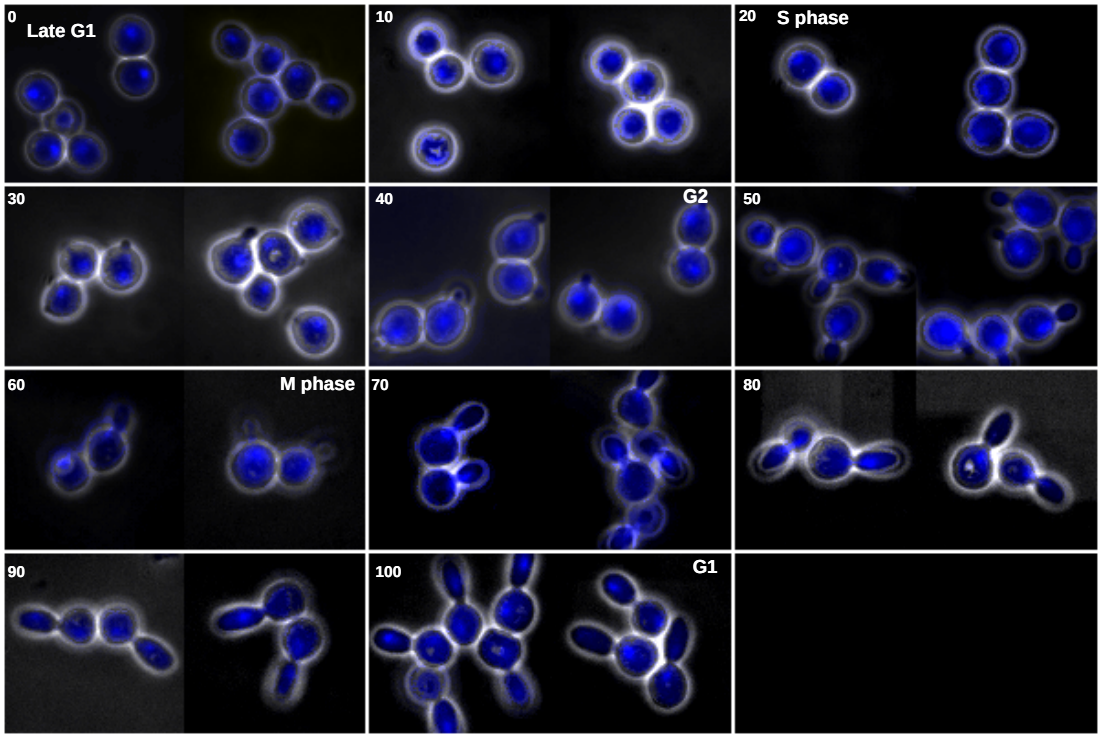
PolyA-enriched RNA fractions were obtained from the culture corresponding to each time-point in both synchronization experiments, reverse transcribed to cDNA with random hexamers and hybridized to tiling arrays.

Figure 2-7. (a) cell density during time-course growth after release.

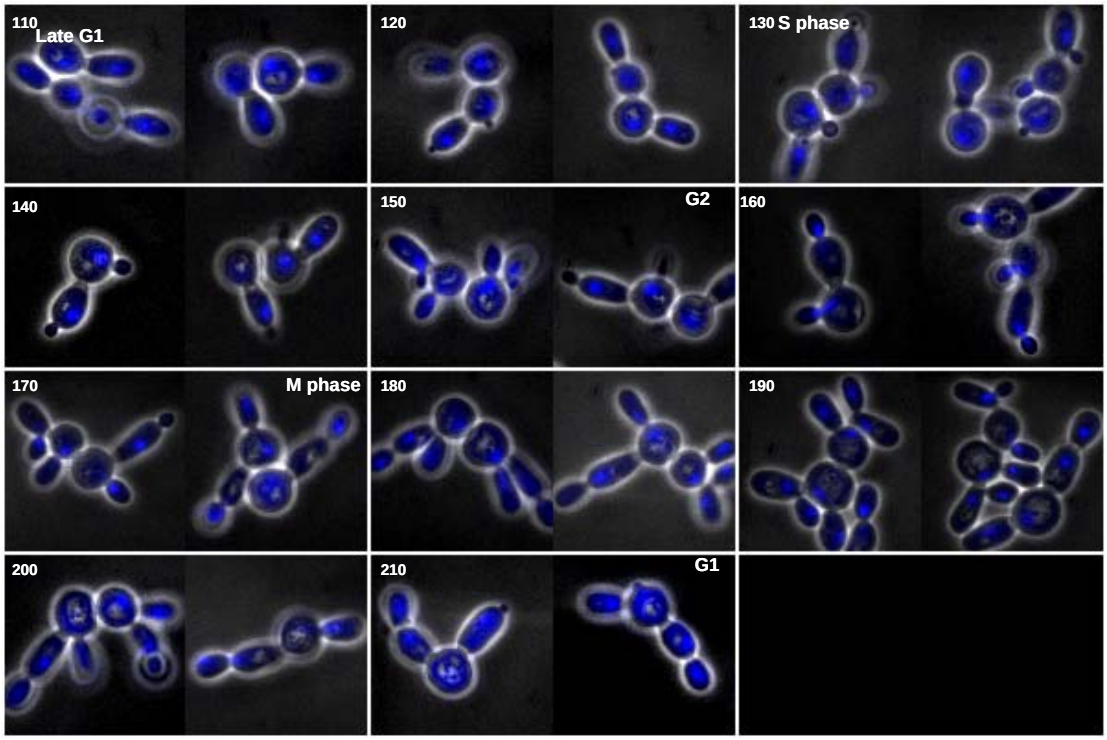


(b). Electron microscopy of nuclear position during mitotic progression. Images are presented in 10 min resolution to make bud size changes and nuclear position apparent. Nucleus appears as a bright blus spot migrating closer to the bud site with progression through the cell cycle phases.

1st cell cycle



2nd cell cycle



To account for different probe affinities the hybridization signal data was normalized to W303 and S96 genomic DNA hybridizations, respectively.

4. Protocol optimization

Five points of concern in processing of a hybridization sample have put the suitability of the available ready-to-use protocol from our previous paper in question.

1. Given the large size of sample library per time-course and impossibility to repeat hybridization sample processing for any selected time-point due to inbuilt absence of technical replicas in the experiment, particular attention had to be paid to consistency in probe preparation throughout the whole collection, such that the yield and the quality of cDNA probes are in the same range.
2. Another point of concern were low yields of total RNA material from initial samples immediately recovering from arrest and the ones with $OD_{600} < 0.7$, which corresponds to ~1 complete cell cycle of both time-course experiments. Low total RNA yields entailed low yields of polyA-RNA, which in turn required optimization of reverse transcription (RT) reaction to generate enough of cDNA for subsequent hybridization to the arrays.
3. The former issue was aggravated by addition of actinomycin D (ActD) into the RT reaction. Reverse transcriptase can generate spurious second-strand cDNAs in the course of reaction, which will hybridize to their complementary strand and appear as false positive antisense transcripts (Gubler 1987a; b; Johnson *et al.* 2005). ActD selectively blocks second-strand cDNA synthesis in the RT reaction due to its specific inhibition of DNA-dependent, but not RNA-dependent DNA-synthesis (Muller *et al.* 1971; Ruprecht *et al.* 1973). Most likely ActD acts through binding deoxyguanosine residues in single- and double-stranded DNA, preventing either the annealing of DNA during priming or elongation in DNA-dependent DNA synthesis (Goldberg *et al.* 1962; 1963). RT reaction in the presence of ActD will synthesizes cDNA, which is free of antisense artifacts (Perocchi *et al.* 2007), but the overall yield of cDNA material will decrease.
4. Consistency in yield and quality of cDNA samples could not be guaranteed by standard phenol:chlorophorm sample purification after RT reaction, followed by phase separation on phase lock gel (PLG) columns and cDNA recovery by ethanol precipitation, since sometimes cDNA samples were lost during passage of RT reaction mix through the PLG columns or cDNA sample was not fully recovered after its precipitation from aqueous phase collected from the column. Due

to minimal quantities of material, significant variation in the yields of precipitated cDNA may occur from sample to sample.

5. Finally, since the aim of my study was to register fine changes in expression intensity of a given transcript over time and deduce periodic behavior it was essential that the amount of final material hybridized to the chip stays exactly the same for all time-points, such that we register *bona fide* fluctuations in transcriptional levels, but not experimental noise of hybridization intensity. With DNase I, used for fragmentation in conventional protocols, it is only possible to control the amount of input cDNA, but the output of fragmented material may vary due to inherent endonuclease activity of the enzyme. Moreover, to obtain the required size range (25-75 bps) of cDNA fragments, the duration of the digest reaction is chosen empirically as a result of several trial digests. Hence, there should be enough of cDNA material to perform trial digests in order to determine the exact time of digestion, which is lacking as explained in points 2 and 3. Even more alerting is the risk to over-digest the sample to very small fragments not suitable for hybridization or even short nucleotide stretches, thus loosing a time-point completely without a possibility to synthesize a new cDNA for it, as stated in point 1.

Considering all the points above, the existing protocol was not suitable for carrying out the cell cycle study and after a series of optimizations I have significantly modified it to obtain better yield, quality and overall consistency in sample preparation.

In particular, decreased amounts of polyA-RNA (~6 ug) together with increased amounts of dNTPs (0.4 mM) in comparison to the previous protocol improved cDNA yield in the RT reaction in the presence of ActD. Instead of conventional phenol:chlorophorm purification the RT reaction mixtures were purified on Affymetrix cDNA Clean-up Columns and eluted with pre-warmed water to further increase the yields of cDNA. Basic improvement and a high degree of consistency in sample yield, purity and quality was observed after purification on Affymetrix columns as controlled by lab chip electrophoresis on the Agilent BioAnalyzer. Finally, endonuclease activity - based fragmentation was substituted with restriction – like reaction (http://www.affymetrix.com/support/technical/index.affx). For this I used a mixture of enzymes classically involved in base-excision repair pathway - Uracil-DNA glycosylase (UDGase) and Apyrimidinic / Apurinic endonuclease (APE). UDG catalyses the first step in this pathway, hydrolysing the N-glycosidic bond connecting the base to the deoxyribose sugar. The enzyme is exquisitely specific for the removal of uracil, but no other base, from DNA, but not from RNA. Abasic sugar is subsequently removed by APE (Fig. 2-8). This digestion method allows for strict control of the

amount of final output cDNA from the reaction, which ultimately equals the input cDNA quantity, and has no risk of over-digestion of cDNA fragments. Digested cDNA fragments are strictly consistent in quantity and size range after digestion. To incorporate uracil into cDNA, which determines exact cleavage sites, the RT reaction was carried out in the presence of dUTP in the dNTP mixture. Exact titration of the dUTP:dTTP ratio in the dNTPs mix was essential to obtain fragments in the desired size range and required additional thorough refinement.

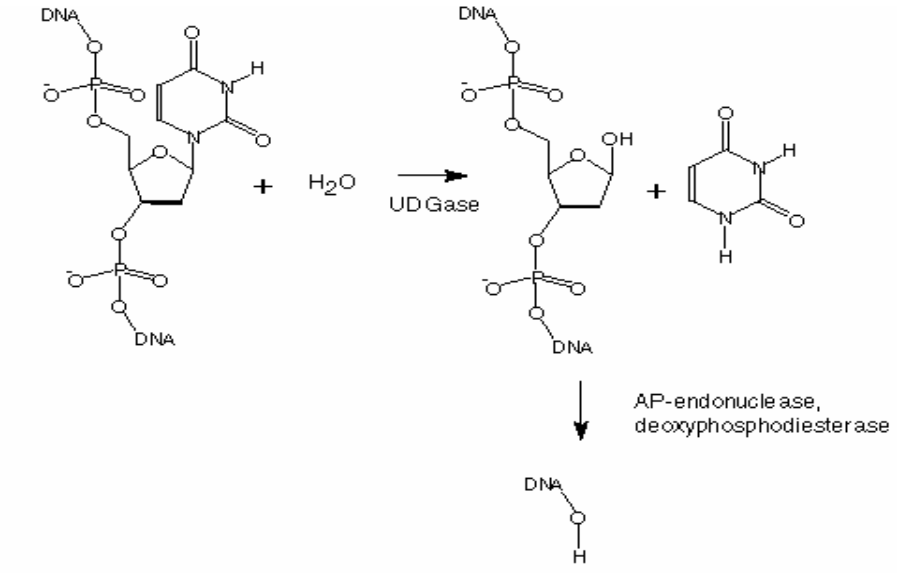


Figure 2-8. 2 steps of DNA fragmentation. UDGase performs apurinization of DNA and APE – deoxyphosphodiesterization.

Data analysis

5. Quality control, normalization and segmentation.

Together with Matt Ritchie from CRI (UK) we have performed computational analysis of the data, including normalization of cDNA signal, quality control assessment and segmentation along genomic coordinates. It was mostly done as described for the whole-genome transcriptome study in rich media (David *et al.* 2006) in chapter 2.1. Boxplots of normalized PM intensities for each synchronization experiment are shown in (Fig. 2-9).

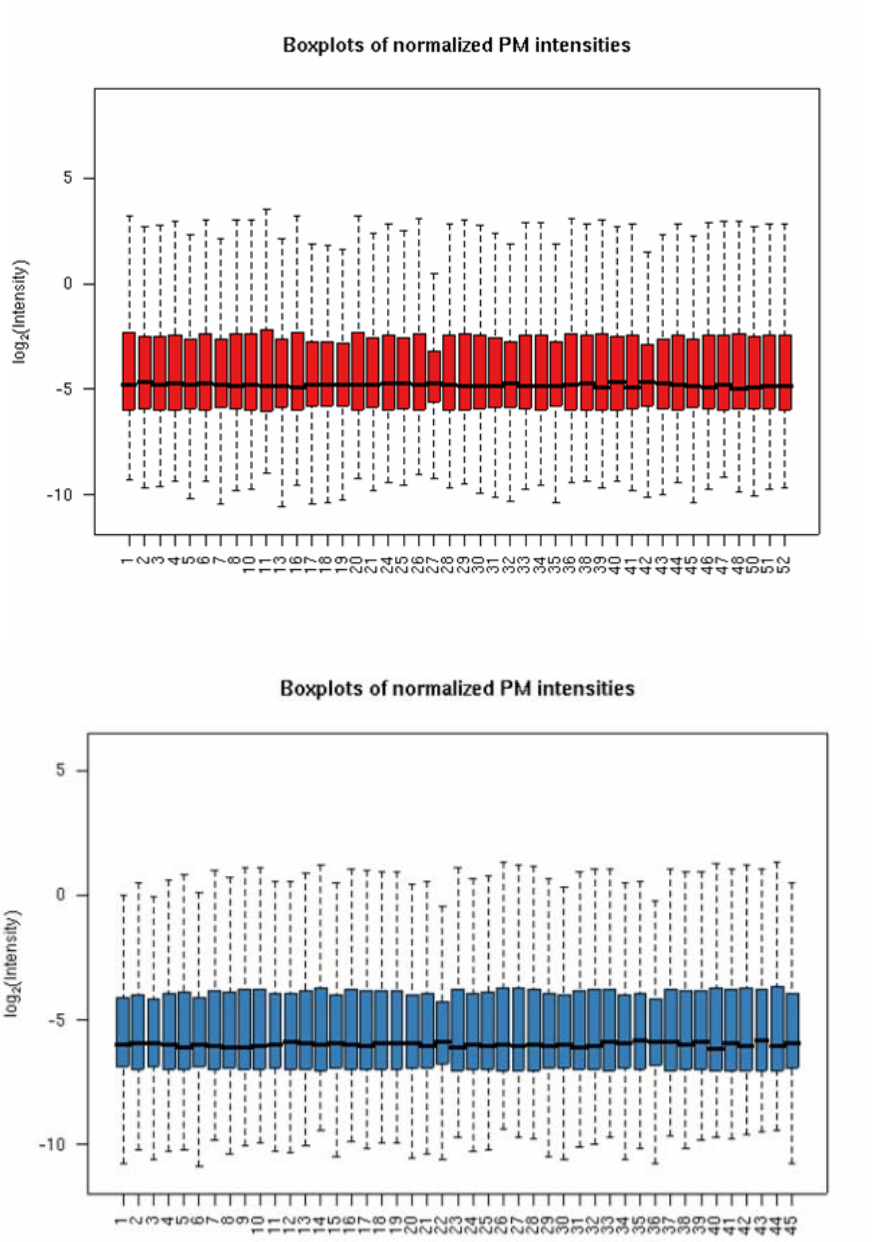
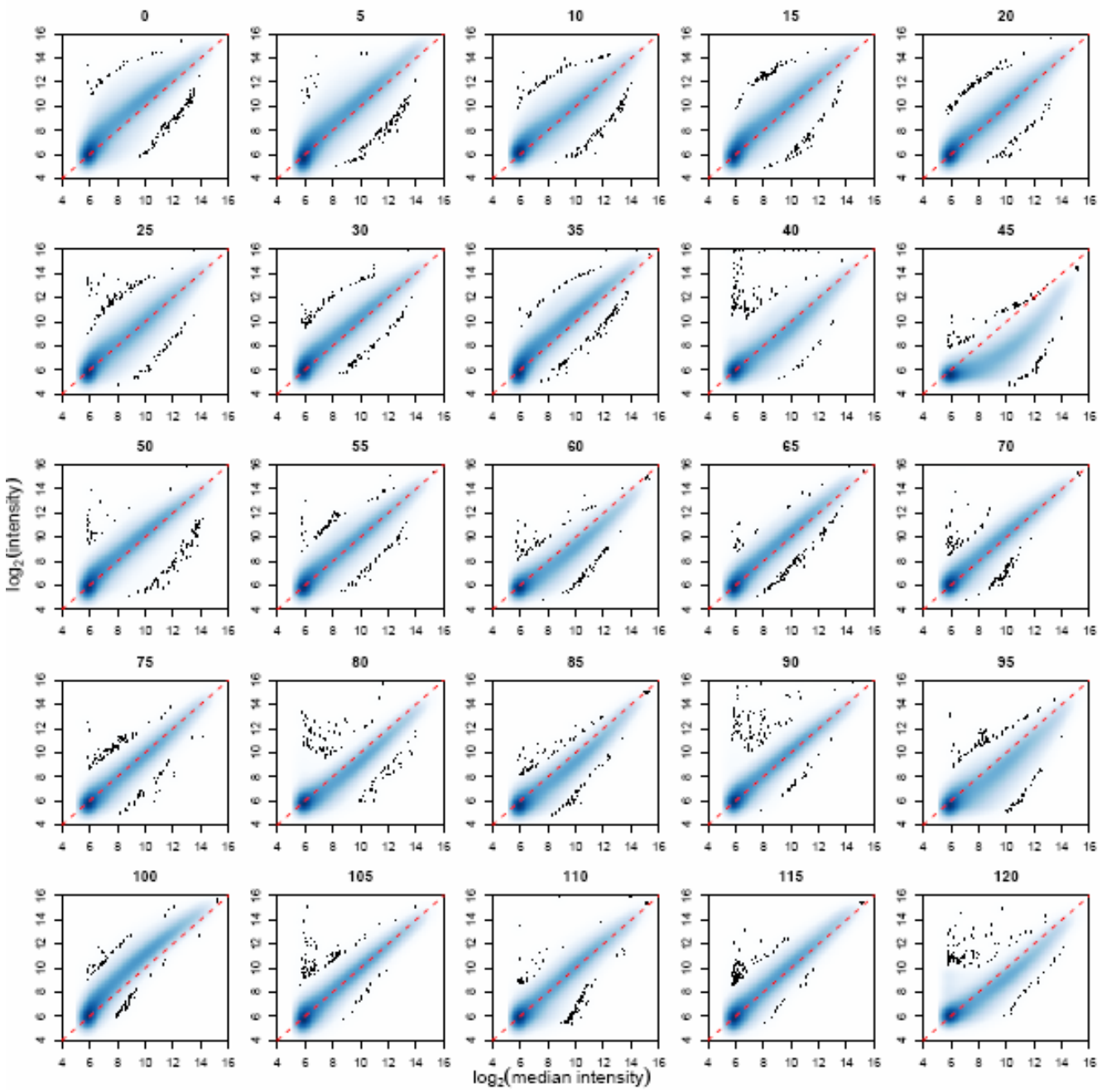
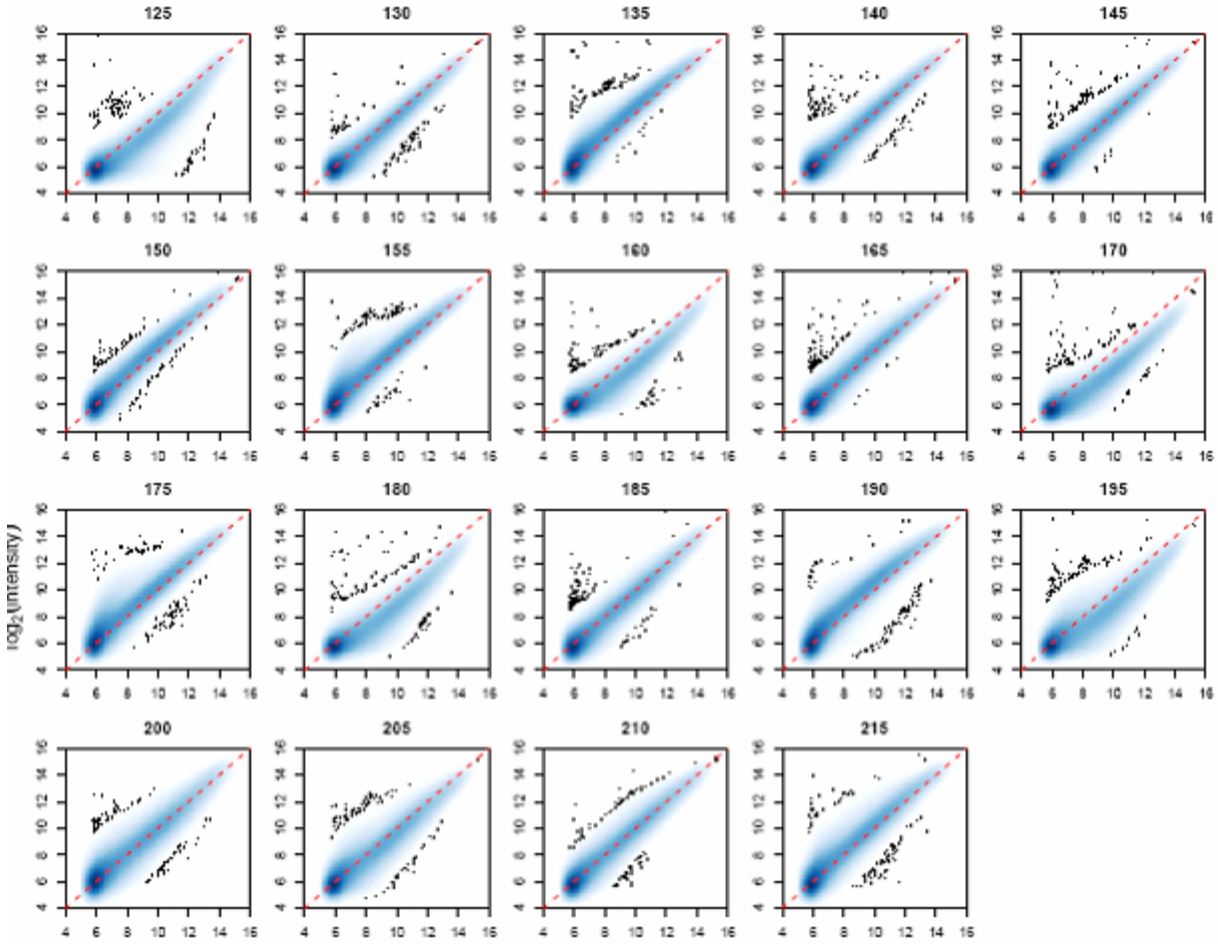


Figure 2-9 (a) Boxplots of PM intensities for cdc28 (top panel) and alpha factor (bottom panel) time-course. X-axes – log2 of sugnal intensity; Y-axes – subsequent time-points hybridizations. Horizontal black line indicates median signal intensity for each chip.

Figure 2-9 (b) The smoothed scatter plots. Plots show the median intensity (across all arrays for a given data set) on the x-axis, versus the intensity from a given array on the y-axis. The intensities are on the log-2 scale, and only the PM probes are used in the plot. The darker the shade of blue, the more points there are in that region of the plot. The bunch of points at the bottom left is probes with the lowest intensities.





Standard segmentation algorithm applied in our previous study performed well on non-periodic transcripts; however, a certain number of less abundant periodic transcripts were underrepresented. We have increased the default number of segments in the algorithm, which improved the overall sensitivity of transcript detection. The drawback of this approach was that some features were divided into several smaller segments with slightly different expression intensity levels, however, it didn't influence the overall feature categorization process and / or calculation of average expression levels of a given feature (see data analysis). Segmentation was performed according to the model: y_ij = mu_j+e_ij, where i is the probe index and j is index for the array. This is a slight variant of the model used in the PNAS (David *et al.* 2006) and explained in (Huber *et al.* 2006), where there is an extra subscript for array to explain changes in level over time. Additionally, the data was visually inspected to ensure comprehensive detection of all expressed sequences.

The output of segmentation is a table of all identified transcripts, their start and end coordinates, and their average transcription level observed over time.

Applying the same criteria as in our transcriptome study (David *et al.* 2006) identified transcripts were categorized into those that overlap existing annotation, novel isolated and antisense (Fig. 2-10, table 2-3). Profiles for all genomic regions are provided in a database that is searchable by gene symbol or chromosomal coordinate (http://www.ebi.ac.uk/huber-srv/cgi-bin/scercycle.pl).

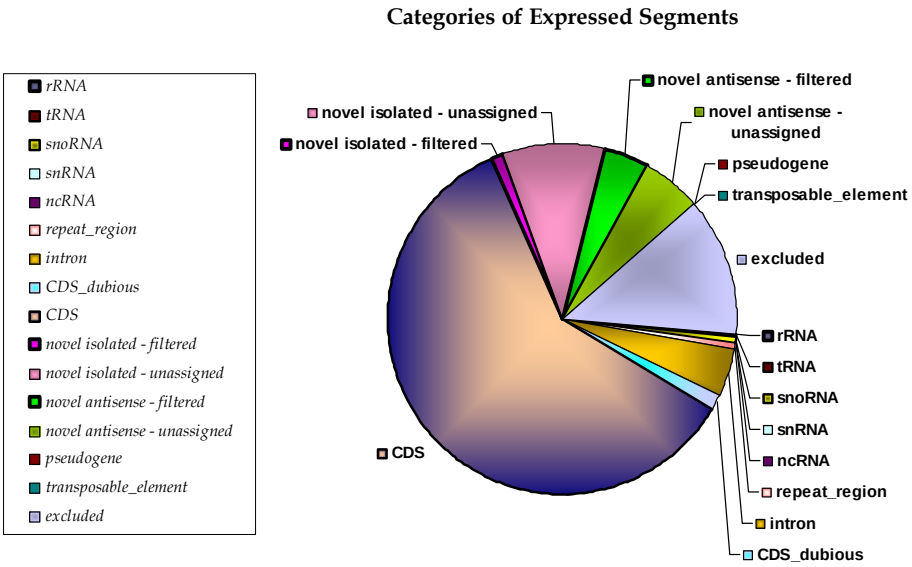


Figure 2-10: Number of expressed segments in each category.

Table 2-3: Number of expressed segments in each category

category	number of segments
rRNA	16
tRNA	20
snoRNA	60
snRNA	4
ncRNA	6
repeat_region	72
intron	552
CDS_dubious	210

category	number of segments
CDS	7645
novel isolated - filtered	135
novel isolated - unassigned	1226
novel antisense - filtered	523
novel antisense - unassigned	695
pseudogene	7
transposable_element	8
excluded	1639

6. Detecting periodic transcripts.

The major objective of transcription profiling in mitosis was to identify periodic genes and other cell cycle – regulated transcripts. First, to forestall computational analysis we have plotted the expression levels and calculated autocorrelation function for known periodic genes, peaking in G1 (Fig. 2-11). Apparent cycling of this selected set of genes has served an additional verification of the quality of my data and set us forward for in depth analyses.

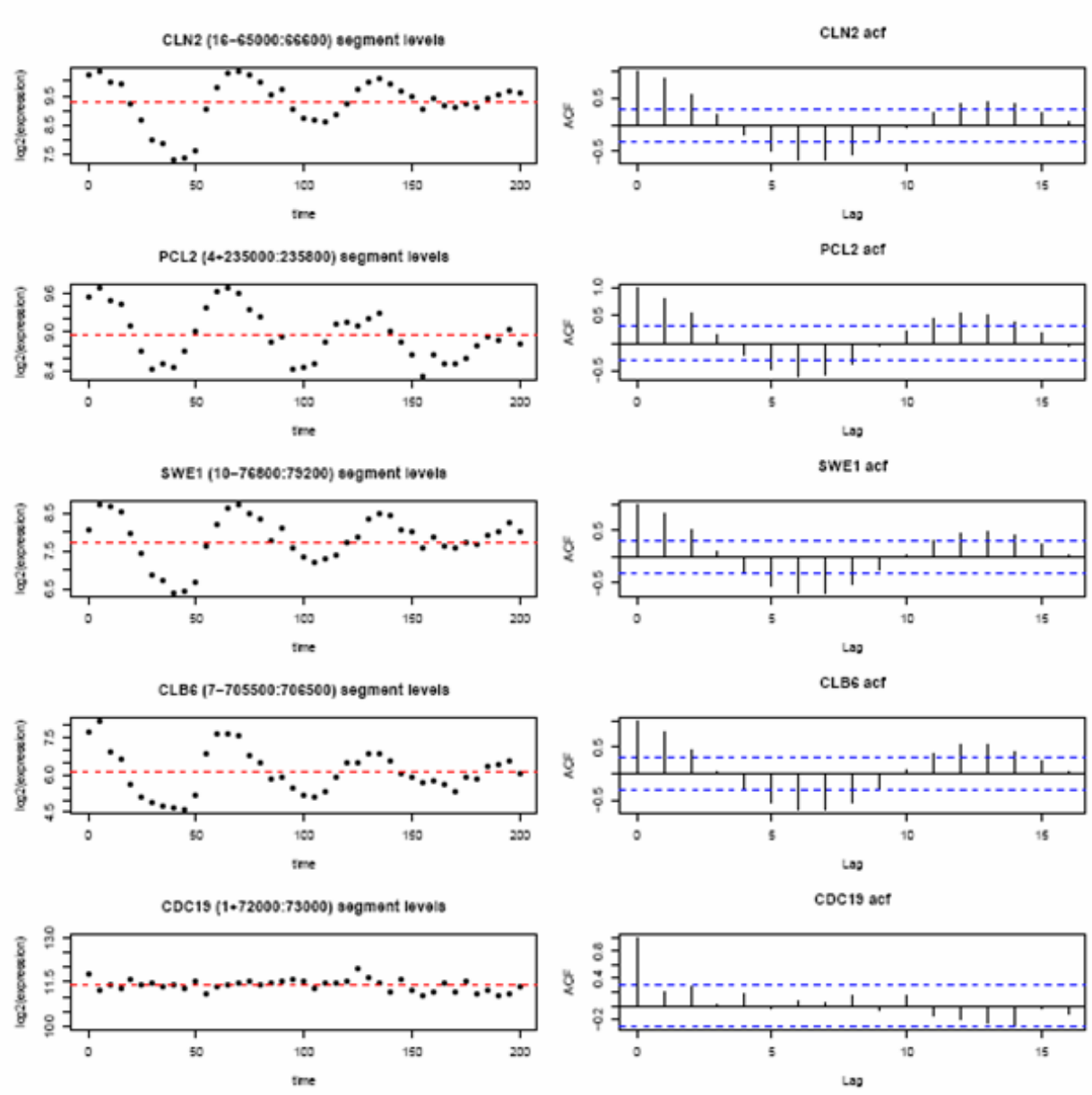


Figure 2-11. The autocorrelation function (ACF) for selected cycling genes. ACF was generated for alpha factor dataset used for the quality control of data. ACF describes the correlation between the process at different points in time. Here ACF clearly shows periodic pattern fitting a sin curve.

The average segment levels from each time-point were analyzed for periodic expression patterns applyong two different methods.

First, the data from the cdc28 and alpha factor time-course were analyzed separately using the approach of Ahdesmaki et al (Ahdesmaki *et al.* 2005), which calculates p-values for a robust nonparametric version of Fisher's g-test (Ueda *et al.* 2002; Wichert *et al.* 2004). Fisher's g-test calculates the p-value(s) and estimates nonrandomness of the dominating frequency in the periodogram from the signal-to-noise ratio (Ptitsyn *et al.* 2006). This test is useful to detect hidden periodicities of unknown frequency in a data set and is widely applied to analysis of microarray data (Wichert *et al.* 2004).

For each segment, the test was carried out at the frequencies corresponding to the estimated cycle time from each experiment (90 minutes for cdc28 and 65 minutes for alpha factor). The method of Ahdesmaki et al. has the benefit of being insensitive to outliers in the time series. Outliers frequently arise in microarray experiments due to technical artefacts affecting particular arrays or for other biological reasons. For example early time-points may be outliers for transcripts affected by heat shock response in the cdc28 experiment. Similarly, a lack of sychronisation at early or late time-points may produce outliers in both experiments. The existence of outliers in our data set led us to apply this method in the hope that fewer false positive periodically expressed segments, driven by outliers, would be identified.

This testing procedure is implemented in the R package GeneCycle:
(http://www.strimmerlab.org/software/genecycle/).

Next Ahdesmaki et al.'s method was applied to the combined cdc28 and alpha-factor dataset. Setting the arbitrary cut-off scores of $p<0.05$ for the low-confidence set and $p< 0.01$ for the high-confidence set we have detected a number of periodic transcripts in all three categories for verified ORFs (CDS), novel isolated genes and antisense transcripts (table 2-4). For convenience of terms, these lists for any particular category are further referred to as A0.05 and A0.01, where A stands for "Ahdesmaki et al".

Statistical analysis and comparison with available benchmark sets (de Lichtenberg *et al.* 2005b) have shown that we are able to identify at least 615 periodic genes. This is more than identified initially by Cho et al. (Cho *et al.* 1998) and a bit less that detected by Spellman et al., who applied 3 different synchronization approaches and incorporated the results obtained by Cho et al. (Spellman *et al.* 1998).

The second method was developed to tackle the inconsistencies in periodicity detection of previously applied computational approaches and applied for the re-analysis of all hitherto existing datasets (de Lichtenberg *et al.* 2005b). As discussed earlier in the introduction, it is a permutation based approach and

its score is composed of two terms - magnitude of regulation, and the periodicity of the expression profile. The combined score ensures that high-ranking genes are both significantly expressed and periodic. Based on comparisons with existing benchmark sets of known periodic genes, a high-confidence set was defined at top-300 ORFs, a medium-confidence at top-500, and the lower-confidence at top-800 set. For the highest scoring 300, 500 and 800 genes the estimated false discovery rate (FDR) is $3x10^{-6}$, $4x10^{-4}$ and $8x10^{-3}$, respectively.

Lars Juhl Jensen has applied this Fourier transform based algorithm to my combined cdc28 and alpha-factor dataset and the output of his analysis presents a table of all identified transcripts, phase and amplitude of cycling as well as periodicity and magnitude P-values, which ultimately determine the cut-off threshold for each category of transcripts. The length of the cell cycle was determined for each synchronization experiment and the peak time for each transcript was derived in percentage of the cell cycle duration (table 2-4).

Method	**Ahdesmaki**		**de Lichtenberg**		
p-value range	**p<0.05**	**p<0.01**	**top300**	**top500**	**top800**
CDS	1463	808	300	500	800
novel isolated filtered	27	12	10	14	29
novel antisense filtered	49	28	17	37	60

Table 2-4. Number of periodic segments identified by two statistical approaches, Ahdesmaki and de Lichtenberg.

Lars Juhl Jensen's analysis and comparison with available benchmark sets (de Lichtenberg *et al.* 2005b) identified at least 500 periodic genes with high confidence, which is more than discovered in the dataset by Cho (Cho *et al.* 1998) and about the same number identified by Pramila et al. (Pramila *et al.* 2006), and we are as sensitive in transcript detection as the groups that performed previous studies.

Both methods show a good degree of overlap for periodic genes with high-significance p-value (Fig. 2-12a). In particular, there is a 90% overlap between periodic ORFs with high-confidence p-value in top300 list and cell cycle – regulated genes in the A0.01 list, 75% of overlap between top500 and A0.01. Top800 genes show only 56% overlap with A0.01 list. Comparison with A0.05 was considered redundant given the low-significance of periodicity p-value for the genes identified in this dataset.

Among the comparison of A0.01 with top500 list, 125 ORFs (from top500) and 241 coding sequences (from A0.01) were manually validated for periodic pattern and a combined dataset was created (table **A1**). This set comprises 375 cycling ORFs, which overlap between top500 and A0.01, 95 cycling ORFs

from the 125 remaining non-matching coding sequences of top500 and 127 periodic genes from 241 ORFs of A0.01 set. We considered visual inspection to enrich the combined dataset a valid approach since we were operating within high- and medium-confidence datasets for both methods. The global pattern of periodic transcription of all 597 ORFs is summarized in figure 2-12b.

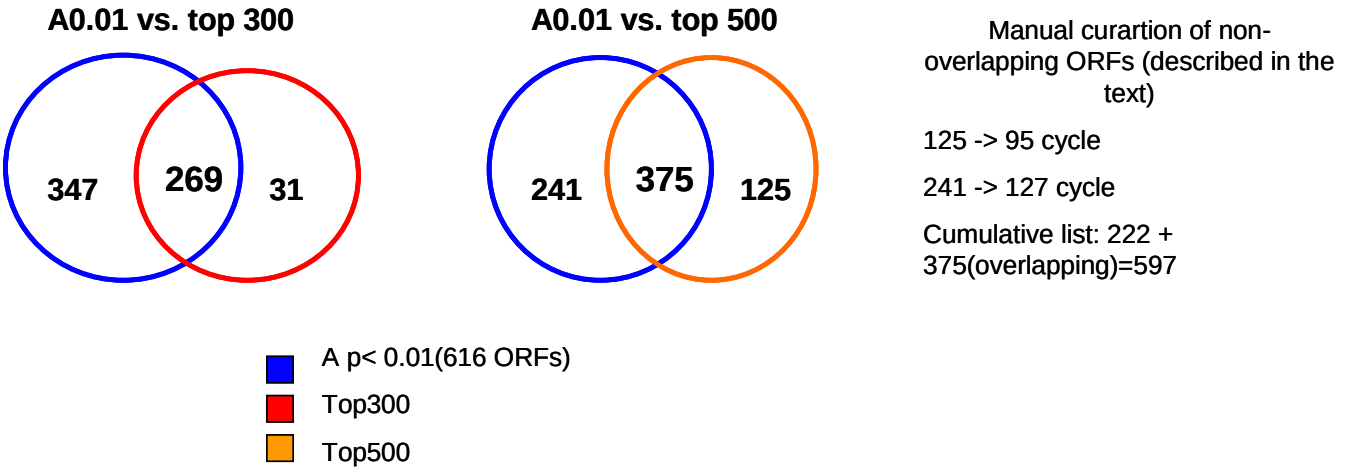


Figure 2-12a. Overlap between periodic genes identified by methods of Ahdesmaki (blue circle) and deLichtenberg (shades of red).

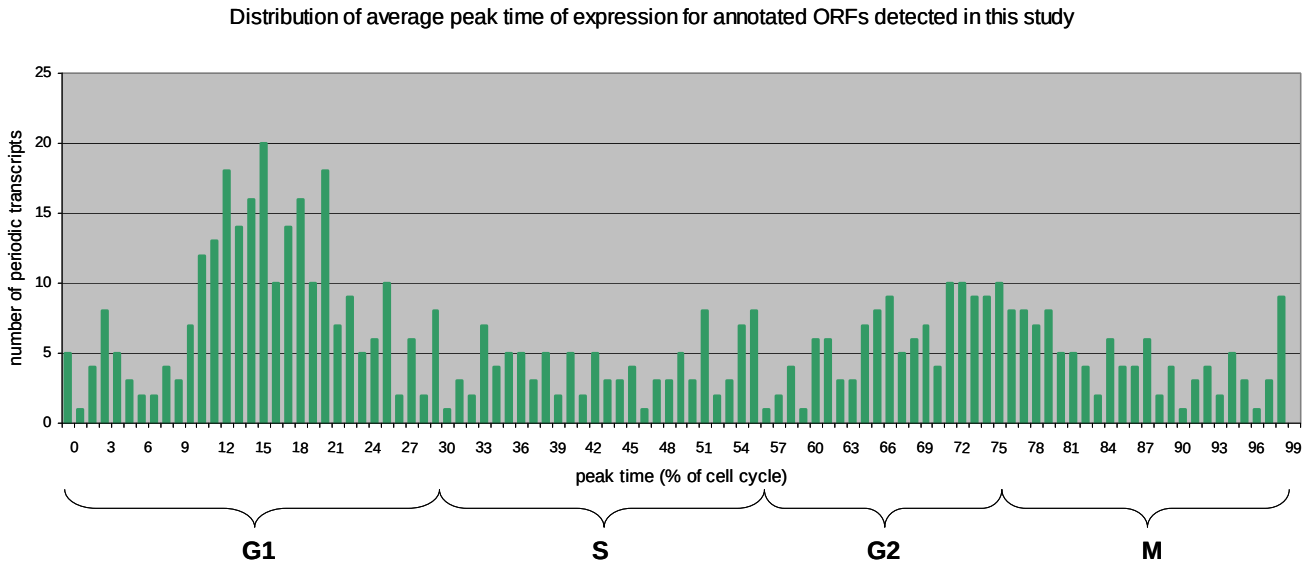


Figure 2-12 b. Distribution of peak time of periodic ORFs over time of cell cycle progression

Overall the method of de Lichtenberg et al. appears to be more specific, whereas the method of Ahdesmaki et al., which does not penalize magnitude of periodicity, is more sensitive allowing for detection of periodic genes in the context of high or low expression background of a given feature.

7. Enrichment for periodic ORFs

Previous studies have identified a large number of periodic transcripts. Although the periodic signal is strong in most data sets (Shedden and Cooper 2002; Wichert *et al.* 2004), the experimental noise is also considerable, as can be seen from the poor overlap between the gene sets identified as periodic in different experiments within budding yeast (de Lichtenberg *et al.* 2005b; Johansson *et al.* 2003; Luan and Li 2004; Shedden and Cooper 2002; Zhao *et al.* 2001). The agreement is extremely poor when different computational methods are applied to the same data. Combining several datasets, nearly 1800 different genes have been proposed to be periodic - that is almost every third gene in the *S.cerevisiae* genome. Recently previous studies were reconciled by re-normalizing the data using signal-dependent non-linear Qspline method (Workman *et al.* 2002) and employing permutation-based method developed by de Lichtenberg et al., which entailed "top-periodic" lists with different significance score cut-offs (de Lichtenberg *et al.* 2005b) as described above.

I have compared periodic genes identified by the de Lichtenberg method to benchmark sets employed by de Lichtenberg et al. Our study is only outperformed by Gauthier et al., who have combined all available analyses of cell cycle-regulated genes performed to date, including that of de Lichtenberg et al. (Gauthier *et al.* 2008). Adding the fraction of periodic genes discovered by Ahdesmaki approach we enrich for more periodic ORFs (Fig. 2-13).

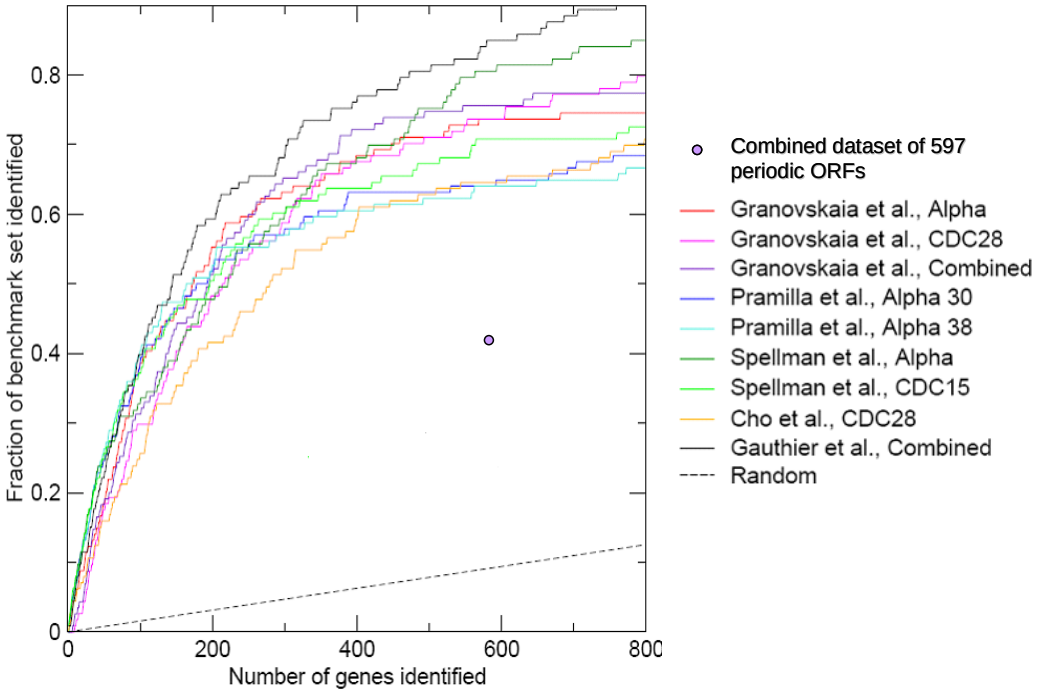


Figure 2-13. Specificity vs sensitivity ROC-like plot. Comparison of the results produced in our study (de Lichtenberg analysis method) with existing datasets. A point indicated 41% of B1+B2 benchmark set identified by all 597 periodic ORFs in our combined dataset.

I have enriched Spellman's list for 188 genes, which are enriched on the GO categories of genes involved in cell cycle regulated processes (table 2-5).

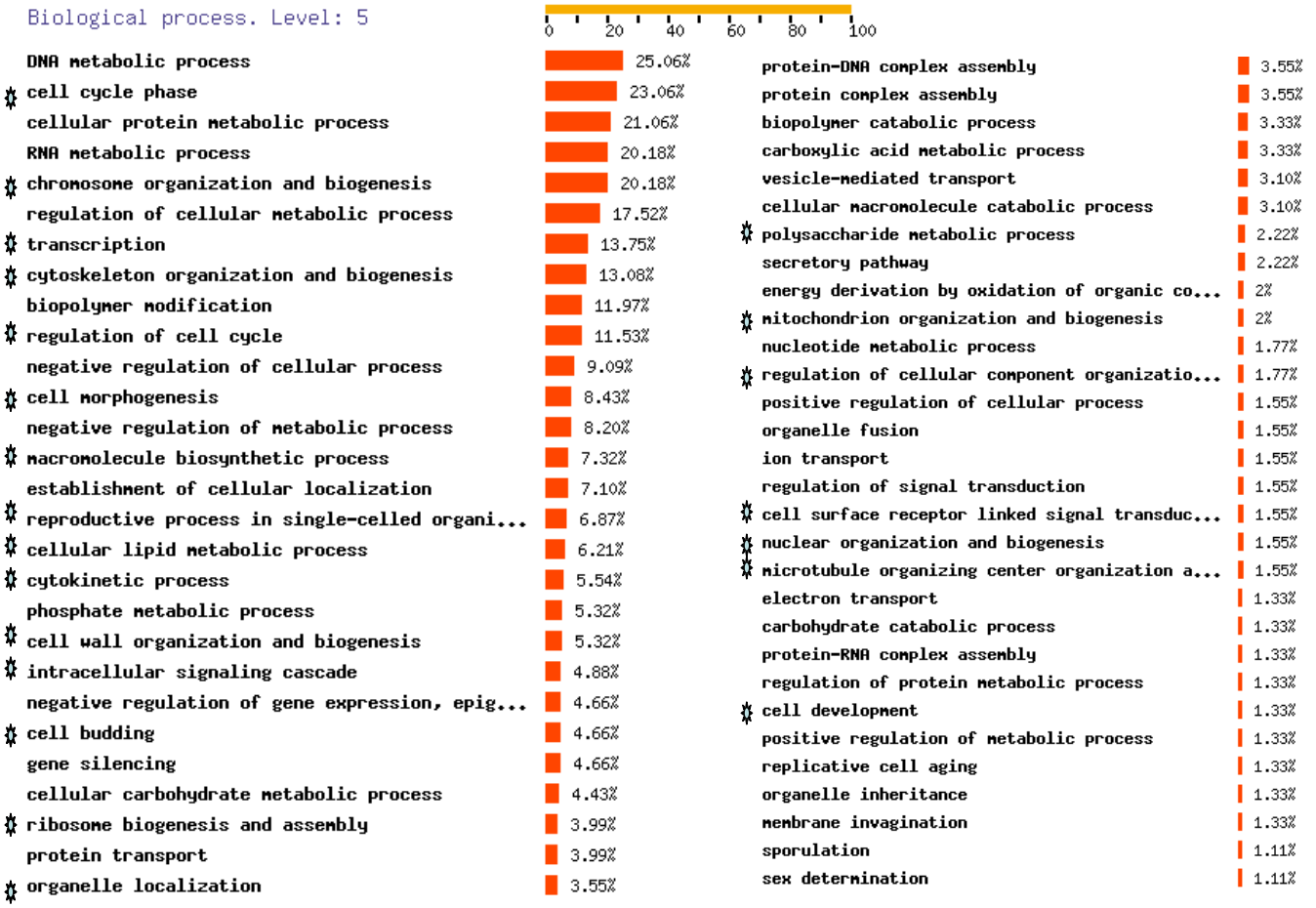


Table 2-5. GO categories. Biological processes in which 188 genes unique for our dataset are involved. The stars indicate direct cell cycle events.

Together with Lars Juhl Jensen we have taken a systematic look at whether splicing plays a role in regulation of cell cycle. For this he has matched 274 intron-containing ORFs annotated in SGD with 600 periodic genes identified by him in previous studies (de Lichtenberg *et al.* 2005b). His search identified seven intron-containing candidates, which may have a role in regulation of the cell cycle: RFA2 (a subunit of DNA replication factor A), TUB1 (alpha tubulin), CIN2 (a putative tubulin folding factor), MOB1 (a component of the mitotic exit network), ECM33 (a GPI-anchored protein that may be involved in bud growth), PMI40 (mannose-6-phosphate isomerase, i.e. cell wall biogenesis), and SRC1 (involved in sister chromatid segregation). The two new datasets presented an opportunity to systematically address the question of whether these and any other intron-containg genes appear to be cell cycle regulated through splicing. Surprisingly, no correlation was registered in splicing and periodicity. Introns either were not expressed at all or repeated the periodic expression pattern of exons.

8. Detection of periodic antisense transcripts.

As shown in Table 2-3, 523 transcripts fall into novel antisense filtered category. Interestingly, part of the antisense transcripts show cycling pattern along with their sense counterparts, suggesting their possible regulatory role in cell cycle – dependent expression.

In order to detect periodic antisense transcripts we have applied "de Lichtenberg" approach similarly as it was done for the CDSs. Conditional "top300", "top500" and "top800" lists were generated, in this case referring to the specific range of p-value cut-offs, which determined the end position of each list for cycling ORFs analysis, but not to the actual number of periodic antisense transcripts. In parallel, "Ahdesmaki" method yielded A0.01 and A0.05 lists of cycling antisense transcripts.

To ensure comprehensive analysis of periodic antisense transcripts all lists were scrutinized in detail. As for the ORFs, A0.01 was overlapped with "top-500" list for the cycling antisense transcripts. Subsequently 7 non-matching antisense transcripts from A0.01 were manually validated and 4 out of 7 transcripts, which displayed a certain visible degree of periodicity, were selected. Likewise, out of 13 left out antisense transcripts of the top-500 set, 8 displayed clear cycling. Both cycling small sets were added to the overlapping 28 periodic antisense transcripts, comprising 40 antisense features altogether. Some antisense segments overlap 2 or more sense features on the opposite strand. These segments were combined together, such that total number of sense features, for which periodic expression is registered on the opposite channel amounts to 37 ORFs (table **A**2). The distribution of antisense periodicity over complete cell cycle is shown in Fig. 2-14a. Notably, not all sense features opposite periodic antisense cycle and those, which do, are expressed in different cell cycle phases.

Periodic antisense transcripts overlap features, which fall into different GO categories, including regulation of meiosis, purine and pyrimidine biosynthesis, cell wall organization, etc., which suggests their possible involvement in the regulation of cell cycle progression. Twice as many antisense transcripts express anti-correlation in periodicity to their opposite cell cycle regulated genes as the ones cycling in-phase with their sense counterparts (10 and 5 transcripts, respectively); some do not show any correlation (Fig. 2-14b).

Remarkably, several important regulatory genes show expression on the antisense channel. I would like to emphasize several of the most interesting cases.

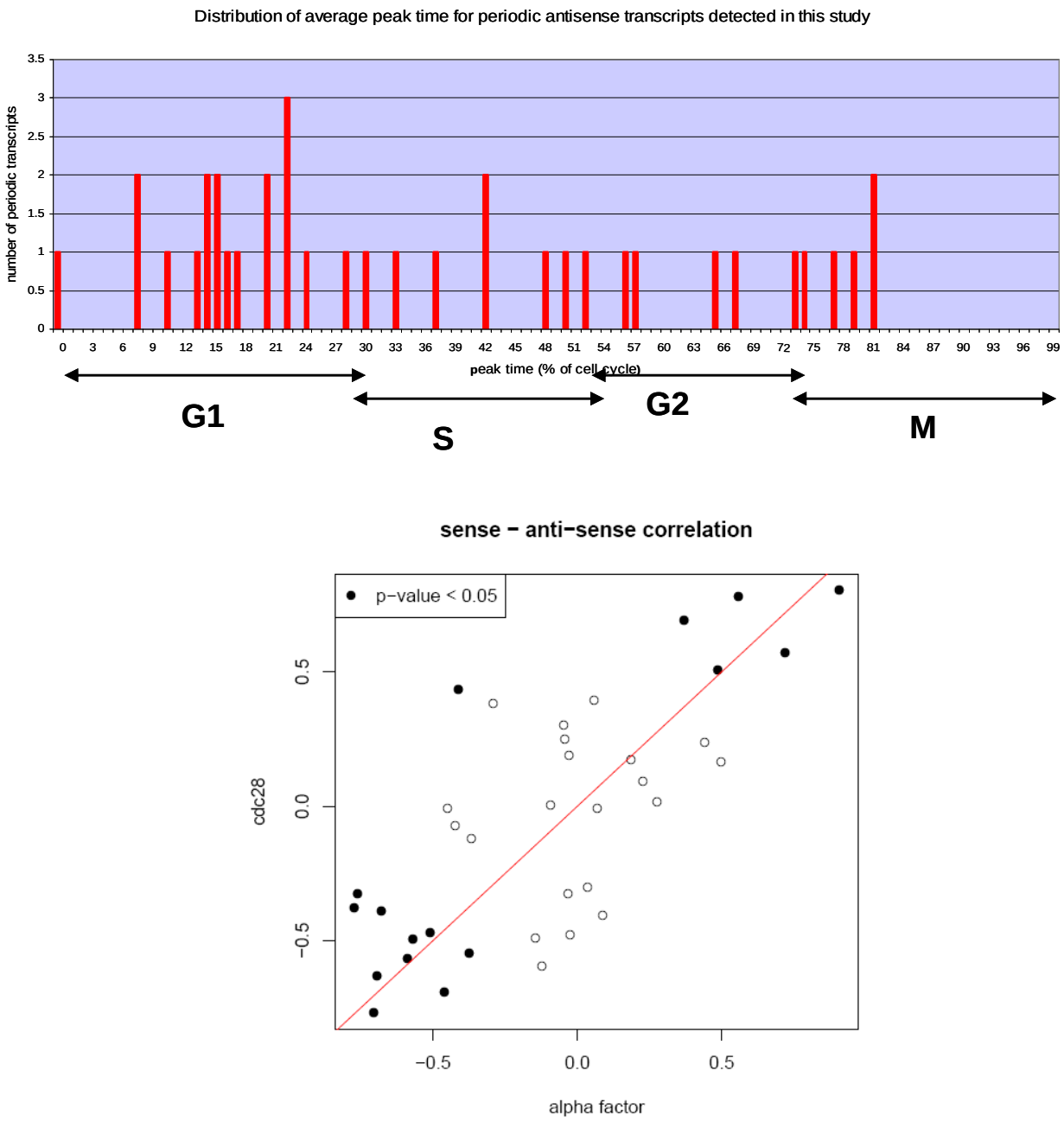


Figure 2-14 (a). Distribution of peak time of periodic antisense transcripts over time of cell cycle progression; (b) correlation between coherent and anti-correlated cycling in sense-antisense pairs.-0.5 and +0.5 representnegative and posititve correlation coefficients of sense and antisense phasing.10 SAPs at the bottom of the plot are anti-correlated, 5 pairs cycle in phase and 22 have no correlation.

FAR1

Far 1 (for "Factor arrest) is a bifunctional cyclin-dependent protein kinase inhibitor that is required to arrest cell cycle and establish cell polarity during yeast mating (http://stke.sciencemag.org/cgi/cm/stkecm) (Tyers and Futcher 1993). It shuttles between the nucleus and the cytosol. In the nucleus it mediates pheromone-imposed G1 arrest by directly inhibiting the cyclin-dependent kinase Cdc28 (Fig. 2-6). Far1$^-$ mutants are insensitive to arrest despite having an intact signal transduction pathway. Pheromone stimulation accelerates nuclear export of FAR1, and in turn FAR1 ferries polarity establishment proteins, such as Cdc24, into the cytosol. The interaction of FAR1 with G-protein betta - gamma subunits (Gβγ) Ste4/Ste18 localizes Cdc24 to the tip of the mating projection and serves as a landmark for orienting the cytoskeleton during polarized cell growth, presumably by mediating efficient and highly localized generation of the guanozine triphosphate (GTP) – bound state of Cdc42 (Butty *et al.* 1998; O'Shea and Herskowitz 2000). A null mutation in CLN2 gene, which codes for a G1 cyclin, reverses the effect of far1 null mutation: far1$^-$ cln2$^-$ strains arrest in response to α factor, suggesting that FAR1 directly inhibits Cln2. The latter property of FAR1 presumably contributes to regulation of cell cycle progression and ensures the block of G1 cyclin to make cells progress to S-phase. Expression of FAR1 is increased 4- to 5-fold upon α factor stimulation. After the arrest cells exhibit peaks of FAR1 in G1/S transition almost coherently with expression of three major cyclins driving this phase: Cln1, Cln2 and Cln3. Pronounced expression of FAR1 antisense counterpart is registered right after the peak of FAR1 expression (Fig. 2-15a, 2-16), suggesting transcriptional control of FAR1 functionality in the cell.

TAF2

Transcription initiation by RNA polymerase II involves the assembly of general transcription factors (GTFs) on the core promoter to form a preinitiation complex (PIC). The first step in PIC assembly is binding of the GTF TFIID to the TATA box. TFIID is a multi-subunit complex consisting of the TATA box binding protein (TBP) and a set of TBP-associated factors (TAFs) (Albright and Tjian 2000; Burley and Roeder 1996; Li *et al.* 2000). Like other components of the transcription machinery, TAFs are highly conserved from yeast to humans. In yeast, 14 TAFs have been identified, 13 of which are required for viability (Li *et al.* 2000). TAF2 is one of the 13 essential TBP – associated factors of transcription complex TFIID.

Promoters in yeast have different transcriptional requirements for TAFs and can be divided onto those whose transcription is dependent upon all or almost all TAFs (TAF-dependent), those whose transcription does not require any TAFs (TAF – independent), and those whose transcription is dependent upon a subset of TAFs. It has been shown that only 3% of S. cerevisiae promoters are dependent on TAF2 (Shen *et al.* 2003). Individually TAF2 does not single out a clearly constrained DNA sequence. However, in a trimeric complex with TBP, TAF(II)250-TAF(II)150 complex select sequences that match the Initiator (Inr) consensus at the appropriate distance from the TATA-box. Inr recognition can determine the responsiveness of a promoter to an activator. TAF2 specifically binds to four-way junction DNA, suggesting that promoter binding by TFIID may involve recognition of DNA structure as well as primary sequence (Chalkley and Verrijzer 1999).

TAF2 peaks at the beginning of S-phase when extensive transcription initiation takes place (Fig. 2-15b). Its antisense counterpart is expressed in anti-correlated fashion (Fig. 2-16), suggesting its regulatory role.

CTF4

Cohesion between sister chromatids mediated by a multisubunit complex called cohesin is established during DNA replication and is essential for the orderly segregation of chromatids during anaphase. In budding yeast, a specialized replication factor C (RF-C) comprising Ctf18/Dcc1/Ctf8 and the DNA-polymerase-alpha-associated protein Ctf4 are required to maintain sister-chromatid cohesion in cells arrested for long periods in mitosis. These genes are also essential for proper chromosome segregation in meiosis. CTF4 exhibits genetic and physical ties to replication fork constituents. It has been shown that absence of either CTF4 or CTF18, also involved in sister chromatid separation causes sister chromatid cohesion failure and leads to a preanaphase accumulation of cells that depends on the spindle assembly checkpoint (Hanna *et al.* 2001). Transcription for CTF4 antisense channel occurs in anti-correlated fashion (Fig. 2-15c, 2-16).

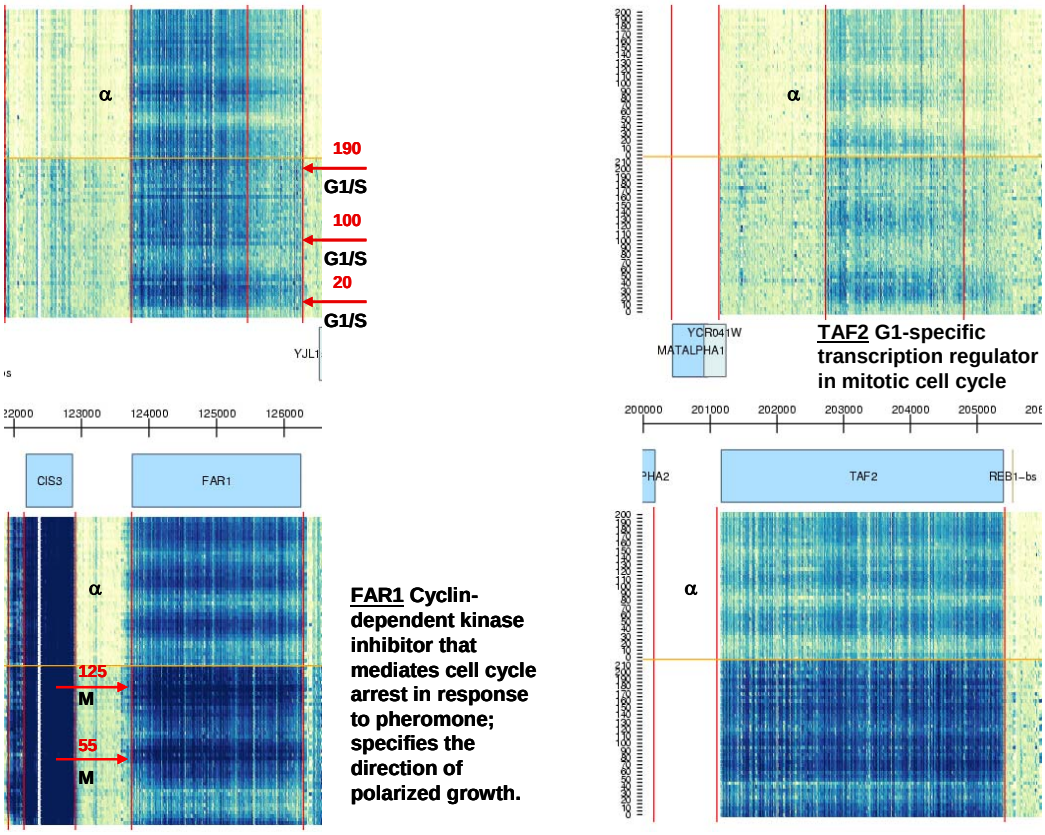
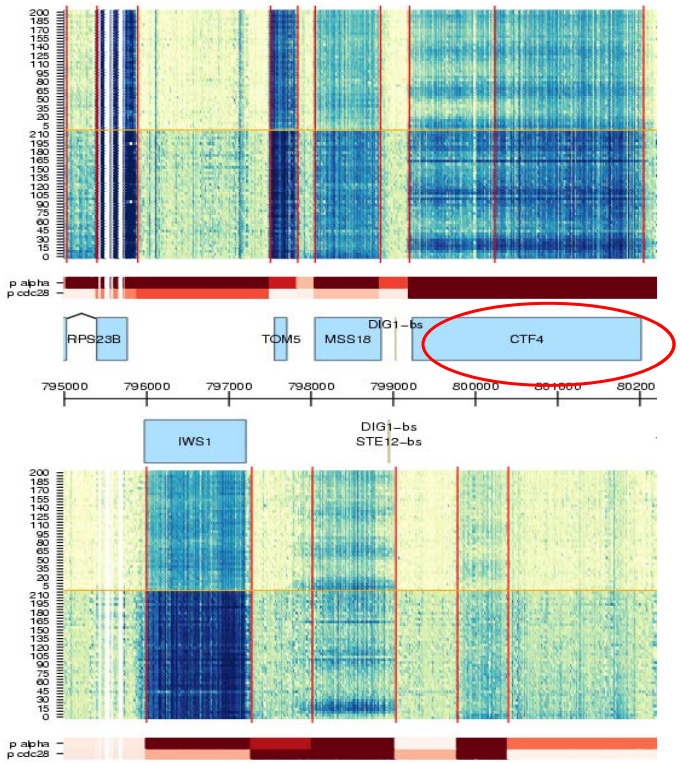


Figure 2-15. Heatmaps of periodic expression for sense ant antisense channels for FAR1, TAF2 and CTF4. Each horizontal line represents a time-point hybridization. Sense and antisense channels are to the top and bottom from the chromosome coordinate line, defined depending on the coding strand for ORF. Alpha and cdc28 datasets are devided by a horizontal orange line, top panel representing alpha dataset for both channels.



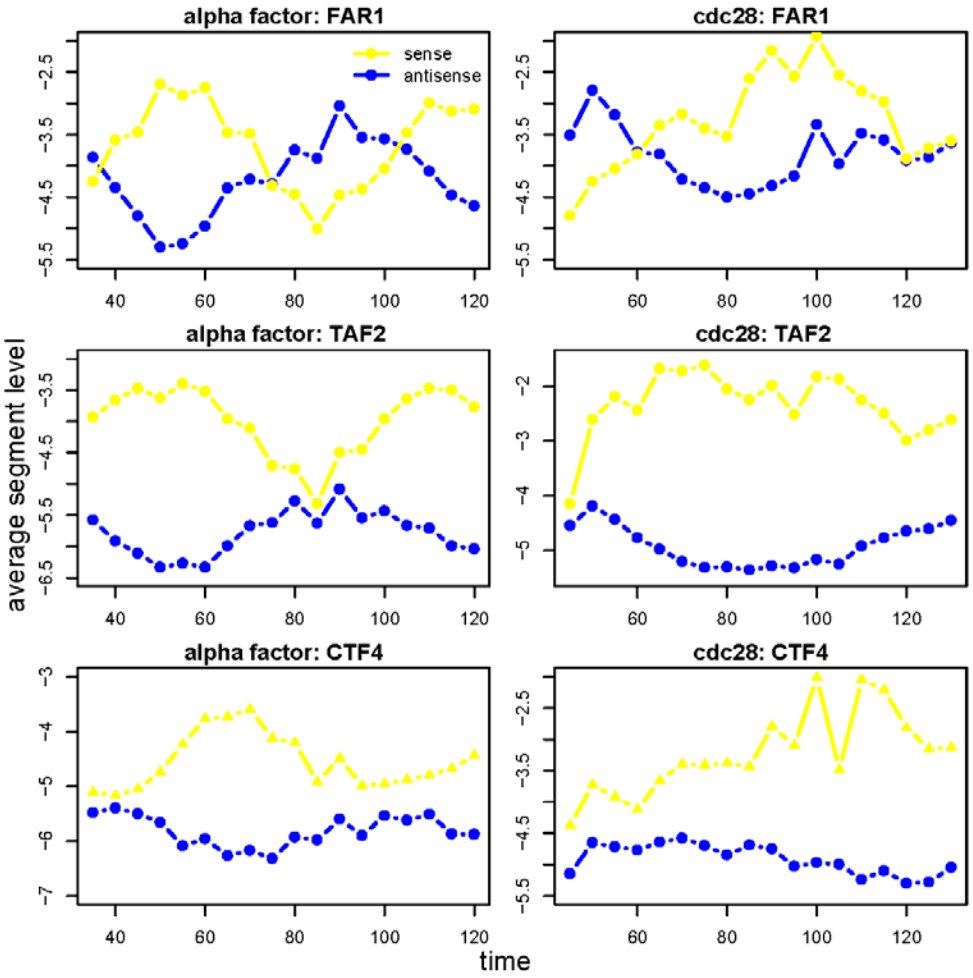


Figure 2-16. Correlation pattern of sense / anti-sense periodic cycling for Far1, Taf2 and CTF4. Two cell cycles shown for alpha-factor dataset and one for cdc28.

9. Promoter analysis and search for transcription factor binding sites within or upstream of the antisense features.

To obtain additional proof that the antisense features we identified as periodic are strongly cell-cycle regulated it is necessary to analyze their upstream intergenic regions. It has been shown for fission yeast that the more strongly cell cycle–regulated genes have longer than average upstream regions. For instance, the top ranked 200 periodic genes had upstream intergenic regions of about 1,200-bp median length, versus a genome-wide median length of 900 bp. the same phenomenon was observed for the cell cycle regulated genes of *S. cerevisiae* (Oliva *et al.* 2005). The longer-than-average promoters found for cell cycle–regulated genes suggested that these promoters might be above average in complexity.

The importance of studying intergenic regions of the cell cycle-regultaed genes is emphasized by the fact that among different mechanisms of antisense action in yeast, transcription interference appears to be central (Munroe and Zhu 2006). As discussed in the introduction, antisense RNAs may control the expression of their sense counterpart by having a stronger promoter region, to which a common transcription factor is recruited in the first place. Thereby transcription initiation of the sense message is not allowed until a certain amount of antisense message "buffer" is ensured (Lapidot and Pilpel 2006). Conversely, the initiation of transcription on one strand may help activate transcription on another strand by rearranging the local chromosomal structure (Camblong *et al.* 2007). In this regard, it is critical to search for common TF-motifs for the sense and antisense messages.

We first aimed to look at motifs for known TFs. Joern Toedling together with Matt Ritchie have analyzed the transcript and regulatory sequences (-800bp upstream and +800bp downstream of transcript) of 37 antisense (a/s) transcripts, after careful analysis reported as cell cycle-dependent (table **A**2). The choice of taking a sequence of -800pb upstream and +800bp downstream of the transcript was based on reports, describing a human promoter structure, that in eukaryotes the regulatory region may exceed that far or even further (Munroe and Zhu 2006). With the compact yeast genome, the regulatory regions may be smaller – around 600bp. 104 genes whose transcription was reported as cell cycle-regulated before and used as a positive control by Spellman (Spellman *et al.* 1998) were used as a control. Position-specific score matrices from MacIsaac (MacIsaac *et al.* 2006), an updated version of the ones provided by Harbison (Harbison *et al.* 2004) were taken for known transcription factor (TF) binding specificities. To assess whether TF binding specificities are overrepresented in the regulatory region of periodic transcripts, we also needed to assess these TF's binding specificities in a 'negative control set of transcripts' regulatory regions. For such a 'control set of transcripts' regulatory region, we took

a) the set of all transcripts in the Alpha cell cycle data

b) the set of novel antisense transcripts whose expression is not periodically cycling with the cell cycle.

With (a) we can say whether certain TFs are specific for the periodic transcripts of interest compared with all transcripts in the cell-cycle data, while with (b) we can say whether certain TFs seem to be specific for periodicity of antisense transcripts.

For TFBS search we have mostly used Python-based software suite TAMO, which provides interfaces to a number of previously published motif-finding algorithms, i.e. AlignAce, Meme, Weeder, and two new such algorithms (Fratkin *et al.* 2006).

Having compared 37 periodic antisense transcripts versus all transcripts in the 'alpha dataset' we discovered that Mbp1 and Swi6 TFs are overrepresented for the given list of antisense transcripts. As a control 252 non-periodic novel antisense sequences were compared versus all transcripts in 'alpha dataset' and no significantly ($p<0.001$) enriched TFs binding specificities were registered for this comparison. To assess the significance of the a/s results 104 of Spellman's known periodic transcripts were compared versus all transcripts in 'alpha dataset'. These were mostly enriched for Mbp1, Swi6 and Mcm1.

For the analysed 37 periodic a/s transcripts the Mbp1 motif is only present in 14 out of the 37 sequences, but the fraction of all transcripts that show this motif is far smaller than for these periodic a/s. Mbp1 is known to be involved in cell-cycle related regulation of transcription. In addition, Swi6, a known dimerization partner of Mbp1 is over-represented in these sequences.

Mbp1p is a DNA-binding protein that forms the MBF complex (Mlu1 cell cycle box [MCB] Binding Factor) with Swi6p, which acts as a trans-activator for MBF function. MBF is a sequence-specific transcription factor that regulates gene expression during the G1/S transition of the cell cycle. Many of the genes activated or repressed by MBF are involved in DNA synthesis and DNA repair (for example, *CDC21*, *CDC8*, *CDC9*, and also G1 cyclins). Mbp1p can bind DNA directly without Swi6p (Fig. 2-17).

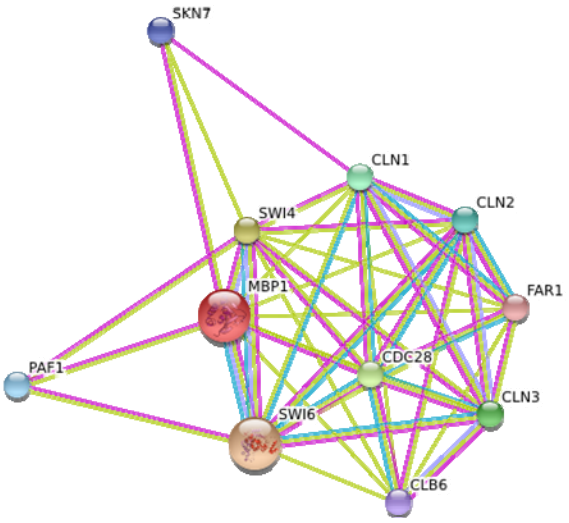


Figure 2-17. Functional associations of Mbp1 transcription factor with proteins, which transcription it initiates in complex with Swi6. Proteins are shown as nodes. (adapted from STRING).

These preliminary results suggest that the periodicity of these a/s is regulated by the same TFs as for known periodic genes. As shown in Fig. 2-18, most of the antisense presumably regulated by Mbp1 peak in the late G1 -> S phase transition together with the late G1 cluster genes, part of which are also regulated by MBF complex. Notably, however, out of the 14 sense counterparts for the 14 antisense transcripts enriched for Mbp1 TFBS, only 5 are cell cycle-regulated. Two are not expressed and the other 7 are expressed in a non-periodic fashion. One of the 5 sense ORFs is co-regulated with its antisense channel, whereas the other 4 cycle in anti-correlated mode. We could not find any TF that specifically seems to regulate a/s transcription in general. However, we have 124 well-annotated TFs in this analysis known hitherto. Perhaps for some TFs the binding specificity is not known or the TF may only bind DNA indirectly.

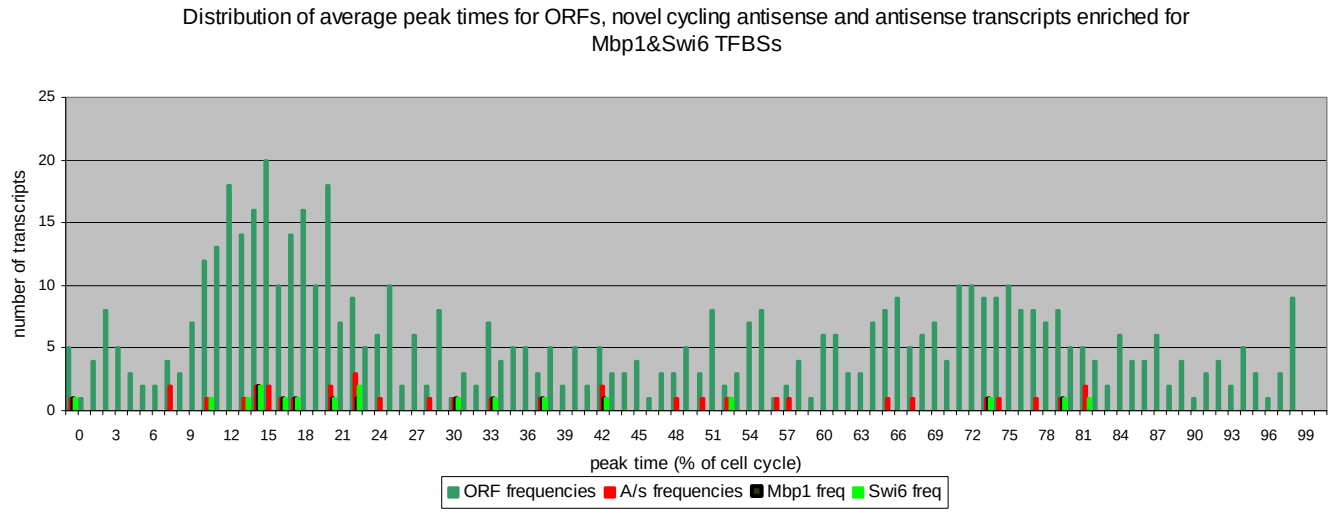


Figure 2-18. Distribution of peak times for ORFs (green), antisense transcripts (red), Mbp1 enriched TFBS antisense (black) and Swi6 enriched TFBS antisense (bright green).

We have also performed *DeNovo* motif finding for this set of 37 periodic antisense transcripts.

Some of the newfound Motifs identified by algorithms Meme and AlignAce in the regulatory sequences of the 37 periodic a/s were not deemed to be specific for the 37 a/s when searching for their occurrence with all Alpha cell-cycle transcripts. None of these had a striking resemblance to a known motif, when comparing them using the STAMP data base.

Interestingly, there appear more than 100 antisense transcripts expressed only at the first several points after release from the temperature-based cdc28 arrest, expression of which fades away as the cells progress through the cell cycle and does not oscillate further on. Such pattern resembles that of response

to stress and may reveal a non-characterized class of non-coding RNAs mediating heat shock response in yeast.

A number of high-confidence antisense transcripts were selected to further validate their functional role in the cell and in regulation of mitosis. These ORFs include: FAR1, TAF2, YMR178W, YLR050C, CTF4, MSS18, CHS7, etc. Their antisense counterparts were cloned into the plasmid with constitutive promoter in order to monitor fine-regulation of sense messages in response to antisense over-expression. There is an array of approaches to apply in order to ascertain the functionality of these selected candidates.

10. Detection of Novel Isolated cycling transcripts

135 novel isolated transcripts were chategorized for our dataset. A number of these oscillate with mitotic progression. The same two approaches were applied to the novel isolated filtered category list to distinguish periodic novel transcripts. A0.01 list (12 entries) was matched with top-500 list (14 entries) and visually inspected for cycling. Manually curated combined list of periodic novel isolated transcripts comprises 11 transcripts (table 2-7). As seen from figure 2-19, the peaks of expression of a few cycling novel isolated transcripts match very well the waves of excitiation in mitotic progression observed for annotated ORFs.

chr	strand	start	end	length	n	level	cycling	peak
2	+	254640	254864	225	26	0.88	+	34
2	+	430975	431591	617	72	0.9	+	83
4	+	49921	50857	937	111	0.26	+	98
4	+	80713	82633	1921	228	0.64	+	19
4	-	443805	444173	369	46	0.06	+	20
4	-	506510	507078	569	69	0.56	low magnitude, vaguely	76
4	-	1018103	1019103	1001	107	0.1	vaguely	13
4	-	1340831	1342151	1321	154	0.72	low magnitude	11
5	-	187029	188061	1033	122	1.85	+, low magnitude	16
7	+	166923	167347	425	46	2.1	+	21
16	+	286673	286993	321	38	0.52	+	69

Table 2-7. Novel Isolated cycling transcripts.

To elucidate the possible function of these novel transcripts and to clarify their importance for cell cycle regulation, first, protein signatures for these transcripts were examined. None of the cycling novel isolated transcripts turned out to have one.

Deletion strains of 10 novel periodic transcripts were generated in both strain backgrounds to check if a new transcript of interest gives a distinct cell cycle related phenotype. Growth curves of the knock-out (KO) strains do not show significant lagging in cell doubling time after non-synchronous growth in rich media for 28 hours at 30^0C and 37^0C, except for 2 mutants of YNN553 (KO#4 and #7) (Fig. 2-20).

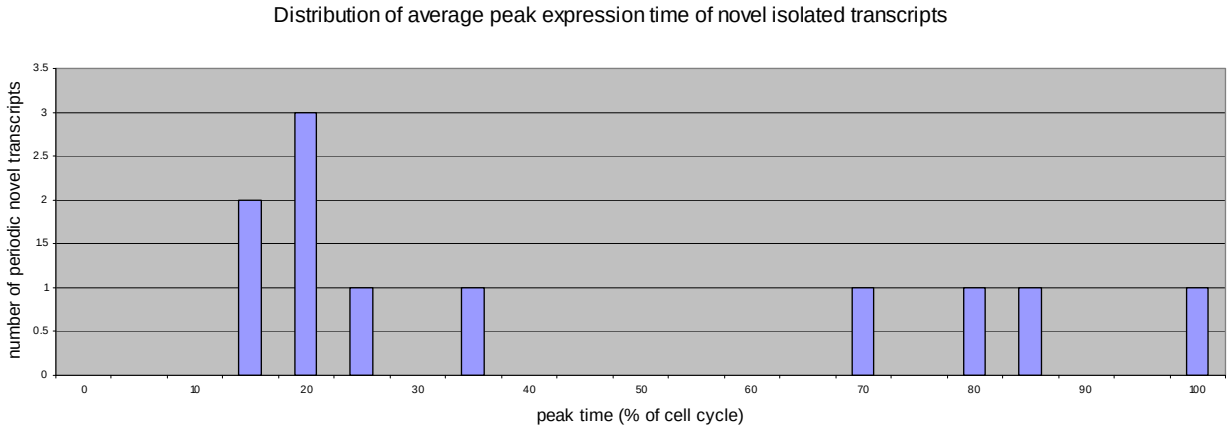
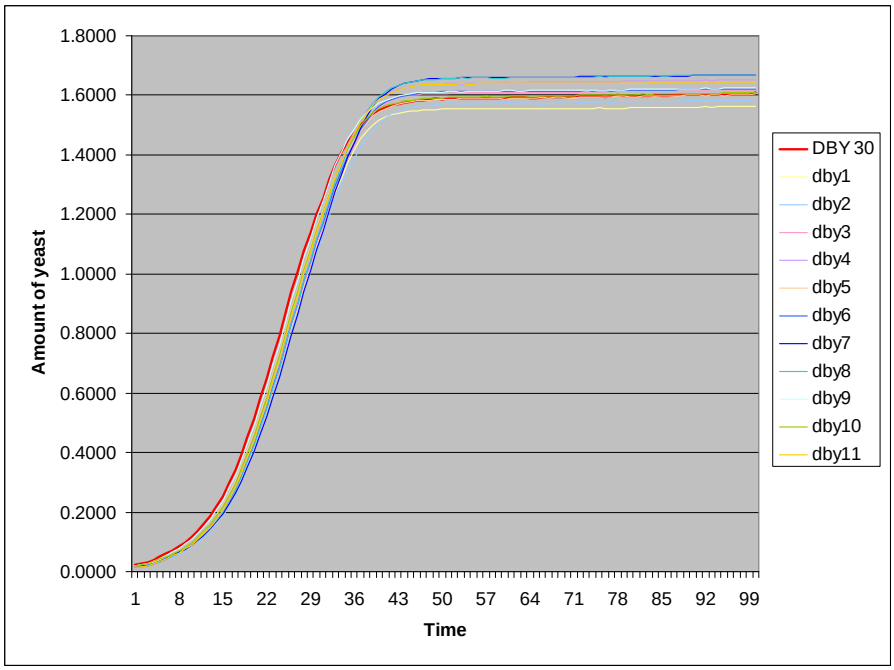
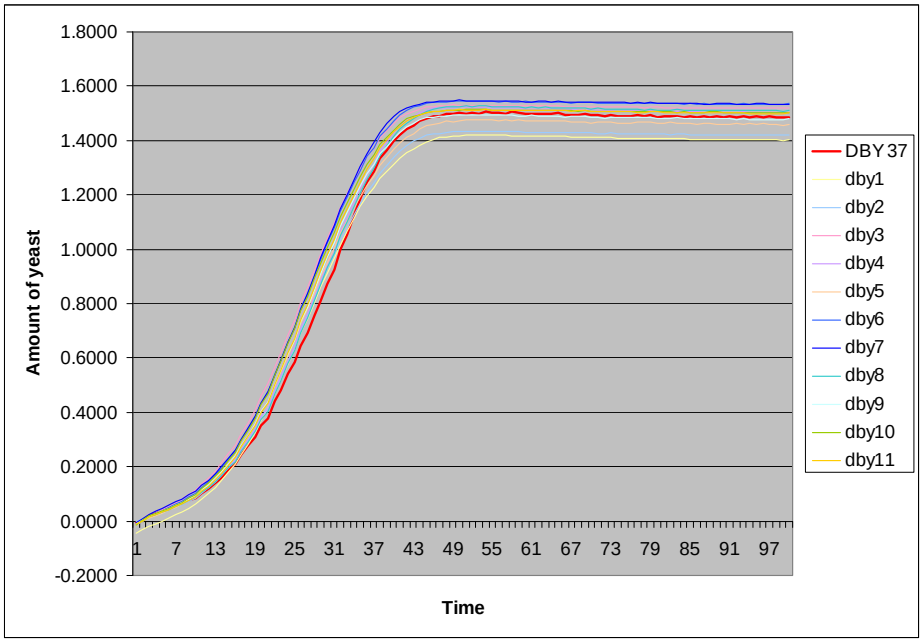
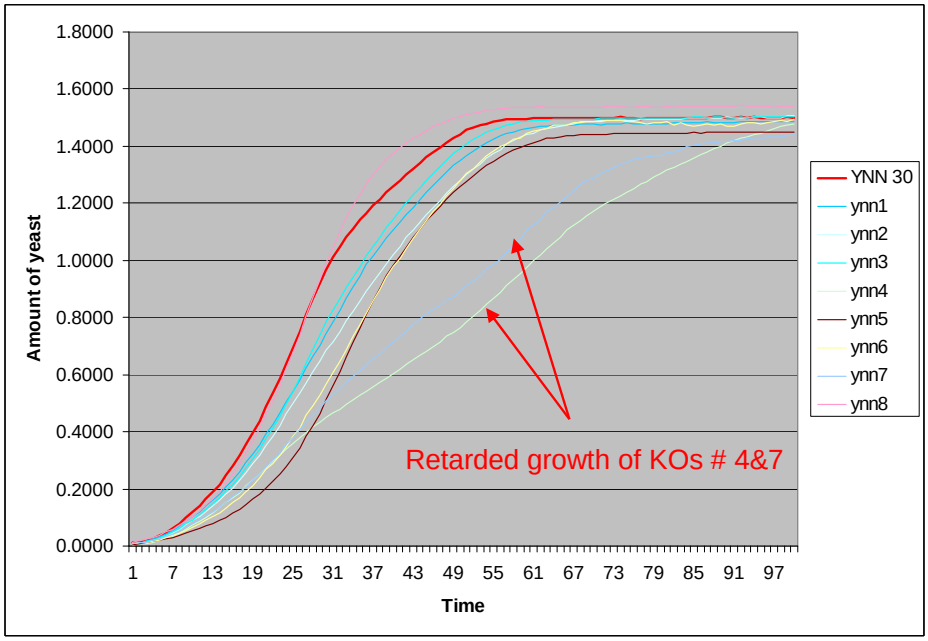


Figure 2-19. Distribution of cycling peaks of novel isolated transcripts in the phases of mitosis.

Figure 2-20. Growth curves of novel isolated cycling knock-outs of DBY8724 strain. Growth of DBY8724 at 30C (upper panel) and 37C (middle panel) and YNN553 at 30C (bottom panel).







11. Bidirectional promoters in yeast

Adjacent gene pairs in the yeast genome have a stronger tendency to express concurrently than random pairs do (Cohen *et al.* 2000). Examples of such expression are observed in our cell cycle dataset (Fig. 2-20). Early single gene studies of promoter specificity have revealed that (TBP) alone cannot define the productive orientation of general factor assembly on a promoter. The rotational and translational selectivity of TBP binding could be enhanced by general transcription factors TFIIA and TFIIB,

nevertheless they are not sufficient, at least individually, to confer a unique polarity to the preinitiation complex (Cox *et al.* 1997). Sharing of regulatory elements within the intergenic region of adjacent gene pairs was often considered the major mechanism responsible for co-expression and binding of common transcription factors and was thought to drive concurrent expression in both directions (Kruglyak and Tang 2000). Although previous studies suggested that the sharing of a common upstream activating sequences (UAS) plays an important role in regulating coexpressed pairs, and that divergent pairs are more likely to share the same regulatory system, the co-expression level (defined by correlation coefficient) of divergent pairs is not significantly higher than that of tandem pairs with a similar intergenic distance (Tsai *et al.* 2007). Hence, it is still in debate to what extent common transcription factors (TFs) contribute to the co-expression of adjacent genes.

Tsai et al. have comprehensively examined the intergenic regions between adjacent co-expressed genes to inquire whether these pairs frequently share common TFs. The conservation of adjacent pairs in five yeast species was investigated. By using the information for TF binding sites in promoter regions available from the MYBS database (http://cg1.iis.sinica.edu.tw/~mybs/), the ratios of TF-sharing pairs among all the adjacent co-expressed pairs in the yeast genomes were analyzed. This study provided clear evidence that sharing of the common TFs is not an exclusive component of the driving force in co-regulation of adjacent gene pairs in yeast.

A new study suggests a model, in which promoters are not intrinsically directional and can support inappropriately oriented transcription and initiate it from cryptic start sites, when chromatin structure is perturbed (Whitehouse *et al.* 2007). In this light, nucleosome positioning is crucial to induce transcription required by a physiological context of the cell as well as to block unwanted transcrioption. Isw2 is one such ATP-dependent chromatin remodelling enzyme, and belongs to a family of proteins that are highly evolutionarily conserved. In multicellular eukaryotes, Isw2 homologues have been implicated in the regulation of transcription, global chromosome structure, DNA replication, cell cycle progression, ribosomal DNA silencing and cohesin loading. In budding yeast, Isw2 acts as a gene repressor by overriding the underlying nucleosome positioning signals of DNA, repositioning nucleosomes over unfavourable DNA sequences to establish a chromatin configuration that is repressive to transcription.

Presumably, during evolution yeasts have developed a sophisticated regulatory system that integrates both TF-based and non-TF based mechanisms(s) for concurrent regulation of neighboring genes in response to various environmental changes.

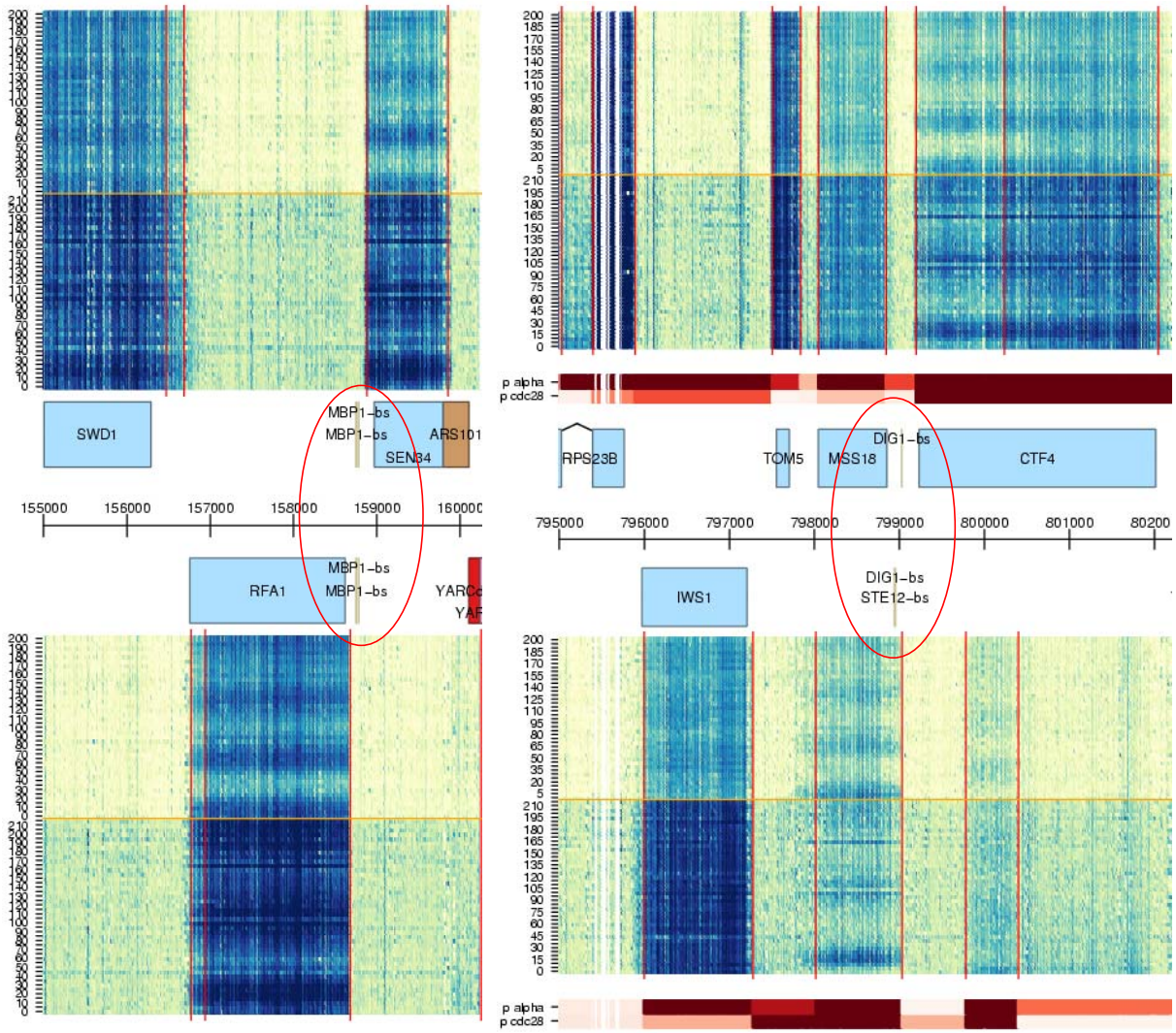


Figure 2-21. Examples of bidirectional promoters.

3. DISCUSSION

I. A high-resolution map of transcription in the yeast genome

Understanding complex functional mechanisms requires the global and parallel analysis of different cellular processes. Microarray methodology presents an integrative platform, which serves to achieve this aim. It has transformed molecular biology leading the way from studies of an individual function of a gene or protein, or at best, interactions within a signaling pathway to a more global scrutiny of cellular activity at different layers of its realization. Adaptation of microarrays in different research areas entailed the realization that detailed molecular analysis of interaction pathways is insufficient; a thorough molecular and structural analysis on all levels of "–omes" is a prerequisite for the elucidation of the complex and interrelated processes that occur in biological systems. The basis of microarray technology built on unbiased sample screening and data accumulation, has led to a more comprehensive characterization of biological systems, which emphasizes their complexity and multilevel regulation.

Transcriptional profiling with microarrays has been perhaps, the most widely used application of chip technology in the recent years. It has revealed abundant transcription from eukaryotic genomes unaccounted for by protein coding genes. A high-resolution genome-wide survey of transcription in a well annotated genome can help to relate transcriptional complexity to function. Having a high-density oligonucleotide tiling array tool in hand, we have developed the methodology of whole-genome transcriptional profiling in yeast. By quantifying RNA expression on both strands of the complete genome of *Saccharomyces cerevisiae* with tiling array, our study identified the boundary, structure, and level of coding and noncoding transcripts. A total of 85% of the genome is expressed in rich media. Apart from expected transcripts, we found operon-like transcripts, transcripts from neighboring genes not separated by intergenic regions, and genes with complex transcriptional architecture where different parts of the same gene are expressed at different levels. We mapped the positions of 3' and 5' UTRs of coding genes and identified hundreds of RNA transcripts distinct from annotated genes. These nonannotated transcripts, on average, have lower sequence conservation and lower rates of deletion phenotype than protein coding genes. Many other transcripts overlap known genes in antisense orientation, and for these pairs global correlations were discovered: UTR lengths correlated with gene function, localization, and requirements for regulation; antisense transcripts overlapped 3' UTRs more than 5' UTRs; UTRs with overlapping antisense tended to be longer; and the presence of antisense associated with gene function. These findings may suggest a regulatory role of antisense transcription in *S. cerevisiae*. Moreover, the

data show that even this well studied genome has transcriptional complexity far beyond current annotation.

Attempting to further elucidate the functionality of antisense transcripts in yeast we have performed transcriptional profiling during a tightly regulated and well-studied process in yeast – cell division cycle.

II. Profiling periodic transcription of the cell cycle – regulated genes.

Cell cycle–dependent mRNA fluctuations have been observed for genes involved in many cellular processes, including control of mRNA transcription, responsiveness to external stimuli, and subcellular localization of proteins. Genetic studies have revealed that the activity of cell cycle–regulatory proteins is required for normal DNA repair, meiosis, and multicellular development (Cho *et al.* 1998). The numerous biological changes associated with the cell cycle make it an attractive model for the study of genome-wide regulation of gene activity. Parallel identification of all of the genes in a genome that are coordinately regulated during such a process provides a consistent internal standard for comparison of gene activity over time and makes it possible to search statistically for candidate regulatory sequences.

DNA microarrays and other genome-wide methods have provided global perspectives on the system underlying cell cycle–regulated gene expression and its integration with other aspects of the cell-cycle machinery.

Unprecedented resolution capacity of yeast whole-genome tiling arrays allowed me to perform whole-genome transcription profiling during mitosis in *S.cerevisiae* augmenting previous studies (Cho *et al.* 1998; Pramila *et al.* 2006; Spellman *et al.* 1998) in the resolved details of periodic expression of identified cell cycle–regulated transcripts and most importantly in discovering novel oscillating features.

Using two statistical methods of periodicity analysis, which complement each other well in specificity and sensitivity of periodic transcripts detection, we have generated a comprehensive list of genes regulated in the cell cycle. We have enriched the Spellman's list of periodic annotated genes for 188 transcripts. They are involved in cell cycle regulation, metabolic processing, gene silencing, transcription, and other processes central to mitotic regulation. 7 of them are directly involved in DNA damage and spindle assembly check-points. As shown earlier (Fig. 2-13), the number of genes identified

in our study yields only to that of Gauthier et al. (Gauthier *et al.* 2008), who had combined computational re-analysis of all hitherto existing datasets and thus built their conclusions on a much broader range of experiments and statistical analyses. This proves the quality and degree of fine-resolution of our experimental data and the in-depth detailed computational analysis. Comparison with the cell cycle-regulated genes identified by Gauthier et al. on the basis of all previously available datasets (Gauthier *et al.* 2008) (www.cyclebase.org) identified 223 ORFs not detected as periodic by previous studies. 154 periodic ORFs are in common between the 188 ORFs not found in Spellman's list and the 223 ORFs not detected as periodic by Gauthier.

In consent with previous studies on transcriptional regulation during specific mitotic phases, we have shown prevalence of periodic expression of annotated genes in 3 distinct periods of cell cycle progression: late G1/S transition, G2/M transition and exit of M phase of mitosis (Fig. 2-12b). These three major transcriptional waves roughly coincide with three main cell-cycle transitions (initiation of DNA replication, entry into mitosis, and exit from mitosis) and fall extremely well into the recent computational model of Lovrics and colleagues (Lovrics *et al.* 2006). They have employed the idea that the cell cycle is driven by biochemical kinetics of molecular interactions controlling cyclins' synthesis and degradation, and have converted this kinetic mechanism into a set of differential equations, which describe the time-courses of three major classes of cyclin-dependent kinase activities (Chen *et al.* 2000).

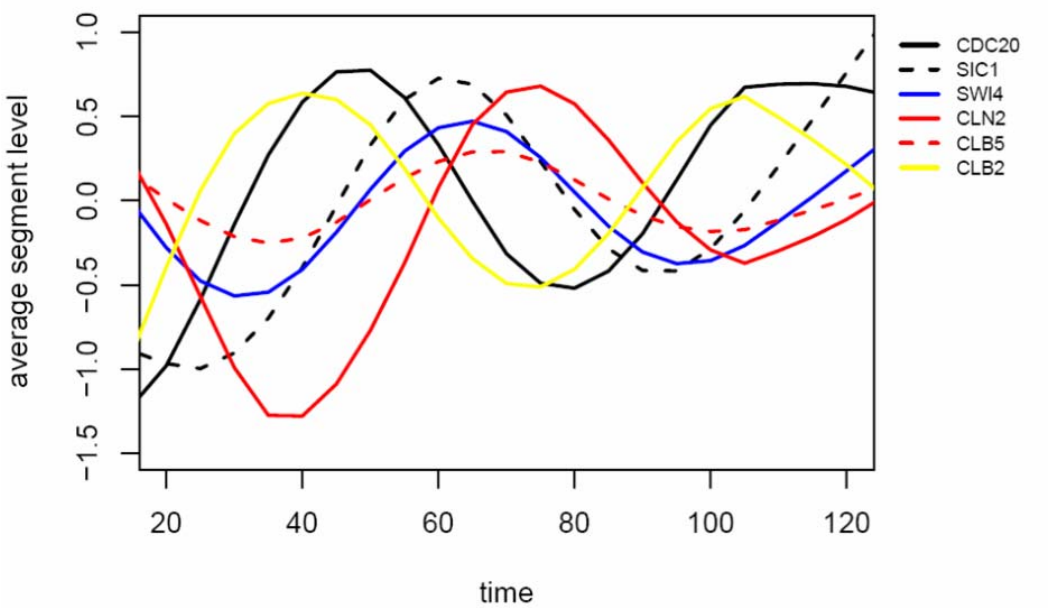


Figure 3-1. Waves of excitation and relaxation of individual cell cycle regulated genes.

By calculating Jacobean eigenvalues for kinetic differential equations, which determined the equilibrium of cell cycle trajectories in conversion from steady state to the transcriptionally active state, which

stipulates transition to the next phase of mitosis, they proposed that several waves of excitation and relaxation of transcriptional machinery drive the mitotic progression. In this model, excitation represents irreversible transition points of mitotic progression (Lovrics *et al.* 2006); in relaxation stage all kinetic processes are in steady state. The first excitation period at START transition flips the control system from a G1 state into S phase with a concomitant initiation of bud growth. The second excitation is responsible for the initiation of mitosis, which is premature in budding yeast. The third excitation changes the morphology of bud growth from polarized to isotropic. Finally the last excitation at mitotic exit pushes the system back to G1 phase. Excitations are reflected in transcriptional fluctuations of single genes, important for irreversible progression through the distinct phases of the cell cycle (Fig. 3-1). Thus, at the beginning of the cell cycle in early G1 all the major cyclins (Cln2, Clb5 and Clb2) are absent, but the level of G1 stabilizers (Sic1 and Cdh1) is high. Later on, due to steep increase in expression of major transcription factor complexes MCB and SCB, the levels of Cln2 and Clb5 – G1 cyclins raise dramatically stipulating irreversible transition through START. Activities of Cln2/Cdc28 and Clb5/ Cdc28 trigger bud formation and DNA replication, respectively, the two processes, which occur in parallel in yeast. As SCB transcription ceases Cln2 levels also start to gradually decrease, thus switching the pattern of bud growth from polarized to isotropic, i.e. in all directions. Simultaneously an M-specific transcription factor Mcm1 is activated, which turns on Clb2 expression. Clb2/Cdc28 in turn activate Cdh1 and the system slowly returns back to steady state at early G1 (Nasmyth 1996).

Clearly, many cell-cycle proteins are regulated at multiple and partly redundant levels, and it is the overall regulation comprised of transcriptional activation / inhibition equilibrium as well as post-transcriptional modifications, and localization determinants, that is important. It has been suggested that parsimony may explain much of cell cycle–regulated transcription (Bahler 2005), whereby genes are expressed when there is a special need for their products at a particular phase in the cell cycle.

We have provided a detailed tapestry of finely resolved transcriptional activity of ~600 periodic genes, which fall into different GO categories and are in control of or are involved in various processes in the cell. Notably, the identified genes are not only regulated by the core cell cycle-driving proteome themselves, but in turn stipulate periodic expression of their downstream effectors and oscillating subunits of key complexes. Potentially, they could also be regulated by ncRNAs or via epigenetic regulations. Moreover, our results provide clear evidence that transcriptional control should not be considered in isolation. As in the whole-genome transcriptome analysis in rich media growth, we have identified a significant number of novel antisense transcripts. Some of these antisense features cycle along with their sense counterparts.

Computational and experimental studies in yeast, *Drosophila* and mammals (Boutz *et al.* 2007; David *et al.* 2006; Samanta *et al.* 2006; Schoenfelder *et al.* 2007), supported by recent discoveries on a single-gene level (Camblong *et al.* 2007; Hongay *et al.* 2006; Uhler *et al.* 2007), provide clear evidence that regulation of transcription by *cis*-acting untranslated antisense RNAs is widespread in the eukaryotic cell. In fact, it appears more prevalent and important than previously realized and significantly contributes to shaping of the cellular general transcriptional profile. Given the finesse of the cellular protein regulatory network, which mainly relies on post-transcriptinal mechanisms, regulation through antisense RNA may almost seem redundant and a waste of energy for the cell. However, in more complex processes, like cell division cycle, with many layers of fine-tuning and regulation at the level of gene transcription, protein production, localization, modification and degradation, non-coding RNA can be regarded as a complement to existing tiers of regulatory networks. Indeed, if most of the cellular responses to various stimuli were determined at the expense of transcription factors and their turnover, the promoters of the genes would have to be remarkably complex to ensure sophisticated regulation of multiple parallel processes of the cell well-being. As discussed in the introduction, increase in organismal complexity requires multiple parallel signaling and regulatory networks functioning in the cell (Mattick and Gagen 2001). The analysis of sequenced genomes revealed that the relative amount of non-protein-coding sequence increases in consistency with complexity. This is presumably due to the expansion of *cis*-acting regulatory elements and genes specifying *trans*-acting non-protein-coding RNAs (Taft *et al.* 2007).

Hence, a systematic examination of the role of antisense regulation in the context of a tightly controlled process such as mitosis is suitable to elucidate the extent of ncRNA expression in yeast and shed some light on the global mechanism of their action.

It has been accepted that post-transcriptional regulation plays a principal role in mitotic progression and regulation by means of intricate stimulatory and inhibitory impulses. This strictly controlled chain of input-output switches is determined by kinases, phosphatases, activators, inhibitors and cellular ferries, expressed just in time when they are needed and degraded almost immediately upon fulfillment of their mission via ubiquitin pathway. The changes of and regulation at the protein level is accompanied by complementary changes in the mRNA levels of the corresponding genes, therefore studying transcriptional wiring of the cell cycle is very meaningful to reveal the control system, functioning in parallel. Along these lines, antisense regulation might offer a higher degree of flexibility to serve the needs of differential gene expression and serves a “buffering layer” for protein signaling pathways.

Out of ~260 antisense transcripts that we discovered 37 display periodic pattern. Interestingly, half of them are expressed coincidentally with peak of expression intensity of periodic ORFs, whereas the other half peaks at the periods of relaxation of transcriptional machinery which drives phase transition, and only a few are expressed at the exit of M phase (Fig. 2-14a). This strongly suggests regulatory role of antisense transcripts during the cell cycle. We can assume that during S and G2/M phases, when the DNA damage and spindle assembly check-point controls are executed antisense provide additional mechanism of inhibition of gene responsible for irreversible entry into M phase and ensures that no premature progression occurs.

We have performed the search for known transcription factor binding sites upstream of cycling antisense features and found that 14 antisense transcripts are enriched for Mbp1 and Swi6 TFBS, which coincides remarkably with the observed pattern of peak expression at late G1 and G1/S transition (Fig. 2-18). These results suggest that the periodicity of these antisense is regulated by the same TFs as for known periodic genes. We could not find any TF that specifically seems to regulate a/s transcription in general. We also performed *de novo* motif finding for cycling antisense and matched the motifs that were found against the known transcription factors' sequences. This analysis has shown slight enrichment for RLR1 and AZF1 transcription factors, which was not significant enough.

Overall there are clear indications of antisense regulatory role in mitosis; the exact mechanisms of it remain to be elucidated by single-gene studies.

In addition to novel antisense transcripts we have discovered novel isolated cycling transcripts. The peaks of their expression match very well the waves of excitiation in mitotic progression discussed for annotated ORFs . This suggests that the newly discovered cyling features could be *bona fide* non-protein-coding genes, involved in regulation of cell cycle or be regulated themselves by periodic genes. Knock-out constructs of these transcripts display no phenotype for the DBY 8724 strain and two KOs show retarded growth in YNN553 background. Lack of apparent phenotypes does not per se exclude the involvement of these trascripts in mitotic processes as they can be downstream of oscillating effectors.

4. OUTLOOK

Among different alternatives for empirical transcriptome mapping, whole-genome tiling arrays have been the most comprehensive tool for studies of global gene expression and provided a more unbiased view of the transcriptional activity within genomes. *S.cerevisiae* tiling arrays were instrumental for shaping our systematic understanding of dynamic events during complex cellular processes, such as cellular responses to stress, mitosis and meiosis as well as in uncovering epigenetic modifications and local chromosomal structure associated with various changes in the whole system.

Patterns of the cell cycle transcription have been extensively surveyed and characterized in detail integrating various experimental approaches and mathematical models. The current work is distinct in the fine resolution of the mitotic transcription data it provides. Our study contributes to the systematic understanding of mitotic progression in yeast and complements previous studies enriching for new periodic ORFs, compared to other lists. Moreover, it adds a unique layer of the non-protein-coding gene expression in the cell cycle to the complexity of mitotic machinery by the discovery of periodic novel isolated unannotated and novel antisense transcripts.

Periodic expression of antisense transcripts, in particular of the ones, found opposite some key cell cycle-regulated ORFs, and the scale of global antisense expression in yeast during mitosis provide indirect evidence for the functional role of non-protein-coding genes in regulating important processes in this model organism and raises several key questions to answer in the next few years to come. What are the exact functions of particular antisense regulatory transcripts in the cell? Do different classes of antisense transcripts become transcriptionally active in different processes of the cell life cycle? It has become clear, that although yeast lack all vestiges of the protein machinery required for RNA silencing or translation inhibition by miRNA or siRNA-like mechanisms it hides a lot of other resources, via which antisense regulation could be realized. The recent examples comprise transcriptional interference and epigenetic modifications. Do any other mechanisms of ncRNA regulation exist in yeast? What are the roles of cycling novel unannotated transcripts? Do they enrich the class of non-coding RNA genes in yeast? What role does chromatin structure and regulation play in regulating antisense transcription?

The relative ease of performing single cell-based assays in *S. cerevisiae* makes it a well-suited model for answering the questions above. To continue this work, we are generating a number of antisense constructs for over-expression in synchronous as well as asynchronous yeast culture to monitor its effect on sense transcription and translation. Over-expression of antisense opposite some important genes

regulating pheromone-response pathway and transition through mitotic phases would be the first important step to gain the information on periodic antisense functionality during the cell cycle.

Advancements in microarray technology bring about more sophisticated mathematical and statistical approaches to analyze the new findings. Design of such methods is particularly topical in surveys of complex highly-regulated cellular processes. Fourier based methods have proven successful for the analysis of time series of oscillating transcripts. Both recent comprehensive computational analyses of the available cell cycle datasets (de Lichtenberg *et al.* 2005b; Gauthier *et al.* 2008), however, emphasize that one single method does not suffice to generate a comprehensive unbiased list of periodic genes in the cell cycle. Therefore, it is important to develop an integrated mathematical model, which will allow integrating the results from different groups.

The global view on a biological system obtained by microarray technology in future will expand towards experimentally more complex systems and develop into experimental multiplexing by analyzing different processes on a single system platform. Despite gene expression technologies having greatly matured over the past years, it has become clear that hybridization-based approaches have obvious limitations in cross-species comparisons (Gilad *et al.* 2005; Oshlack *et al.* 2007). Alternatively, sequencing-based approaches are used to measure gene expression. High-throughout automated sequencing may significantly reduce the costs of whole-genome studies thereby allowing for screens of multiple strains at a time, mutant for certain antisense or having a deletion around a novel isolated transcript. Application of this technology will also speed up answering the challenge to uncover hitherto clandestine regulatory roles of non-coding genes in mitosis.

5. MATERIALS AND METHODS

1. Tiling Array Design

The array was designed in collaboration with Affymetrix (Santa Clara, CA) (PN 520055). With a 5 micron feature size, 25 bp probes are tiled on this array on sense and anti-sense strands with 8 bp spacing with a 4bp shift between the strands. This array is unique in interrogating every single nucleotide of the genome 6 times and in being the first array containing the whole eukaryotic genome on it. Such design enables identification of all transcripts, including those, lacking coding capacity.

2. Total RNA extraction, Poly(A)-RNA enrichment and cDNA preparation.

An *S.cerevisiae* S288c background strain S96 (MATa *gal2 lys5*) was grown in 100 ml of rich media (2% Difco peptone, 1% yeast extract, 2% dextrose, supplemented by 0.003% adenine hemisulfate) to mid-exponential phase (OD_{600}~1.0). Total RNA was isolated by the standard hot phenol method. Briefly, yeast cells were disrupted by vortexing with glass beads in hot acid phenol. After centrifugation RNA was re-extracted from the aqueous phase with phenol:chlorophorm:isoamyl alcohol in Phase-Lock-Gel (PLG) system (Eppendorf), precipitated under cold EthOH and resuspended in DEPC-H_2O.

Poly(A)-RNA fraction was enriched from 1mg of total RNA by two rounds of passage through Oligotex columns with oligo-(dT) beads (Qiagen's Oligotex mRNA Midi Kit). Total RNA and double-enriched poly(A)-RNA fraction were treated with RNase-free DNaseI using Ambion's Turbo DNA-free Kit for 25 min. at 37^0C according to manufacturer's instructions. DNaseI was inactivated by a DNaseI inactivating reagent. Poly(A)-RNA was reverse transcribed to single-stranded cDNA for microarray hybridization in 200ul reaction as follows: 9ug of RNA was mixed with random hexamers (Invitrogen) in the ratio 2:1 and incubated at 70^0C for 10 min, transferred on ice and mixed with 40 ul 5x first strand synthesis buffer (Invitrogen), 20 ul 0.1M DTT (Invitrogen), 5 ul 10mM dNTPs (Invitrogen) and 10ul of Superscript reverse transcriptase II (Invitrogen). The mixture was incubated at 42^0C for 1 hour. SSRTII was inactivated by addition of 200ul phenol:chlorophorm:isoamylalcohol (25:24:1) and ss-cDNA was purified by PLG system and subsequent salt precipitation under EthOH. ss-cDNA was digested to 50-100 bp fragments with 0.1 U of DNase I (Invitrogen) in 1× One-Phor-All buffer (Pharmacia) containing 1.5 mM $CoCl_2$ (Roche) at 37°C. After heat inactivation of the DNaseI, the cDNA fragments were 3'end-labeled in the same buffer by the addition of 1.5 ul of Terminal Transferase (25 U/ul) (Roche) and 1.5 ul 10 mM biotin-N6-ddATP (Molecular Probes) for 2 hours at 37°C. 3 ug of ss-cDNA were hybridized per tiling array.

3. Genomic DNA preparation.

For DNA hybridization S96 yeast strain was grown in rich media overnight to saturation and the whole-genomic DNA was purified on a Qiagen column using Qiagen's Genomic DNA Kit. 15ug of genomic DNA was fragmented with 0.2U of DNaseI (Invitrogen) for 5 min at 370C as described above, labeled and hybridized to tiling array.

Three replicate hybridizations (biological) of poly(A), two of total RNA, and three of genomic DNA were performed.

4. Probe Annotation.

Probe sequences were aligned to the genome sequence of *S. cerevisiae* strain S288c (SGD of August 7, 2005). Perfect match probes were further analyzed.

5. Normalization.

RNA hybridization intensities were adjusted by Nij = (Xij – Bj(Ai)) / Ai, where X_{ij} is the RNA intensity of the *i*th probe on the *j*th array, A_i is the geometric mean of the intensities from the DNA hybridizations, $B_j(A)$ is a continuous function that parameterizes the estimated background of probes with gain A, and N_{ij} is the adjusted intensity. Probes were grouped into 20 strata defined by the 5%, 10%, 15%, ..., 100% quantiles of A_i. Within each stratum, and for each array j, the midpoint of the shorth of the intensities of the probes for which no genomic feature was annotated on either strand was calculated. Linear interpolation yielded the function B_j. Dead probes (the 5% of probes with lowest signal in the DNA hybridization) were discarded. The values N_{ij} were background-adjusted and transformed to $\log_2$ scale by using VSN (Huber *et al.* 2002).

6. Segmentation.

Segments of approximately constant hybridization signal were defined by using a dynamic programming algorithm that, for each chromosome strand separately, minimizes the cost function

$$G(t_1, \ldots, t_S) = \sum_{s=1}^{S} \sum_{j=1}^{J} \sum_{i>=t_s}^{i<t_{s+1}} (y_{ij} - \bar{y}_{sj})^2,$$

where y_{ij} is the VSN-normalized signal of the ith probe on the jth replicate array, $\bar{y}_{sj}$ is the arithmetic mean of the signal values of array j in segment s, S is the number of segments, and t_1, ..., t_S are the segment boundaries (Picard *et al.* 2005). For each chromosome, S was chosen such that the average segment length was 1,500 nt. S, the only parameter of the segmentation algorithm, controls the sensitivity–specificity tradeoff and was chosen to yield high sensitivity.

All analyses were performed with custom-written software in the language and statistics environment R (Team 2007) and BIOCONDUCTOR (Gentleman 2005). For additional details on analyses and experimental procedure, see *Supporting Text*, which is published as supporting information on the PNAS web site (David *et al.* 2006).

7. Cell Cycle Synchronization.

50 ml of W101 (MAT**a** *ade2-1 trp1-1 leu2-3,112 his3-11, 15 ura3 can1-100* [psi1]) background temperature sensitive *cdc28-13* mutant *S.cerevisiae* strain K3445 (YNN553) was grown over the day in the shaking water-bath at 25^0C and diluted in 3x1.6L cultures for overnight growth in the air incubator at 25^0C. In the morning the cultures of OD600 ~ 0.2 were mixed together, distributed into 45x100ml samples and arrested in late G1 at START by shifting the temperature from 25^0C to 38^0C. After 3.5 hours (equals to more than two complete cell cycles) the cells were transferred back to permissive temperature to re-initiate cell division and samples were collected every 5 minutes for 215 minutes. The cultures were centrifuged and snap-frozen in liquid nitrogen. The degree of synchrony was monitored by assessing the number of budding cells and measuring the bud size. Nuclear position was determined by Hoechst staining with fluorescence microscopy.

To arrest *bar1* strain DBY2487 (MAT**a** *GAL2 ura3 bar1::URA3* (Spellman *et al.* 1998) in G1 at START α-factor pheromone was added to a final concentration of 600ng/ml. After 2 hours of arrest cells were released by washing and recovered in fresh preconditioned medium to facilitate initiation of mitosis. Samples were collected every 5 minutes for 220 minutes (equals to three cell cycles). The degree of synchrony was monitored by assessing the number of budding cells. Nuclear position was determined by Hoechst staining with fluorescence microscopy.

8. Hybridization sample processing

Total RNA fraction was isolated from the culture corresponding to each time-point as described above. Poly(A) - enriched RNA was obtained by single passage through the Oligotex Oligo-dT Column (Qiagen), treated with DNaseI (Ambion), and reverse transcribed to cDNA with random hexamers similarly as described above. Each RT reaction was carried out in replica and comprised 6ug of poly(A)-RNA, 3ug RH6, 1ul of 6 mg/ml Actinomycin D, 0.4 mM dNTPs containing dUTP, and 1600 Untis of SuperScript II (Invitrogen) in a total volume of 200 ul. The synthesis was carried out at 42^0C for 1h10min, the enzyme was inactivated at 70^0C for 10 min and RNA in the RNA:cDNA hetero-duplex digested by RNAseH for 15 min at 37^0C with subsequent inactivation for 15 min at 70^0C . Replica cDNA samples were further applied to Affy Clean-up column (Affymetrix) and eluted together. Purified cDNA was fragmented and labeled with WT Terminal Labeling Kit (Affymetrix) according to manufacturer's instructions and 3.3ug of each time-point sample was hybridized to tiling arrays.

9. Array Normalization and Segmentation

We preprocessed the arrays with the Robust Multi-chip Average (RMA) approach. RMA is used for gene-centric Affymetrix arrays when the probesets are defined in advance. It comprises three steps: 1) background adjustment, 2) quantile normalization, 3) summarization by median polishing. RMA uses only the information from the perfect match (PM) probe intensities. This method summarizes the intensities from the probes which make up a probe set in a robust way so that the summarized expression values are insensitive to outlier probes. The values between arrays are normalized to make them more comparable, assuming similar intensity value distribution on all arrays and arranging them in ascending order. Segmentation was performed similarly as above.

10. Detection of periodic genes.

The data from the cdc28 and alpha factor time-course were analyzed separately using the approach of Ahdesmaki et al (Ahdesmaki *et al.* 2005), which calculates p-values for a robust nonparametric version of Fisher's g-test.

For each segment, the test was carried out at the frequencies corresponding to the estimated cycle time from each experiment (90 minutes and 65 minutes for cdc28 and alpha factor, respectively). P-values were obtained by permuting each time series 1000 times to obtain an empirical distribution of the robust

g-statistics and comparing the g-statistic from the original time-series to this reference distribution. Due to the large number of tests performed for each time-course, p-values were adjusted for multiple testing using the FDR method of Benjamini and Hochberg (Benjamini 1995). Combined p-values for the two series were obtained by multiplying the p-values from the individual experiments.
Additionally combined dataset was analyzed by the Fourier transform proposed by de Lichtenberg et al. (de Lichtenberg *et al.* 2005b), which takes into account both magnitude of oscillation of the expressed transcript and periodicity of expression intensity pattern for each time-point.
The high- and medium- confidence cut-off score periodic gene lists were further curated manually to generate a combined list of cell cycle regulated transcripts.

11. Transcription Factor Binding Sites (TFBS) analysis.

We used the Python-based software suite TAMO for this analysis with some additional R scripts.

A string is said to match a motif if under that motif it achieves a score of 0.8 times the maximal score any string in the set of control sequences could have under the motif. Typically, this fraction of the maximal score is set to 0.7 to 0.9. Our setting, 0.8 is neither a very stringent nor a very relaxed setting.

Over-representation of known TF binding specificities are assessed by the group-specificity score by J. Hughes (Hughes *et al.* 2000) and G. Church (Zhu *et al.* 2002).

In addition, we look for new statistically over-represent sequence motifs in the regulatory sequences of interest. TAMO provides interfaces to a number of previously published motif-finding algorithms, i.e. AlignAce, Meme, Weeder, and two new such algorithms.

6. CITATIONS

References

Ahdesmaki, M., H. Lahdesmaki, R. Pearson*, et al.* (2005). "Robust detection of periodic time series measured from biological systems." BMC Bioinformatics **6**: 117.

Albright, S. R. and R. Tjian (2000). "TAFs revisited: more data reveal new twists and confirm old ideas." Gene **242**(1-2): 1-13.

Allen, J., H. M. Davey, D. Broadhurst*, et al.* (2003). "High-throughput classification of yeast mutants for functional genomics using metabolic footprinting." Nat Biotechnol **21**(6): 692-6.

Allen, T. A., S. Von Kaenel, J. A. Goodrich and J. F. Kugel (2004). "The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock." Nat Struct Mol Biol **11**(9): 816-21.

Aravin, A. A., N. M. Naumova, A. V. Tulin*, et al.* (2001). "Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline." Curr Biol **11**(13): 1017-27.

Argaman, L., R. Hershberg, J. Vogel*, et al.* (2001). "Novel small RNA-encoding genes in the intergenic regions of Escherichia coli." Curr Biol **11**(12): 941-50.

Bach, S., N. Talarek, T. Andrieu*, et al.* (2003). "Isolation of drugs active against mammalian prions using a yeast-based screening assay." Nat Biotechnol **21**(9): 1075-81.

Bader, G. D., A. Heilbut, B. Andrews*, et al.* (2003). "Functional genomics and proteomics: charting a multidimensional map of the yeast cell." Trends Cell Biol **13**(7): 344-56.

Bahler, J. (2005). "Cell-cycle control of gene expression in budding and fission yeast." Annu Rev Genet **39**: 69-94.

Bailly, E., S. Cabantous, D. Sondaz*, et al.* (2003). "Differential cellular localization among mitotic cyclins from Saccharomyces cerevisiae: a new role for the axial budding protein Bud3 in targeting Clb2 to the mother-bud neck." J Cell Sci **116**(Pt 20): 4119-30.

Bartel, D. P. (2004). "MicroRNAs: genomics, biogenesis, mechanism, and function." Cell **116**(2): 281-97.

Bass, B. L. (2002). "RNA editing by adenosine deaminases that act on RNA." Annu Rev Biochem **71**: 817-46.

Bean, J. M., E. D. Siggia and F. R. Cross (2006). "Coherence and timing of cell cycle start examined at single-cell resolution." Mol Cell **21**(1): 3-14.

Benjamini, Y. a. H., Y. (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." Journal of the Royal Statistical Society Series B **57**: 289– 300.

Bentwich, I., A. Avniel, Y. Karov*, et al.* (2005). "Identification of hundreds of conserved and nonconserved human microRNAs." Nat Genet **37**(7): 766-70.

Bernard, A., L. Ferhat, F. Dessi*, et al.* (1999). "Q/R editing of the rat GluR5 and GluR6 kainate receptors in vivo and in vitro: evidence for independent developmental, pathological and cellular regulation." Eur J Neurosci **11**(2): 604-16.

Bernstein, E., A. A. Caudy, S. M. Hammond and G. J. Hannon (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." Nature **409**(6818): 363-6.

Bertone, P., V. Stolc, T. E. Royce*, et al.* (2004). "Global identification of human transcribed sequences with genome tiling arrays." Science **306**(5705): 2242-6.

Bertone, P., V. Trifonov, J. S. Rozowsky*, et al.* (2006). "Design optimization methods for genomic DNA tiling arrays." Genome Res **16**(2): 271-81.

Bitko, V. and S. Barik (2001). "Phenotypic silencing of cytoplasmic genes using sequence-specific double-stranded short interfering RNA and its application in the reverse genetics of wild type negative-strand RNA viruses." BMC Microbiol **1**: 34.

Blackwood, D. H., A. Fordyce, M. T. Walker*, et al.* (2001). "Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family." Am J Hum Genet **69**(2): 428-33.

Bloom, J. and F. R. Cross (2007a). "Multiple levels of cyclin specificity in cell-cycle control." Nat Rev Mol Cell Biol **8**(2): 149-60.

Bloom, J. and F. R. Cross (2007b). "Novel role for Cdc14 sequestration: Cdc14 dephosphorylates factors that promote DNA replication." Mol Cell Biol **27**(3): 842-53.

Blumenthal, T. and K. S. Gleason (2003). "Caenorhabditis elegans operons: form and function." Nat Rev Genet **4**(2): 112-20.

Boer, V. M., J. H. de Winde, J. T. Pronk and M. D. Piper (2003). "The genome-wide transcriptional responses of Saccharomyces cerevisiae grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur." J Biol Chem **278**(5): 3265-74.

Bohmert, K., I. Camus, C. Bellini*, et al.* (1998). "AGO1 defines a novel locus of Arabidopsis controlling leaf development." Embo J **17**(1): 170-80.

Boisvert, F. M., S. van Koningsbruggen, J. Navascues and A. I. Lamond (2007). "The multifunctional nucleolus." Nat Rev Mol Cell Biol **8**(7): 574-85.

Bolstad, B. M., R. A. Irizarry, M. Astrand and T. P. Speed (2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias." Bioinformatics **19**(2): 185-93.

Bonoli, M., M. Graziola, V. Poggi and A. Hochkoeppler (2006). "RNA complementary to the 5' UTR of mRNA triggers effective silencing in Saccharomyces cerevisiae." Biochem Biophys Res Commun **339**(4): 1224-31.

Booher, R. N., R. J. Deshaies and M. W. Kirschner (1993). "Properties of Saccharomyces cerevisiae wee1 and its differential regulation of p34CDC28 in response to G1 and G2 cyclins." Embo J **12**(9): 3417-26.

Borsani, O., J. Zhu, P. E. Verslues*, et al.* (2005). "Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis." Cell **123**(7): 1279-91.

Boutz, P. L., G. Chawla, P. Stoilov and D. L. Black (2007). "MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development." Genes Dev **21**(1): 71-84.

Brantl, S. and E. G. Wagner (2000). "Antisense RNA-mediated transcriptional attenuation: an in vitro study of plasmid pT181." Mol Microbiol **35**(6): 1469-82.

Breeden, L. L. (1997). "Alpha-factor synchronization of budding yeast." Methods Enzymol **283**: 332-41.

Breeden, L. L. (2000). "Cyclin transcription: Timing is everything." Curr Biol **10**(16): R586-8.

Burley, S. K. and R. G. Roeder (1996). "Biochemistry and structural biology of transcription factor IID (TFIID)." Annu Rev Biochem **65**: 769-99.

Burton, J. L. and M. J. Solomon (2001). "D box and KEN box motifs in budding yeast Hsl1p are required for APC-mediated degradation and direct binding to Cdc20p and Cdh1p." Genes Dev **15**(18): 2381-95.

Butty, A. C., P. M. Pryciak, L. S. Huang*, et al.* (1998). "The role of Far1p in linking the heterotrimeric G protein to polarity establishment proteins during yeast mating." Science **282**(5393): 1511-6.

Calderon, A. J. and J. A. Lavergne (2005). "RNA interference: a novel and physiologic mechanism of gene silencing with great therapeutic potential." P R Health Sci J **24**(1): 27-33.

Calin, G. A. and C. M. Croce (2006). "Genomics of chronic lymphocytic leukemia microRNAs as new players with clinical significance." Semin Oncol **33**(2): 167-73.

Camblong, J., N. Iglesias, C. Fickentscher*, et al.* (2007). "Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae." Cell **131**(4): 706-17.

Carninci, P., T. Kasukawa, S. Katayama*, et al.* (2005). "The transcriptional landscape of the mammalian genome." Science **309**(5740): 1559-63.

Castrillo, J. I., A. Hayes, S. Mohammed*, et al.* (2003). "An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry." Phytochemistry **62**(6): 929-37.

Castrillo, J. I. and S. G. Oliver (2004). "Yeast as a touchstone in post-genomic research: strategies for integrative analysis in functional genomics." J Biochem Mol Biol **37**(1): 93-106.

Catt, K. J., J. P. Harwood, G. Aguilera and M. L. Dufau (1979). "Hormonal regulation of peptide receptors and target cell responses." Nature **280**(5718): 109-16.

Cavaille, J., K. Buiting, M. Kiefmann*, et al.* (2000). "Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization." Proc Natl Acad Sci U S A **97**(26): 14311-6.

Cawley, S., S. Bekiranov, H. H. Ng*, et al.* (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.

Cerutti, H. (2003). "RNA interference: traveling in the cell and gaining functions?" Trends Genet **19**(1): 39-46.

Chalkley, G. E. and C. P. Verrijzer (1999). "DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator." Embo J **18**(17): 4835-45.

Chan, S. W., D. Zilberman, Z. Xie*, et al.* (2004). "RNA silencing genes control de novo DNA methylation." Science **303**(5662): 1336.

Chang, F. and M. Peter (2003). "Yeasts make their mark." Nat Cell Biol **5**(4): 294-9.

Chen, C. N., L. Porubleva, G. Shearer*, et al.* (2003). "Associating protein activities with their genes: rapid identification of a gene encoding a methylglyoxal reductase in the yeast Saccharomyces cerevisiae." Yeast **20**(6): 545-54.

Chen, J., M. Sun, L. D. Hurst*, et al.* (2005). "Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts." Trends Genet **21**(6): 326-9.

Chen, J., M. Sun, W. J. Kent*, et al.* (2004a). "Over 20% of human transcripts might form sense-antisense pairs." Nucleic Acids Res **32**(16): 4812-20.

Chen, K. C., A. Csikasz-Nagy, B. Gyorffy*, et al.* (2000). "Kinetic analysis of a molecular model of the budding yeast cell cycle." Mol Biol Cell **11**(1): 369-91.

Chen, S., A. Zhang, L. B. Blyn and G. Storz (2004b). "MicC, a second small-RNA regulator of Omp protein expression in Escherichia coli." J Bacteriol **186**(20): 6689-97.

Cheng, J., P. Kapranov, J. Drenkow*, et al.* (2005). "Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution." Science **308**(5725): 1149-54.

Cho, R. J., M. J. Campbell, E. A. Winzeler*, et al.* (1998). "A genome-wide transcriptional analysis of the mitotic cell cycle." Mol Cell **2**(1): 65-73.

Ciejek, E. and J. Thorner (1979). "Recovery of S. cerevisiae a cells from G1 arrest by alpha factor pheromone requires endopeptidase action." Cell **18**(3): 623-35.

Cismowski, M. J., Laff, G.M., Solomon, M.J., Reed, S.I. (1995). "KIN28 encodes a C-terminal domain kinase that controls mRNA transcription in Saccharomyces cerevisiae but lacks cyclin-dependent kinase-activating kinase (CAK) activity." Molecular and cellular biology **15**(6): 2983-92.

Clancy, M. J., M. E. Shambaugh, C. S. Timpte and J. A. Bokar (2002). "Induction of sporulation in Saccharomyces cerevisiae leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene." Nucleic Acids Res **30**(20): 4509-18.

Clark, B. F. (2006). "The crystal structure of tRNA." J Biosci **31**(4): 453-7.

Clemenson, C. and M. C. Marsolier-Kergoat (2006). "The spindle assembly checkpoint regulates the phosphorylation state of a subset of DNA checkpoint proteins in Saccharomyces cerevisiae." Mol Cell Biol **26**(24): 9149-61.

Clouet d'Orval, B., M. L. Bortolin, C. Gaspin and J. P. Bachellerie (2001). "Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNATrp intron guides the formation of two ribose-methylated nucleosides in the mature tRNATrp." Nucleic Acids Res **29**(22): 4518-29.

Cohen, B. A., R. D. Mitra, J. D. Hughes and G. M. Church (2000). "A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression." Nat Genet **26**(2): 183-6.

Cook, P. R. (2003). "Nongenic transcription, gene regulation and action at a distance." J Cell Sci **116**(Pt 22): 4483-91.

Cooper, S. and K. Shedden (2003). "Microarray analysis of gene expression during the cell cycle." Cell Chromosome **2**(1): 1.

Cornell, M., N. W. Paton, C. Hedeler*, et al.* (2003). "GIMS: an integrated data storage and analysis environment for genomic and functional data." Yeast **20**(15): 1291-306.

Costa, F. F. (2005). "Non-coding RNAs: new players in eukaryotic biology." Gene **357**(2): 83-94.

Cox, J. M., M. M. Hayward, J. F. Sanchez*, et al.* (1997). "Bidirectional binding of the TATA box binding protein to the TATA box." Proc Natl Acad Sci U S A **94**(25): 13475-80.

Crick, F. H. (1958). "On protein synthesis." Symp Soc Exp Biol **12**: 138-63.

Cross, F. (1995). "Transcriptional regulation by a cyclin-cdk." Trends Genet **11**(6): 209-11.

Dahary, D., O. Elroy-Stein and R. Sorek (2005). "Naturally occurring antisense: transcriptional leakage or real overlap?" Genome Res **15**(3): 364-8.

Darzacq, X., B. E. Jady, C. Verheggen*, et al.* (2002). "Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs." Embo J **21**(11): 2746-56.

David, L., W. Huber, M. Granovskaia*, et al.* (2006). "A high-resolution map of transcription in the yeast genome." Proc Natl Acad Sci U S A **103**(14): 5320-5.

Davies, W., A. R. Isles, T. Humby and L. S. Wilkinson (2007). "What Are Imprinted Genes Doing in the Brain?" Epigenetics **2**(4).

Davies, W., A. R. Isles and L. S. Wilkinson (2005). "Imprinted gene expression in the brain." Neurosci Biobehav Rev **29**(3): 421-30.

de Lichtenberg, U., L. J. Jensen, S. Brunak and P. Bork (2005a). "Dynamic complex formation during the yeast cell cycle." Science **307**(5710): 724-7.

de Lichtenberg, U., L. J. Jensen, A. Fausboll*, et al.* (2005b). "Comparison of computational methods for the identification of cell cycle-regulated genes." Bioinformatics **21**(7): 1164-71.

de Lichtenberg, U., T. S. Jensen, L. J. Jensen and S. Brunak (2003). "Protein feature based identification of cell cycle regulated proteins in yeast." J Mol Biol **329**(4): 663-74.

Decatur, W. A. and M. J. Fournier (2002). "rRNA modifications and ribosome function." Trends Biochem Sci **27**(7): 344-51.

DeCerbo, J. and G. G. Carmichael (2005). "SINEs point to abundant editing in the human genome." Genome Biol **6**(4): 216.

DeChiara, T. M. and J. Brosius (1987). "Neural BC1 RNA: cDNA clones reveal nonrepetitive sequence content." Proc Natl Acad Sci U S A **84**(9): 2624-8.

Delihas, N. and S. Forst (2001). "MicF: an antisense RNA gene involved in response of Escherichia coli to global stress factors." J Mol Biol **313**(1): 1-12.

Delneri, D., I. Colson, S. Grammenoudi*, et al.* (2003). "Engineering evolution to study speciation in yeasts." Nature **422**(6927): 68-72.

Dennis, P. P. and A. Omer (2005). "Small non-coding RNAs in Archaea." Curr Opin Microbiol **8**(6): 685-94.

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

Dirick, L., T. Bohm and K. Nasmyth (1995). "Roles and regulation of Cln-Cdc28 kinases at the start of the cell cycle of Saccharomyces cerevisiae." Embo J **14**(19): 4803-13.

Dong, X., K. H. Zavitz, B. J. Thomas*, et al.* (1997). "Control of G1 in the developing Drosophila eye: rca1 regulates Cyclin A." Genes Dev **11**(1): 94-105.

Dostie, J., Z. Mourelatos, M. Yang*, et al.* (2003). "Numerous microRNPs in neuronal cells containing novel microRNAs." Rna **9**(2): 180-6.

Dubey, A. K., C. S. Baker, K. Suzuki*, et al.* (2003). "CsrA regulates translation of the Escherichia coli carbon starvation gene, cstA, by blocking ribosome access to the cstA transcript." J Bacteriol **185**(15): 4450-60.

Dudley, N. R., J. C. Labbe and B. Goldstein (2002). "Using RNA interference to identify genes required for RNA interference." Proc Natl Acad Sci U S A **99**(7): 4191-6.

Duhring, U., I. M. Axmann, W. R. Hess and A. Wilde (2006). "An internal antisense RNA regulates expression of the photosynthesis gene isiA." Proc Natl Acad Sci U S A **103**(18): 7054-8.

Eddy, S. R. (2001). "Non-coding RNA genes and the modern RNA world." Nat Rev Genet **2**(12): 919-29.

Eguchi, Y. and J. Tomizawa (1991). "Complexes formed by complementary RNA stem-loops. Their formations, structures and interaction with ColE1 Rom protein." J Mol Biol **220**(4): 831-42.

Eisen, M. B., P. T. Spellman, P. O. Brown and D. Botstein (1998). "Cluster analysis and display of genome-wide expression patterns." Proc Natl Acad Sci U S A **95**(25): 14863-8.

Elbashir, S. M., W. Lendeckel and T. Tuschl (2001a). "RNA interference is mediated by 21- and 22-nucleotide RNAs." Genes Dev **15**(2): 188-200.

Elbashir, S. M., J. Martinez, A. Patkaniowska*, et al.* (2001b). "Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate." Embo J **20**(23): 6877-88.

Espinoza, C. A., T. A. Allen, A. R. Hieb*, et al.* (2004). "B2 RNA binds directly to RNA polymerase II to repress transcript synthesis." Nat Struct Mol Biol **11**(9): 822-9.

Euskirchen, G. (2004). "Integrative approaches in molecular medicine." Pharmacogenomics **5**(4): 357-60.

Fernandez-Bellot, E. and C. Cullin (2001). "The protein-only theory and the yeast Saccharomyces cerevisiae: the prions and the propagons." Cell Mol Life Sci **58**(12-13): 1857-78.

Ficarro, S. B., M. L. McCleland, P. T. Stukenberg*, et al.* (2002). "Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae." Nat Biotechnol **20**(3): 301-5.

Fiehn, O. and W. Weckwerth (2003). "Deciphering metabolic networks." Eur J Biochem **270**(4): 579-88.

Finkelstein, D. B. and S. Strausberg (1979). "Metabolism of alpha-factor by a mating type cells of Saccharomyces cerevisiae." J Biol Chem **254**(3): 796-803.

Fire, A., S. Xu, M. K. Montgomery*, et al.* (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-11.

Folichon, M., V. Arluison, O. Pellegrini*, et al.* (2003). "The poly(A) binding protein Hfq protects RNA from RNase E and exoribonucleolytic degradation." Nucleic Acids Res **31**(24): 7302-10.

Fratkin, E., B. T. Naughton, D. L. Brutlag and S. Batzoglou (2006). "MotifCut: regulatory motifs finding with maximum density subgraphs." Bioinformatics **22**(14): e150-7.

Furuno, M., K. C. Pang, N. Ninomiya*, et al.* (2006). "Clusters of internally primed transcripts reveal novel long noncoding RNAs." PLoS Genet **2**(4): e37.

Ganot, P., M. L. Bortolin and T. Kiss (1997). "Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs." Cell **89**(5): 799-809.

Garrels, J. I., C. S. McLaughlin, J. R. Warner*, et al.* (1997). "Proteome studies of Saccharomyces cerevisiae: identification and characterization of abundant proteins." Electrophoresis **18**(8): 1347-60.

Gauthier, N. P., M. E. Larsen, R. Wernersson*, et al.* (2008). "Cyclebase.org--a comprehensive multi-organism online database of cell-cycle experiments." Nucleic Acids Res **36**(Database issue): D854-9.

Gavin, A. C., M. Bosche, R. Krause*, et al.* (2002). "Functional organization of the yeast proteome by systematic analysis of protein complexes." Nature **415**(6868): 141-7.

Geissmann, T. A. and D. Touati (2004). "Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator." Embo J **23**(2): 396-405.

Gentleman, R., Carey, V., Huber, W., Irizarry, R. & Dudoit, S. (2005). Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Heidelberg, Springer.

Gerber, A. P., D. Herschlag and P. O. Brown (2004). "Extensive association of functionally and cytotopically related mRNAs with Puf family RNA-binding proteins in yeast." PLoS Biol **2**(3): E79.

Gerdes, K., J. S. Jacobsen and T. Franch (1997). "Plasmid stabilization by post-segregational killing." Genet Eng (N Y) **19**: 49-61.

Ghaemmaghami, S., W. K. Huh, K. Bower*, et al.* (2003). "Global analysis of protein expression in yeast." Nature **425**(6959): 737-41.

Giaever, G., D. D. Shoemaker, T. W. Jones*, et al.* (1999). "Genomic profiling of drug sensitivities via induced haploinsufficiency." Nat Genet **21**(3): 278-83.

Gilad, Y., S. A. Rifkin, P. Bertone*, et al.* (2005). "Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles." Genome Res **15**(5): 674-80.

Gingeras, T. R. (2007). "Origin of phenotypes: genes and transcripts." Genome Res **17**(6): 682-90.

Goffeau, A. (1996). "1996: a vintage year for yeast and Yeast." Yeast **12**(16): 1603-5.

Goldberg, I. H., M. Rabinowitz and E. Reich (1962). "Basis of actinomycin action. I. DNA binding and inhibition of RNA-polymerase synthetic reactions by actinomycin." Proc Natl Acad Sci U S A **48**: 2094-101.

Goldberg, I. H., M. Rabinowitz and E. Reich (1963). "Basis of actinomycin action. II. Effect of actinomycin on the nucleoside triphosphate-inorganic pyrophosphate exchange." Proc Natl Acad Sci U S A **49**: 226-9.

Gottesman, S. (2002). "Stealth regulation: biological circuits with small RNA switches." Genes Dev **16**(22): 2829-42.

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

Graber, J. H., C. R. Cantor, S. C. Mohr and T. F. Smith (1999). "Genomic detection of new yeast pre-mRNA 3'-end-processing signals." Nucleic Acids Res **27**(3): 888-94.

Graber, J. H., G. D. McAllister and T. F. Smith (2002). "Probabilistic prediction of Saccharomyces cerevisiae mRNA 3'-processing sites." Nucleic Acids Res **30**(8): 1851-8.

Grad, Y., J. Aach, G. D. Hayes*, et al.* (2003). "Computational and experimental identification of C. elegans microRNAs." Mol Cell **11**(5): 1253-63.

Green, S. R. and A. D. Johnson (2004). "Promoter-dependent roles for the Srb10 cyclin-dependent kinase and the Hda1 deacetylase in Tup1-mediated repression in Saccharomyces cerevisiae." Mol Biol Cell **15**(9): 4191-202.

Griffin, T. J., S. P. Gygi, T. Ideker*, et al.* (2002). "Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae." Mol Cell Proteomics **1**(4): 323-33.

Grishok, A., A. E. Pasquinelli, D. Conte*, et al.* (2001). "Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing." Cell **106**(1): 23-34.

Grishok, A., J. L. Sinskey and P. A. Sharp (2005). "Transcriptional silencing of a transgene by RNAi in the soma of C. elegans." Genes Dev **19**(6): 683-96.

Grishok, A., H. Tabara and C. C. Mello (2000). "Genetic requirements for inheritance of RNAi in C. elegans." Science **287**(5462): 2494-7.

Gubler, U. (1987a). "Second-strand cDNA synthesis: classical method." Methods Enzymol **152**: 325-9.

Gubler, U. (1987b). "Second-strand cDNA synthesis: mRNA fragments as primers." Methods Enzymol **152**: 330-5.

Haasch, D., Y. W. Chen, R. M. Reilly*, et al.* (2002). "T cell activation induces a noncoding RNA transcript sensitive to inhibition by immunosuppressant drugs and encoded by the proto-oncogene, BIC." Cell Immunol **217**(1-2): 78-86.

Hagiya, M., K. Yoshida and N. Yanagishima (1977). "The release of sex-specific substances responsible for sexual agglutination from haploid cells of Saccharomyces cerevisiae." Exp Cell Res **104**(2): 263-72.

Haley, B. and P. D. Zamore (2004). "Kinetic analysis of the RNAi enzyme complex." Nat Struct Mol Biol **11**(7): 599-606.

Hall, I. M., G. D. Shankaranarayana, K. Noma*, et al.* (2002). "Establishment and maintenance of a heterochromatin domain." Science **297**(5590): 2232-7.

Hamilton, A. J. and D. C. Baulcombe (1999). "A species of small antisense RNA in posttranscriptional gene silencing in plants." Science **286**(5441): 950-2.

Hammond, S. M., E. Bernstein, D. Beach and G. J. Hannon (2000). "An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells." Nature **404**(6775): 293-6.

Han, Y. and D. Grierson (2002). "Relationship between small antisense RNAs and aberrant RNAs associated with sense transgene mediated gene silencing in tomato." Plant J **29**(4): 509-19.

Hanna, J. S., E. S. Kroll, V. Lundblad and F. A. Spencer (2001). "Saccharomyces cerevisiae CTF18 and CTF4 are required for sister chromatid cohesion." Mol Cell Biol **21**(9): 3144-58.

Hannon, G. J. (2002). "RNA interference." Nature **418**(6894): 244-51.

Harbison, C. T., D. B. Gordon, T. I. Lee*, et al.* (2004). "Transcriptional regulatory code of a eukaryotic genome." Nature **431**(7004): 99-104.

Hartwell, L. H., L. Hood, M. L. Goldberg*, et al.* (2004a). Genetics: From Genes to Genomes, McGraw-Hill: New York, NY.**:** 264.

Hartwell, L. H., L. Hood, M. L. Goldberg*, et al.* (2004b). Genetics: From Genes to Genomes, McGraw-Hill: New York.

Harvey, S. L., A. Charlet, W. Haas*, et al.* (2005). "Cdk1-dependent regulation of the mitotic inhibitor Wee1." Cell **122**(3): 407-20.

Hastings, M. L., C. Milcarek, K. Martincic*, et al.* (1997). "Expression of the thyroid hormone receptor gene, erbAalpha, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels." Nucleic Acids Res **25**(21): 4296-300.

Hayes, A., N. Zhang, J. Wu*, et al.* (2002). "Hybridization array technology coupled with chemostat culture: Tools to interrogate gene expression in Saccharomyces cerevisiae." Methods **26**(3): 281-90.

He, F., X. Li, P. Spatrick*, et al.* (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, L., J. M. Thomson, M. T. Hemann*, et al.* (2005). "A microRNA polycistron as a potential human oncogene." Nature **435**(7043): 828-33.

Heidrich, N. and S. Brantl (2003). "Antisense-RNA mediated transcriptional attenuation: importance of a U-turn loop structure in the target RNA of plasmid pIP501 for efficient inhibition by the antisense RNA." J Mol Biol **333**(5): 917-29.

Heix, J., A. Vente, R. Voit*, et al.* (1998). "Mitotic silencing of human rRNA synthesis: inactivation of the promoter selectivity factor SL1 by cdc2/cyclin B-mediated phosphorylation." Embo J **17**(24): 7373-81.

Hekstra, D., A. R. Taussig, M. Magnasco and F. Naef (2003). "Absolute mRNA concentrations from sequence-specific calibration of oligonucleotide arrays." Nucleic Acids Res **31**(7): 1962-8.

Hereford, L. M., M. A. Osley, T. R. Ludwig, 2nd and C. S. McLaughlin (1981). "Cell-cycle regulation of yeast histone mRNA." Cell **24**(2): 367-75.

Hild, M., B. Beckmann, S. A. Haas*, et al.* (2003). "An integrated gene annotation and transcriptional profiling approach towards the full gene content of the Drosophila genome." Genome Biol **5**(1): R3.

Hiley, S. L., J. Jackman, T. Babak*, et al.* (2005). "Detection and discovery of RNA modifications using microarrays." Nucleic Acids Res **33**(1): e2.

Hinnebusch, A. G. (2005). "Translational regulation of GCN4 and the general amino acid control of yeast." Annu Rev Microbiol **59**: 407-50.

Hinsby, A. M., L. Kiemer, E. O. Karlberg*, et al.* (2006). "A wiring of the human nucleolus." Mol Cell **22**(2): 285-95.

Ho, Y., A. Gruhler, A. Heilbut*, et al.* (2002). "Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry." Nature **415**(6868): 180-3.

Holen, T., M. Amarzguioui, M. T. Wiiger*, et al.* (2002). "Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor." Nucleic Acids Res **30**(8): 1757-66.

Hollander, M. C., I. Alamo and A. J. Fornace, Jr. (1996). "A novel DNA damage-inducible transcript, gadd7, inhibits cell growth, but lacks a protein product." Nucleic Acids Res **24**(9): 1589-93.

Hongay, C. F., P. L. Grisafi, T. Galitski and G. R. Fink (2006). "Antisense transcription controls cell fate in Saccharomyces cerevisiae." Cell **127**(4): 735-45.

Horak, C. E., N. M. Luscombe, J. Qian*, et al.* (2002). "Complex transcriptional circuitry at the G1/S transition in Saccharomyces cerevisiae." Genes Dev **16**(23): 3017-33.

Horak, C. E. and M. Snyder (2002a). "ChIP-chip: a genomic approach for identifying transcription factor binding sites." Methods Enzymol **350**: 469-83.

Horak, C. E. and M. Snyder (2002b). "Global analysis of gene expression in yeast." Funct Integr Genomics **2**(4-5): 171-80.

Hornstein, E. and N. Shomron (2006). "Canalization of development by microRNAs." Nat Genet **38 Suppl**: S20-4.

Howitz, K. T., K. J. Bitterman, H. Y. Cohen*, et al.* (2003). "Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan." Nature **425**(6954): 191-6.

Hu, G. K., S. J. Madore, B. Moldover*, et al.* (2001). "Predicting splice variant from DNA chip expression data." Genome Res **11**(7): 1237-45.

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

Huber, W., A. von Heydebreck, H. Sultmann*, et al.* (2002). "Variance stabilization applied to microarray data calibration and to the quantification of differential expression." Bioinformatics **18 Suppl 1**: S96-104.

Hughes, J. D., P. W. Estep, S. Tavazoie and G. M. Church (2000). "Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae." J Mol Biol **296**(5): 1205-14.

Huh, W. K., J. V. Falvo, L. C. Gerke*, et al.* (2003). "Global analysis of protein localization in budding yeast." Nature **425**(6959): 686-91.

Hurowitz, E. H. and P. O. Brown (2003). "Genome-wide analysis of mRNA lengths in Saccharomyces cerevisiae." Genome Biol **5**(1): R2.

Huttenhofer, A., J. Brosius and J. P. Bachellerie (2002). "RNomics: identification and function of small, non-messenger RNAs." Curr Opin Chem Biol **6**(6): 835-43.

Huttenhofer, A., M. Kiefmann, S. Meier-Ewert*, et al.* (2001). "RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse." Embo J **20**(11): 2943-53.

Huttenhofer, A. and J. Vogel (2006). "Experimental approaches to identify non-coding RNAs." Nucleic Acids Res **34**(2): 635-46.

Hutvagner, G. and P. D. Zamore (2002). "A microRNA in a multiple-turnover RNAi enzyme complex." Science **297**(5589): 2056-60.

Inada, M. and C. Guthrie (2004). "Identification of Lhp1p-associated RNAs by microarray analysis in Saccharomyces cerevisiae reveals association with coding and noncoding RNAs." Proc Natl Acad Sci U S A **101**(2): 434-9.

Ishizuka, A., M. C. Siomi and H. Siomi (2002). "A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins." Genes Dev **16**(19): 2497-508.

Ito, T., T. Chiba, R. Ozawa*, et al.* (2001). "A comprehensive two-hybrid analysis to explore the yeast protein interactome." Proc Natl Acad Sci U S A **98**(8): 4569-74.

Itoh, T. and J. Tomizawa (1980). "Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonuclease H." Proc Natl Acad Sci U S A **77**(5): 2450-4.

Iyer, V. R., C. E. Horak, C. S. Scafe*, et al.* (2001). "Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF." Nature **409**(6819): 533-8.

Jady, B. E. and T. Kiss (2000). "Characterisation of the U83 and U84 small nucleolar RNAs: two novel 2'-O-ribose methylation guide RNAs that lack complementarities to ribosomal RNAs." Nucleic Acids Res **28**(6): 1348-54.

Jang, J. K., L. Messina, M. B. Erdman*, et al.* (1995). "Induction of metaphase arrest in Drosophila oocytes by chiasma-based kinetochore tension." Science **268**(5219): 1917-9.

Jansen, R. and M. Gerstein (2000). "Analysis of the yeast transcriptome with structural and functional categories: characterizing highly expressed proteins." Nucleic Acids Res **28**(6): 1481-8.

Jeong, B. H., N. H. Kim, E. K. Choi*, et al.* (2005). "Polymorphism at 3' UTR +28 of the prion-like protein gene is associated with sporadic Creutzfeldt-Jakob disease." Eur J Hum Genet **13**(9): 1094-7.

Johansson, D., P. Lindgren and A. Berglund (2003). "A multivariate approach applied to microarray data for identification of genes with cell cycle-coupled transcription." Bioinformatics **19**(4): 467-73.

Johnson, J. M., S. Edwards, D. Shoemaker and E. E. Schadt (2005). "Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments." Trends Genet **21**(2): 93-102.

Jopling, C. L., M. Yi, A. M. Lancaster*, et al.* (2005). "Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA." Science **309**(5740): 1577-81.

Jorgensen, P., I. Rupes, J. R. Sharom*, et al.* (2004). "A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size." Genes Dev **18**(20): 2491-505.

Kalidas, S. and D. P. Smith (2002). "Novel genomic cDNA hybrids produce effective RNA interference in adult Drosophila." Neuron **33**(2): 177-84.

Kampa, D., J. Cheng, P. Kapranov*, et al.* (2004). "Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22." Genome Res **14**(3): 331-42.

Kapranov, P., S. E. Cawley, J. Drenkow*, et al.* (2002). "Large-scale transcriptional activity in chromosomes 21 and 22." Science **296**(5569): 916-9.

Kapranov, P., J. Drenkow, J. Cheng*, et al.* (2005). "Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays." Genome Res **15**(7): 987-97.

Kapranov, P., V. I. Sementchenko and T. R. Gingeras (2003). "Beyond expression profiling: next generation uses of high density oligonucleotide arrays." Brief Funct Genomic Proteomic **2**(1): 47-56.

Katayama, S., Y. Tomaru, T. Kasukawa*, et al.* (2005). "Antisense transcription in the mammalian transcriptome." Science **309**(5740): 1564-6.

Kawano, M., T. Oshima, H. Kasai and H. Mori (2002). "Molecular characterization of long direct repeat (LDR) sequences expressing a stable mRNA encoding for a 35-amino-acid cell-killing peptide and a cis-encoded small antisense RNA in Escherichia coli." Mol Microbiol **45**(2): 333-49.

Kellis, M., N. Patterson, M. Endrizzi*, et al.* (2003). "Sequencing and comparison of yeast species to identify genes and regulatory elements." Nature **423**(6937): 241-54.

Kennerdell, J. R. and R. W. Carthew (1998). "Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway." Cell **95**(7): 1017-26.

Kenzelmann, M., K. Rippe and J. S. Mattick (2006). "RNA: Networks & Imaging." Mol Syst Biol **2**: 44.

Kessler, M. M., Q. Zeng, S. Hogan*, et al.* (2003). "Systematic discovery of new genes in the Saccharomyces cerevisiae genome." Genome Res **13**(2): 264-71.

Kim, V. N. (2005). "MicroRNA biogenesis: coordinated cropping and dicing." Nat Rev Mol Cell Biol **6**(5): 376-85.

King, T. H., B. Liu, R. R. McCully and M. J. Fournier (2003). "Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center." Mol Cell **11**(2): 425-35.

Kishore, S. and S. Stamm (2006). "The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C." Science **311**(5758): 230-2.

Kiss, T. (2002). "Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions." Cell **109**(2): 145-8.

Kiss-Laszlo, Z. and T. Hohn (1996). "Pararetro- and retrovirus RNA: splicing and the control of nuclear export." Trends Microbiol **4**(12): 480-5.

Knee, R. and P. R. Murphy (1997). "Regulation of gene expression by natural antisense RNA transcripts." Neurochem Int **31**(3): 379-92.

Koch, C. and K. Nasmyth (1994). "Cell cycle regulated transcription in yeast." Curr Opin Cell Biol **6**(3): 451-9.

Kohr, G., T. Melcher and P. H. Seeburg (1998). "Candidate editases for GluR channels in single neurons of rat hippocampus and cerebellum." Neuropharmacology **37**(10-11): 1411-7.

Kong, Y. and J. H. Han (2005). "MicroRNA: biological and computational perspective." Genomics Proteomics Bioinformatics **3**(2): 62-72.

Koob, M. D., M. L. Moseley, L. J. Schut*, et al.* (1999). "An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8)." Nat Genet **21**(4): 379-84.

Kruglyak, S. and H. Tang (2000). "Regulation of adjacent yeast genes." Trends Genet **16**(3): 109-11.

Kryndushkin, D. S., I. M. Alexandrov, M. D. Ter-Avanesyan and V. V. Kushnirov (2003). "Yeast [PSI+] prion aggregates are formed by small Sup35 polymers fragmented by Hsp104." J Biol Chem **278**(49): 49636-43.

Kuersten, S. and E. B. Goodwin (2003). "The power of the 3' UTR: translational control and development." Nat Rev Genet **4**(8): 626-37.

Kumar, M. and G. G. Carmichael (1998). "Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes." Microbiol Mol Biol Rev **62**(4): 1415-34.

Kurihara, Y. and Y. Watanabe (2004). "Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions." Proc Natl Acad Sci U S A **101**(34): 12753-8.

Kuwabara, T., J. Hsieh, K. Nakashima*, et al.* (2004). "A small modulatory dsRNA specifies the fate of adult neural stem cells." Cell **116**(6): 779-93.

Kwek, K. Y., S. Murphy, A. Furger*, et al.* (2002). "U1 snRNA associates with TFIIH and regulates transcriptional initiation." Nat Struct Biol **9**(11): 800-5.
Laabs, T. L., D. D. Markwardt, M. G. Slattery*, et al.* (2003). "ACE2 is required for daughter cell-specific G1 delay in Saccharomyces cerevisiae." Proc Natl Acad Sci U S A **100**(18): 10275-80.
Lachner, M. and T. Jenuwein (2002). "The many faces of histone lysine methylation." Curr Opin Cell Biol **14**(3): 286-98.
Lafonataine, D. J. and D. Tollervey (2001). Ribosomal RNA. Encyclopedia of Life Sciences**:** 1-7.
Lagos-Quintana, M., R. Rauhut, W. Lendeckel and T. Tuschl (2001). "Identification of novel genes coding for small expressed RNAs." Science **294**(5543): 853-8.
Lander, E. S., L. M. Linton, B. Birren*, et al.* (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.
Lanz, R. B., N. J. McKenna, S. A. Onate*, et al.* (1999). "A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex." Cell **97**(1): 17-27.
Lapidot, M. and Y. Pilpel (2006). "Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms." EMBO Rep **7**(12): 1216-22.
Lashkari, D. A., J. L. DeRisi, J. H. McCusker*, et al.* (1997). "Yeast microarrays for genome wide parallel genetic and gene expression analysis." Proc Natl Acad Sci U S A **94**(24): 13057-62.
Lau, N. C., L. P. Lim, E. G. Weinstein and D. P. Bartel (2001). "An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans." Science **294**(5543): 858-62.
Lavorgna, G., D. Dahary, B. Lehner*, et al.* (2004). "In search of antisense." Trends Biochem Sci **29**(2): 88-94.
Law, R. H. and R. J. Devenish (1988). "Expression in yeast of antisense RNA to ADE1 mRNA." Biochem Int **17**(4): 673-9.
Lease, R. A. and M. Belfort (2000). "A trans-acting RNA as a control switch in Escherichia coli: DsrA modulates function by forming alternative structures." Proc Natl Acad Sci U S A **97**(18): 9919-24.
Lease, R. A., M. E. Cusick and M. Belfort (1998). "Riboregulation in Escherichia coli: DsrA RNA acts by RNA:RNA interactions at multiple loci." Proc Natl Acad Sci U S A **95**(21): 12456-61.
Lee, R. C. and V. Ambros (2001). "An extensive class of small RNAs in Caenorhabditis elegans." Science **294**(5543): 862-4.
Lee, R. C., R. L. Feinbaum and V. Ambros (1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14." Cell **75**(5): 843-54.
Lee, T. I., N. J. Rinaldi, F. Robert*, et al.* (2002). "Transcriptional regulatory networks in Saccharomyces cerevisiae." Science **298**(5594): 799-804.
Lenz, D. H., K. C. Mok, B. N. Lilley*, et al.* (2004). "The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae." Cell **118**(1): 69-82.
Levine, K., K. Huang and F. R. Cross (1996). "Saccharomyces cerevisiae G1 cyclins differ in their intrinsic functional specificities." Mol Cell Biol **16**(12): 6794-803.
Levine, M. and R. Tjian (2003). "Transcription regulation and animal diversity." Nature **424**(6945): 147-51.
Lew, D. J., N. J. Marini and S. I. Reed (1993). "A suppressor of cln3 for size control." Cell **72**(4): 488-9.
Lew, D. J. and S. I. Reed (1993). "Morphogenesis in the yeast cell cycle: regulation by Cdc28 and cyclins." J Cell Biol **120**(6): 1305-20.
Lew, D. J. and S. I. Reed (1995a). "A cell cycle checkpoint monitors cell morphogenesis in budding yeast." J Cell Biol **129**(3): 739-49.
Lew, D. J. and S. I. Reed (1995b). "Cell cycle control of morphogenesis in budding yeast." Curr Opin Genet Dev **5**(1): 17-23.
Lewis, A. and W. Reik (2006). "How imprinting centres work." Cytogenet Genome Res **113**(1-4): 81-9.

Li, S. G., H. Zhou, Y. P. Luo*, et al.* (2005). "Identification and functional analysis of 20 Box H/ACA small nucleolar RNAs (snoRNAs) from Schizosaccharomyces pombe." J Biol Chem **280**(16): 16446-55.

Li, X. Y., S. R. Bhaumik and M. R. Green (2000). "Distinct classes of yeast promoters revealed by differential TAF recruitment." Science **288**(5469): 1242-4.

Li, Y. Y., L. Qin, Z. M. Guo*, et al.* (2006). "In silico discovery of human natural antisense transcripts." BMC Bioinformatics **7**: 18.

Lillie, S. H. and S. S. Brown (1994). "Immunofluorescence localization of the unconventional myosin, Myo2p, and the putative kinesin-related protein, Smy1p, to the same regions of polarized growth in Saccharomyces cerevisiae." J Cell Biol **125**(4): 825-42.

Lipke, P. N., A. Taylor and C. E. Ballou (1976). "Morphogenic effects of alpha-factor on Saccharomyces cerevisiae a cells." J Bacteriol **127**(1): 610-8.

Lipshutz, R. J., S. P. Fodor, T. R. Gingeras and D. J. Lockhart (1999). "High density synthetic oligonucleotide arrays." Nat Genet **21**(1 Suppl): 20-4.

Liu, A. Y., B. S. Torchia, B. R. Migeon and R. F. Siliciano (1997). "The human NTT gene: identification of a novel 17-kb noncoding nuclear RNA expressed in activated CD4+ T cells." Genomics **39**(2): 171-84.

Liu, J., M. A. Carmell, F. V. Rivas*, et al.* (2004). "Argonaute2 is the catalytic engine of mammalian RNAi." Science **305**(5689): 1437-41.

Llave, C., K. D. Kasschau, M. A. Rector and J. C. Carrington (2002a). "Endogenous and silencing-associated small RNAs in plants." Plant Cell **14**(7): 1605-19.

Llave, C., Z. Xie, K. D. Kasschau and J. C. Carrington (2002b). "Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA." Science **297**(5589): 2053-6.

Lockhart, D. J. and E. A. Winzeler (2000). "Genomics, gene expression and DNA arrays." Nature **405**(6788): 827-36.

Lovrics, A., A. Csikasz-Nagy, I. G. Zsely*, et al.* (2006). "Time scale and dimension analysis of a budding yeast cell cycle model." BMC Bioinformatics **7**: 494.

Lowe, T. M. and S. R. Eddy (1999). "A computational screen for methylation guide snoRNAs in yeast." Science **283**(5405): 1168-71.

Lu, J., G. Getz, E. A. Miska*, et al.* (2005). "MicroRNA expression profiles classify human cancers." Nature **435**(7043): 834-8.

Luan, Y. and H. Li (2004). "Model-based methods for identifying periodically expressed genes based on time course microarray gene expression data." Bioinformatics **20**(3): 332-9.

Lukowiak, A. A., A. Narayanan, Z. H. Li*, et al.* (2001). "The snoRNA domain of vertebrate telomerase RNA functions to localize the RNA within the nucleus." Rna **7**(12): 1833-44.

Maas, S., A. Rich and K. Nishikura (2003). "A-to-I RNA editing: recent news and residual mysteries." J Biol Chem **278**(3): 1391-4.

MacIsaac, K. D., T. Wang, D. B. Gordon*, et al.* (2006). "An improved map of conserved regulatory sites for Saccharomyces cerevisiae." BMC Bioinformatics **7**: 113.

Majdalani, N., S. Chen, J. Murrow*, et al.* (2001). "Regulation of RpoS by a novel small RNA: the characterization of RprA." Mol Microbiol **39**(5): 1382-94.

Majdalani, N., D. Hernandez and S. Gottesman (2002). "Regulation and mode of action of the second small RNA activator of RpoS translation, RprA." Mol Microbiol **46**(3): 813-26.

Majdalani, N., C. K. Vanderpool and S. Gottesman (2005). "Bacterial small RNA regulators." Crit Rev Biochem Mol Biol **40**(2): 93-113.

Makeyev, E. V., J. Zhang, M. A. Carrasco and T. Maniatis (2007). "The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing." Mol Cell **27**(3): 435-48.

Maness, P. F. and G. M. Edelman (1978). "Inactivation and chemical alteration of mating factor alpha by cells and spheroplasts of yeast." Proc Natl Acad Sci U S A **75**(3): 1304-1308.
Marc, P., A. Margeot, F. Devaux*, et al.* (2002). "Genome-wide analysis of mRNAs targeted to yeast mitochondria." EMBO Rep **3**(2): 159-64.
Marker, C., A. Zemann, T. Terhorst*, et al.* (2002). "Experimental RNomics: identification of 140 candidates for small non-messenger RNAs in the plant Arabidopsis thaliana." Curr Biol **12**(23): 2002-13.
Martens, J. A., L. Laprade and F. Winston (2004). "Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene." Nature **429**(6991): 571-4.
Martens, J. A., P. Y. Wu and F. Winston (2005). "Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae." Genes Dev **19**(22): 2695-704.
Martignetti, J. A. and J. Brosius (1993). "BC200 RNA: a neural RNA polymerase III product encoded by a monomeric Alu element." Proc Natl Acad Sci U S A **90**(24): 11563-7.
Martinez, J., A. Patkaniowska, H. Urlaub*, et al.* (2002). "Single-stranded antisense siRNAs guide target RNA cleavage in RNAi." Cell **110**(5): 563-74.
Martone, R., G. Euskirchen, P. Bertone*, et al.* (2003). "Distribution of NF-kappaB-binding sites across human chromosome 22." Proc Natl Acad Sci U S A **100**(21): 12247-52.
Masse, E. and S. Gottesman (2002). "A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli." Proc Natl Acad Sci U S A **99**(7): 4620-5.
Mathieu, O. and J. Bender (2004). "RNA-directed DNA methylation." J Cell Sci **117**(Pt 21): 4881-8.
Mattick, J. S. (2001). "Non-coding RNAs: the architects of eukaryotic complexity." EMBO Rep **2**(11): 986-91.
Mattick, J. S. (2003). "Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms." Bioessays **25**(10): 930-9.
Mattick, J. S. (2004). "RNA regulation: a new genetics?" Nat Rev Genet **5**(4): 316-23.
Mattick, J. S. and M. J. Gagen (2001). "The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms." Mol Biol Evol **18**(9): 1611-30.
Matzke, M., A. J. Matzke and J. M. Kooter (2001a). "RNA: guiding gene silencing." Science **293**(5532): 1080-3.
Matzke, M. A., A. J. Matzke, G. J. Pruss and V. B. Vance (2001b). "RNA-based silencing strategies in plants." Curr Opin Genet Dev **11**(2): 221-7.
McCutcheon, J. P. and S. R. Eddy (2003). "Computational identification of non-coding RNAs in Saccharomyces cerevisiae by comparative genomics." Nucleic Acids Res **31**(14): 4119-28.
McManus, M. T. and P. A. Sharp (2002). "Gene silencing in mammals by small interfering RNAs." Nat Rev Genet **3**(10): 737-47.
McMillan, J. N., R. A. Sia and D. J. Lew (1998). "A morphogenesis checkpoint monitors the actin cytoskeleton in yeast." J Cell Biol **142**(6): 1487-99.
Meguro, M., K. Mitsuya, N. Nomura*, et al.* (2001). "Large-scale evaluation of imprinting status in the Prader-Willi syndrome region: an imprinted direct repeat cluster resembling small nucleolar RNA genes." Hum Mol Genet **10**(4): 383-94.
Meister, G., M. Landthaler, A. Patkaniowska*, et al.* (2004). "Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs." Mol Cell **15**(2): 185-97.
Meister, G. and T. Tuschl (2004). "Mechanisms of gene silencing by double-stranded RNA." Nature **431**(7006): 343-9.
Mello, C. C. and D. Conte, Jr. (2004). "Revealing the world of RNA interference." Nature **431**(7006): 338-42.

Mendenhall, M. D. and A. E. Hodge (1998). "Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast Saccharomyces cerevisiae." Microbiol Mol Biol Rev **62**(4): 1191-243.

Mette, M. F., A. J. Matzke and M. A. Matzke (2001). "Resistance of RNA-mediated TGS to HC-Pro, a viral suppressor of PTGS, suggests alternative pathways for dsRNA processing." Curr Biol **11**(14): 1119-23.

Michael, M. Z., O. C. SM, N. G. van Holst Pellekaan*, et al.* (2003). "Reduced accumulation of specific microRNAs in colorectal neoplasia." Mol Cancer Res **1**(12): 882-91.

Michaud, G. A., M. Salcius, F. Zhou*, et al.* (2003). "Analyzing antibody specificity with whole proteome microarrays." Nat Biotechnol **21**(12): 1509-12.

Mignone, F., C. Gissi, S. Liuni and G. Pesole (2002). "Untranslated regions of mRNAs." Genome Biol **3**(3): REVIEWS0004.

Millar, J. K., J. C. Wilson-Annan, S. Anderson*, et al.* (2000). "Disruption of two novel genes by a translocation co-segregating with schizophrenia." Hum Mol Genet **9**(9): 1415-23.

Misra, S., M. A. Crosby, C. J. Mungall*, et al.* (2002). "Annotation of the Drosophila melanogaster euchromatic genome: a systematic review." Genome Biol **3**(12): RESEARCH0083.

Mitchell, J. R., J. Cheng and K. Collins (1999). "A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end." Mol Cell Biol **19**(1): 567-76.

Mitchell, J. R. and K. Collins (2000). "Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase." Mol Cell **6**(2): 361-71.

Miura, F., N. Kawaguchi, J. Sese*, et al.* (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.

Moll, I., T. Afonyushkin, O. Vytvytska*, et al.* (2003). "Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs." Rna **9**(11): 1308-14.

Moller, T., T. Franch, P. Hojrup*, et al.* (2002). "Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction." Mol Cell **9**(1): 23-30.

Moore, S. A. (1983). "Comparison of dose-response curves for alpha factor-induced cell division arrest, agglutination, and projection formation of yeast cells. Implication for the mechanism of alpha factor action." J Biol Chem **258**(22): 13849-56.

Mourelatos, Z., J. Dostie, S. Paushkin*, et al.* (2002). "miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs." Genes Dev **16**(6): 720-8.

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

Munroe, S. H. (2004). "Diversity of antisense regulation in eukaryotes: multiple mechanisms, emerging patterns." J Cell Biochem **93**(4): 664-71.

Munroe, S. H. and J. Zhu (2006). "Overlapping transcripts, double-stranded RNA and antisense regulation: a genomic perspective." Cell Mol Life Sci **63**(18): 2102-18.

Naef, F. and M. O. Magnasco (2003). "Solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays." Phys Rev E Stat Nonlin Soft Matter Phys **68**(1 Pt 1): 011906.

Nasmyth, K. (1993). "Control of the yeast cell cycle by the Cdc28 protein kinase." Curr Opin Cell Biol **5**(2): 166-79.

Nasmyth, K. (1996). "At the heart of the budding yeast cell cycle." Trends Genet **12**(10): 405-12.

Nasmyth, K. (1999). "Separating sister chromatids." Trends Biochem Sci **24**(3): 98-104.

Nasmyth, K. (2001). "Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis." Annu Rev Genet **35**: 673-745.

Nasr, F., A. M. Becam, S. C. Brown*, et al.* (1995). "Artificial antisense RNA regulation of YBR1012 (YBR136w), an essential gene from Saccharomyces cerevisiae which is important for progression through G1/S." Mol Gen Genet **249**(1): 51-7.

Navarro, P., S. Pichard, C. Ciaudo*, et al.* (2005). "Tsix transcription across the Xist gene alters chromatin conformation without affecting Xist transcription: implications for X-chromosome inactivation." Genes Dev **19**(12): 1474-84.

Nemes, J. P., K. A. Benzow, M. L. Moseley*, et al.* (2000). "The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1)." Hum Mol Genet **9**(10): 1543-51.

Nguyen, V. T., T. Kiss, A. A. Michels and O. Bensaude (2001). "7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes." Nature **414**(6861): 322-5.

Ni, J., A. L. Tien and M. J. Fournier (1997). "Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA." Cell **89**(4): 565-73.

Nicoloso, M., L. H. Qu, B. Michot and J. P. Bachellerie (1996). "Intron-encoded, antisense small nucleolar RNAs: the characterization of nine novel species points to their direct role as guides for the 2'-O-ribose methylation of rRNAs." J Mol Biol **260**(2): 178-95.

Niemitz, E. L., M. R. DeBaun, J. Fallon*, et al.* (2004). "Microdeletion of LIT1 in familial Beckwith-Wiedemann syndrome." Am J Hum Genet **75**(5): 844-9.

Novak, B., A. Csikasz-Nagy, B. Gyorffy*, et al.* (1998). "Model scenarios for evolution of the eukaryotic cell cycle." Philos Trans R Soc Lond B Biol Sci **353**(1378): 2063-76.

Novak, B., J. J. Tyson, B. Gyorffy and A. Csikasz-Nagy (2007). "Irreversible cell-cycle transitions are due to systems-level feedback." Nat Cell Biol **9**(7): 724-8.

Nykanen, A., B. Haley and P. D. Zamore (2001). "ATP requirements and small interfering RNA structure in the RNA interference pathway." Cell **107**(3): 309-21.

O'Donnell, K. A., E. A. Wentzel, K. I. Zeller*, et al.* (2005). "c-Myc-regulated microRNAs modulate E2F1 expression." Nature **435**(7043): 839-43.

Oehlen, L. J. and F. R. Cross (1994). "G1 cyclins CLN1 and CLN2 repress the mating factor response pathway at Start in the yeast cell cycle." Genes Dev **8**(9): 1058-70.

Oehlen, L. J., J. D. McKinney and F. R. Cross (1996). "Ste12 and Mcm1 regulate cell cycle-dependent transcription of FAR1." Mol Cell Biol **16**(6): 2830-7.

Ofengand, J. (2002). "Ribosomal RNA pseudouridines and pseudouridine synthases." FEBS Lett **514**(1): 17-25.

Ogawa, Y. and J. T. Lee (2002). "Antisense regulation in X inactivation and autosomal imprinting." Cytogenet Genome Res **99**(1-4): 59-65.

Oliva, A., A. Rosebrock, F. Ferrezuelo*, et al.* (2005). "The cell cycle-regulated genes of Schizosaccharomyces pombe." PLoS Biol **3**(7): e225.

Olivas, W. M., D. Muhlrad and R. Parker (1997). "Analysis of the yeast genome: identification of new non-coding and small ORF-containing RNAs." Nucleic Acids Res **25**(22): 4619-25.

Oliver, S. G. (2002). "Functional genomics: lessons from yeast." Philos Trans R Soc Lond B Biol Sci **357**(1417): 17-23.

Olsen, P. H. and V. Ambros (1999). "The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation." Dev Biol **216**(2): 671-80.

Olsson, L., M. E. Larsen, B. Ronnow*, et al.* (1997). "Silencing MIG1 in Saccharomyces cerevisiae: effects of antisense MIG1 expression and MIG1 gene disruption." Appl Environ Microbiol **63**(6): 2366-71.

Omer, A. D., S. Ziesche, W. A. Decatur*, et al.* (2003). "RNA-modifying machines in archaea." Mol Microbiol **48**(3): 617-29.

O'Neill, M. J. (2005). "The influence of non-coding RNAs on allele-specific gene expression in mammals." Hum Mol Genet **14 Spec No 1**: R113-20.
Opdyke, J. A., J. G. Kang and G. Storz (2004). "GadY, a small-RNA regulator of acid response genes in Escherichia coli." J Bacteriol **186**(20): 6698-705.
Osato, N., H. Yamada, K. Satoh*, et al.* (2003). "Antisense transcripts with rice full-length cDNAs." Genome Biol **5**(1): R5.
Osborn, M. J. and J. R. Miller (2007). "Rescuing yeast mutants with human genes." Brief Funct Genomic Proteomic **6**(2): 104-11.
O'Shea, E. K. and I. Herskowitz (2000). "The ins and outs of cell-polarity decisions." Nat Cell Biol **2**(3): E39-41.
Oshlack, A., A. E. Chabot, G. K. Smyth and Y. Gilad (2007). "Using DNA microarrays to study gene expression in closely related species." Bioinformatics **23**(10): 1235-42.
Outeiro, T. F. and S. Lindquist (2003). "Yeast cells provide insight into alpha-synuclein biology and pathobiology." Science **302**(5651): 1772-5.
Pasquinelli, A. E., B. J. Reinhart, F. Slack*, et al.* (2000). "Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA." Nature **408**(6808): 86-9.
Payne, W. E. and J. I. Garrels (1997). "Yeast Protein database (YPD): a database for the complete proteome of Saccharomyces cerevisiae." Nucleic Acids Res **25**(1): 57-62.
Penn, S. G., D. R. Rank, D. K. Hanzel and D. L. Barker (2000). "Mining the human genome using microarrays of open reading frames." Nat Genet **26**(3): 315-8.
Perocchi, F., Z. Xu, S. Clauder-Munster and L. M. Steinmetz (2007). "Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D." Nucleic Acids Res **35**(19): e128.
Peters, J. M. (2006). "The anaphase promoting complex/cyclosome: a machine designed to destroy." Nat Rev Mol Cell Biol **7**(9): 644-56.
Picard, F., S. Robin, M. Lavielle*, et al.* (2005). "A statistical approach for array CGH data analysis." BMC Bioinformatics **6**: 27.
Pickford, A. S. and C. Cogoni (2003). "RNA-mediated gene silencing." Cell Mol Life Sci **60**(5): 871-82.
Pillai, R. S., S. N. Bhattacharyya, C. G. Artus*, et al.* (2005). "Inhibition of translational initiation by Let-7 MicroRNA in human cells." Science **309**(5740): 1573-6.
Plasterk, R. H. (2002). "RNA silencing: the genome's immune system." Science **296**(5571): 1263-5.
Pramila, T., W. Wu, S. Miles*, et al.* (2006). "The Forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle." Genes Dev **20**(16): 2266-78.
Prasanth, K. V. and D. L. Spector (2007). "Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum." Genes Dev **21**(1): 11-42.
Pratt, J. M., J. Petty, I. Riba-Garcia*, et al.* (2002). "Dynamics of protein turnover, a missing dimension in proteomics." Mol Cell Proteomics **1**(8): 579-91.
Preall, J. B., Z. He, J. M. Gorra and E. J. Sontheimer (2006). "Short interfering RNA strand selection is independent of dsRNA processing polarity during RNAi in Drosophila." Curr Biol **16**(5): 530-5.
Prescott, E. M. and N. J. Proudfoot (2002). "Transcriptional collision between convergent genes in budding yeast." Proc Natl Acad Sci U S A **99**(13): 8796-801.
Price, C., K. Nasmyth and T. Schuster (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.
Ptitsyn, A. A., S. Zvonic, S. A. Conrad*, et al.* (2006). "Circadian clocks are resounding in peripheral tissues." PLoS Comput Biol **2**(3): e16.
Raamsdonk, L. M., B. Teusink, D. Broadhurst*, et al.* (2001). "A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations." Nat Biotechnol **19**(1): 45-50.

Reed, R. and E. Hurt (2002). "A conserved mRNA export machinery coupled to pre-mRNA splicing." Cell **108**(4): 523-31.
Reed, S. I. (2003). "Ratchets and clocks: the cell cycle, ubiquitylation and protein turnover." Nat Rev Mol Cell Biol **4**(11): 855-64.
Reenan, R. A. (2001). "The RNA world meets behavior: A-->I pre-mRNA editing in animals." Trends Genet **17**(2): 53-6.
Reichow, S. L., T. Hamma, A. R. Ferre-D'Amare and G. Varani (2007). "The structure and function of small nucleolar ribonucleoproteins." Nucleic Acids Res **35**(5): 1452-64.
Reinhart, B. J., F. J. Slack, M. Basson*, et al.* (2000). "The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans." Nature **403**(6772): 901-6.
Ren, B., F. Robert, J. J. Wyrick*, et al.* (2000). "Genome-wide location and function of DNA binding proteins." Science **290**(5500): 2306-9.
Rhoades, M. W., B. J. Reinhart, L. P. Lim*, et al.* (2002). "Prediction of plant microRNA targets." Cell **110**(4): 513-20.
Rinn, J. L., G. Euskirchen, P. Bertone*, et al.* (2003). "The transcriptional activity of human Chromosome 22." Genes Dev **17**(4): 529-40.
Robert, V. J., T. Sijen, J. van Wolfswinkel and R. H. Plasterk (2005). "Chromatin and RNAi factors protect the C. elegans germline against repetitive sequences." Genes Dev **19**(7): 782-7.
Romeo, T. (1998). "Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB." Mol Microbiol **29**(6): 1321-30.
ROSE, A. H. and J. HARRISON, Eds. (1987-1995). The Yeasts. London - Orlando - New York - San Diego - Austin - Boston - Sydney - Tokyo - Toronto, Academic Press.
Ross-Macdonald, P., P. S. Coelho, T. Roemer*, et al.* (1999). "Large-scale analysis of the yeast genome by transposon tagging and gene disruption." Nature **402**(6760): 413-8.
Royce, T. E., J. S. Rozowsky, P. Bertone*, et al.* (2005). "Issues in the analysis of oligonucleotide tiling microarrays for transcript mapping." Trends Genet **21**(8): 466-75.
Ruprecht, R. M., N. C. Goodman and S. Spiegelman (1973). "Conditions for the selective synthesis of DNA complementary to template RNA." Biochim Biophys Acta **294**(1): 192-203.
Ruvkun, G. (2001). "Molecular biology. Glimpses of a tiny RNA world." Science **294**(5543): 797-9.
Ruvkun, G., B. Wightman and I. Ha (2004). "The 20 years it took to recognize the importance of tiny RNAs." Cell **116**(2 Suppl): S93-6, 2 p following S96.
Saha, S., A. B. Sparks, C. Rago*, et al.* (2002). "Using the transcriptome to annotate the genome." Nat Biotechnol **20**(5): 508-12.
Samanta, M. P., W. Tongprasit, H. Sethi*, et al.* (2006). "Global identification of noncoding RNAs in Saccharomyces cerevisiae by modulating an essential RNA processing pathway." Proc Natl Acad Sci U S A **103**(11): 4192-7.
SantaLucia, J., Jr. (1998). "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics." Proc Natl Acad Sci U S A **95**(4): 1460-5.
Scadden, A. D. and C. W. Smith (2001a). "RNAi is antagonized by A-->I hyper-editing." EMBO Rep **2**(12): 1107-11.
Scadden, A. D. and C. W. Smith (2001b). "Specific cleavage of hyper-edited dsRNAs." Embo J **20**(15): 4243-52.
Schadt, E. E., S. W. Edwards, D. GuhaThakurta*, et al.* (2004). "A comprehensive transcript index of the human genome generated using microarrays and computational approaches." Genome Biol **5**(10): R73.
Schattner, P., W. A. Decatur, C. A. Davis*, et al.* (2004). "Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the Saccharomyces cerevisiae genome." Nucleic Acids Res **32**(14): 4281-96.

Schena, M., D. Shalon, R. W. Davis and P. O. Brown (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray." Science **270**(5235): 467-70.
Schoenfelder, S., G. Smits, P. Fraser*, et al.* (2007). "Non-coding transcripts in the H19 imprinting control region mediate gene silencing in transgenic Drosophila." EMBO Rep **8**(11): 1068-73.
Schumacher, M. A., R. F. Pearson, T. Moller*, et al.* (2002). "Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein." Embo J **21**(13): 3546-56.
Schwob, E., T. Bohm, M. D. Mendenhall and K. Nasmyth (1994). "The B-type cyclin kinase inhibitor p40SIC1 controls the G1 to S transition in S. cerevisiae." Cell **79**(2): 233-44.
Scully, R., J. Chen, R. L. Ochs*, et al.* (1997). "Dynamic changes of BRCA1 subnuclear location and phosphorylation state are initiated by DNA damage." Cell **90**(3): 425-35.
Seitz, H., H. Royo, M. L. Bortolin*, et al.* (2004). "A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain." Genome Res **14**(9): 1741-8.
Selinger, D. W., K. J. Cheung, R. Mei*, et al.* (2000). "RNA expression analysis using a 30 base pair resolution Escherichia coli genome array." Nat Biotechnol **18**(12): 1262-8.
Sen, G. L. and H. M. Blau (2005). "Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies." Nat Cell Biol **7**(6): 633-6.
Shedden, K. and S. Cooper (2002). "Analysis of cell-cycle gene expression in Saccharomyces cerevisiae using microarrays and multiple synchronization methods." Nucleic Acids Res **30**(13): 2920-9.
Shen, B. and H. M. Goodman (2004). "Uridine addition after microRNA-directed cleavage." Science **306**(5698): 997.
Shen, W. C., S. R. Bhaumik, H. C. Causton*, et al.* (2003). "Systematic analysis of essential yeast TAFs in genome-wide transcription and preinitiation complex assembly." Embo J **22**(13): 3395-402.
Sherman, M. Y. and P. J. Muchowski (2003). "Making yeast tremble: yeast models as tools to study neurodegenerative disorders." Neuromolecular Med **4**(1-2): 133-46.
Shirayama, M., A. Toth, M. Galova and K. Nasmyth (1999). "APC(Cdc20) promotes exit from mitosis by destroying the anaphase inhibitor Pds1 and cyclin Clb5." Nature **402**(6758): 203-7.
Shoemaker, D. D., E. E. Schadt, C. D. Armour*, et al.* (2001). "Experimental annotation of the human genome using microarray technology." Nature **409**(6822): 922-7.
Sia, R. A., H. A. Herald and D. J. Lew (1996). "Cdc28 tyrosine phosphorylation and the morphogenesis checkpoint in budding yeast." Mol Biol Cell **7**(11): 1657-66.
Siepel, A., G. Bejerano, J. S. Pedersen*, et al.* (2005). "Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes." Genome Res **15**(8): 1034-50.
Sijen, T., J. Fleenor, F. Simmer*, et al.* (2001a). "On the role of RNA amplification in dsRNA-triggered gene silencing." Cell **107**(4): 465-76.
Sijen, T., I. Vijn, A. Rebocho*, et al.* (2001b). "Transcriptional and posttranscriptional gene silencing are mechanistically related." Curr Biol **11**(6): 436-40.
Simon, I., J. Barnett, N. Hannett*, et al.* (2001). "Serial regulation of transcriptional regulators in the yeast cell cycle." Cell **106**(6): 697-708.
Simon, J. A. and T. J. Yen (2003). "Novel approaches to screen for anticancer drugs using Saccharomyces cerevisiae." Methods Mol Biol **223**: 555-76.
Singer, B. S., T. Shtatland, D. Brown and L. Gold (1997). "Libraries for genomic SELEX." Nucleic Acids Res **25**(4): 781-6.
Sleutels, F., G. Tjon, T. Ludwig and D. P. Barlow (2003). "Imprinted silencing of Slc22a2 and Slc22a3 does not need transcriptional overlap between Igf2r and Air." Embo J **22**(14): 3696-704.
Sleutels, F., R. Zwart and D. P. Barlow (2002). "The non-coding Air RNA is required for silencing autosomal imprinted genes." Nature **415**(6873): 810-3.
Song, J. J., S. K. Smith, G. J. Hannon and L. Joshua-Tor (2004). "Crystal structure of Argonaute and its implications for RISC slicer activity." Science **305**(5689): 1434-7.

Sontheimer, E. J. (2005). "Assembly and function of RNA silencing complexes." Nat Rev Mol Cell Biol **6**(2): 127-38.
Sontheimer, E. J. and R. W. Carthew (2005). "Silence from within: endogenous siRNAs and miRNAs." Cell **122**(1): 9-12.
Sparago, A., F. Cerrato, M. Vernucci*, et al.* (2004). "Microdeletions in the human H19 DMR result in loss of IGF2 imprinting and Beckwith-Wiedemann syndrome." Nat Genet **36**(9): 958-60.
Sparks, K. A. and C. L. Dieckmann (1998). "Regulation of poly(A) site choice of several yeast mRNAs." Nucleic Acids Res **26**(20): 4676-87.
Spellman, P. T., G. Sherlock, M. Q. Zhang*, et al.* (1998). "Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization." Mol Biol Cell **9**(12): 3273-97.
Steinmetz, L. M. and R. W. Davis (2004). "Maximizing the potential of functional genomics." Nat Rev Genet **5**(3): 190-201.
Stolc, V., Z. Gauhar, C. Mason*, et al.* (2004). "A gene expression map for the euchromatic genome of Drosophila melanogaster." Science **306**(5696): 655-60.
Storz, G., S. Altuvia and K. M. Wassarman (2005). "An abundance of RNA regulators." Annu Rev Biochem **74**: 199-217.
Sun, M., L. D. Hurst, G. G. Carmichael and J. Chen (2005). "Evidence for a preferential targeting of 3'-UTRs by cis-encoded natural antisense transcripts." Nucleic Acids Res **33**(17): 5533-43.
Svoboda, P., P. Stein, H. Hayashi and R. M. Schultz (2000). "Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference." Development **127**(19): 4147-56.
Tabara, H., M. Sarkissian, W. G. Kelly*, et al.* (1999). "The rde-1 gene, RNA interference, and transposon silencing in C. elegans." Cell **99**(2): 123-32.
Taft, R. J., M. Pheasant and J. S. Mattick (2007). "The relationship between non-protein-coding DNA and eukaryotic complexity." Bioessays **29**(3): 288-99.
Tang, T. H., J. P. Bachellerie, T. Rozhdestvensky*, et al.* (2002). "Identification of 86 candidates for small non-messenger RNAs from the archaeon Archaeoglobus fulgidus." Proc Natl Acad Sci U S A **99**(11): 7536-41.
Team, R. D. C. (2007). The R Manuals. R. D. C. Team.
Thakur, N., V. K. Tiwari, H. Thomassin*, et al.* (2004). "An antisense RNA regulates the bidirectional silencing property of the Kcnq1 imprinting control region." Mol Cell Biol **24**(18): 7855-62.
Thomas, B. J., K. H. Zavitz, X. Dong*, et al.* (1997). "roughex down-regulates G2 cyclins in G1." Genes Dev **11**(10): 1289-98.
Tiedge, H., R. T. Fremeau, Jr., P. H. Weinstock*, et al.* (1991). "Dendritic location of neural BC1 RNA." Proc Natl Acad Sci U S A **88**(6): 2093-7.
Tijsterman, M., R. F. Ketting, K. L. Okihara*, et al.* (2002). "RNA helicase MUT-14-dependent gene silencing triggered in C. elegans by short antisense RNAs." Science **295**(5555): 694-7.
Tjaden, B., R. M. Saxena, S. Stolyar*, et al.* (2002). "Transcriptome analysis of Escherichia coli using high-density oligonucleotide probe arrays." Nucleic Acids Res **30**(17): 3732-8.
Tkacz, J. S. and V. L. MacKay (1979). "Sexual conjugation in yeast. Cell surface changes in response to the action of mating hormones." J Cell Biol **80**(2): 326-33.
Tommasi, S. and G. P. Pfeifer (1999). "In vivo structure of two divergent promoters at the human PCNA locus. Synthesis of antisense RNA and S phase-dependent binding of E2F complexes in intron 1." J Biol Chem **274**(39): 27829-38.
Tonkin, L. A., L. Saccomanno, D. P. Morse*, et al.* (2002). "RNA editing by ADARs is important for normal behavior in Caenorhabditis elegans." Embo J **21**(22): 6025-35.
Toth, A., E. Queralt, F. Uhlmann and B. Novak (2007). "Mitotic exit in two dimensions." J Theor Biol **248**(3): 560-73.

Trotochaud, A. E. and K. M. Wassarman (2004). "6S RNA function enhances long-term cell survival." J Bacteriol **186**(15): 4978-85.
Tsai, H. K., C. P. Su, M. Y. Lu*, et al.* (2007). "Co-expression of adjacent genes in yeast cannot be simply attributed to shared regulatory system." BMC Genomics **8**: 352.
Tufarelli, C., J. A. Stanley, D. Garrick*, et al.* (2003). "Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease." Nat Genet **34**(2): 157-65.
Tuschl, T., J. B. Thomson and F. Eckstein (1995). "RNA cleavage by small catalytic RNAs." Curr Opin Struct Biol **5**(3): 296-302.
Tycowski, K. T., C. M. Smith, M. D. Shu and J. A. Steitz (1996). "A small nucleolar RNA requirement for site-specific ribose methylation of rRNA in Xenopus." Proc Natl Acad Sci U S A **93**(25): 14480-5.
Tyers, M. (2004). "Cell cycle goes global." Curr Opin Cell Biol **16**(6): 602-13.
Tyers, M. and B. Futcher (1993). "Far1 and Fus3 link the mating pheromone signal transduction pathway to three G1-phase Cdc28 kinase complexes." Mol Cell Biol **13**(9): 5659-69.
Tyers, M. and P. Jorgensen (2000). "Proteolysis and the cell cycle: with this RING I do thee destroy." Curr Opin Genet Dev **10**(1): 54-64.
Tyson, J. J. and B. Novak (2001). "Regulation of the eukaryotic cell cycle: molecular antagonism, hysteresis, and irreversible transitions." J Theor Biol **210**(2): 249-63.
Tyson, J. J., B. Novak, K. Chen and J. Val (1995). "Checkpoints in the cell cycle from a modeler's perspective." Prog Cell Cycle Res **1**: 1-8.
Ueda, H. R., W. Chen, A. Adachi*, et al.* (2002). "A transcription factor response element for gene expression during circadian night." Nature **418**(6897): 534-9.
Uetz, P. and R. E. Hughes (2000). "Systematic and large-scale two-hybrid screens." Curr Opin Microbiol **3**(3): 303-8.
Uhler, J. P., C. Hertel and J. Q. Svejstrup (2007). "A role for noncoding transcription in activation of the yeast PHO5 gene." Proc Natl Acad Sci U S A **104**(19): 8011-6.
Valadkhan, S. (2007a). "The spliceosome: a ribozyme at heart?" Biol Chem **388**(7): 693-7.
Valadkhan, S. (2007b). "The spliceosome: caught in a web of shifting interactions." Curr Opin Struct Biol **17**(3): 310-5.
Valente, L. and K. Nishikura (2005). "ADAR gene family and A-to-I RNA editing: diverse roles in posttranscriptional gene regulation." Prog Nucleic Acid Res Mol Biol **79**: 299-338.
Valente, L. and K. Nishikura (2007). "RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions." J Biol Chem **282**(22): 16054-61.
Valverde, C., S. Heeb, C. Keel and D. Haas (2003). "RsmY, a small regulatory RNA, is required in concert with RsmZ for GacA-dependent expression of biocontrol traits in Pseudomonas fluorescens CHA0." Mol Microbiol **50**(4): 1361-79.
Vance, V. and H. Vaucheret (2001). "RNA silencing in plants--defense and counterdefense." Science **292**(5525): 2277-80.
Vanderpool, C. K. and S. Gottesman (2004). "Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system." Mol Microbiol **54**(4): 1076-89.
Vanhee-Brossollet, C. and C. Vaquero (1998). "Do natural antisense transcripts make sense in eukaryotes?" Gene **211**(1): 1-9.
Velculescu, V. E., L. Zhang, W. Zhou*, et al.* (1997). "Characterization of the yeast transcriptome." Cell **88**(2): 243-51.

Verlhac, M. H., J. Z. Kubiak, M. Weber*, et al.* (1996). "Mos is required for MAP kinase activation and is involved in microtubule organization during meiotic maturation in the mouse." Development **122**(3): 815-22.
Vogel, J., V. Bartels, T. H. Tang*, et al.* (2003). "RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria." Nucleic Acids Res **31**(22): 6435-43.
Volpe, T. A., C. Kidner, I. M. Hall*, et al.* (2002). "Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi." Science **297**(5588): 1833-7.
Wagner, E. G., S. Altuvia and P. Romby (2002). "Antisense RNAs in bacteria and their genetic elements." Adv Genet **46**: 361-98.
Wagner, E. G. and R. W. Simons (1994). "Antisense RNA control in bacteria, phages, and plasmids." Annu Rev Microbiol **48**: 713-42.
Wang, J. P., B. G. Lindsay, L. Cui*, et al.* (2005a). "Gene capture prediction and overlap estimation in EST sequencing from one or multiple libraries." BMC Bioinformatics **6**: 300.
Wang, Q., Z. Zhang, K. Blackwell and G. G. Carmichael (2005b). "Vigilins bind to promiscuously A-to-I-edited RNAs and are involved in the formation of heterochromatin." Curr Biol **15**(4): 384-91.
Wang, Y., D. R. Crawford and K. J. Davies (1996). "adapt33, a novel oxidant-inducible RNA from hamster HA-1 cells." Arch Biochem Biophys **332**(2): 255-60.
Wang, Y., K. J. Davies, J. A. Melendez and D. R. Crawford (2003). "Characterization of adapt33, a stress-inducible riboregulator." Gene Expr **11**(2): 85-94.
Washburn, M. P., A. Koller, G. Oshiro*, et al.* (2003). "Protein pathway and complex clustering of correlated mRNA and protein expression analyses in Saccharomyces cerevisiae." Proc Natl Acad Sci U S A **100**(6): 3107-12.
Wassarman, K. M. and G. Storz (2000). "6S RNA regulates E. coli RNA polymerase activity." Cell **101**(6): 613-23.
Wassarman, K. M., A. Zhang and G. Storz (1999). "Small RNAs in Escherichia coli." Trends Microbiol **7**(1): 37-45.
Wassenegger, M. (2000). "RNA-directed DNA methylation." Plant Mol Biol **43**(2-3): 203-20.
Waterhouse, P. M., M. B. Wang and T. Lough (2001). "Gene silencing as an adaptive defence against viruses." Nature **411**(6839): 834-42.
Weilbacher, T., K. Suzuki, A. K. Dubey*, et al.* (2003). "A novel sRNA component of the carbon storage regulatory system of Escherichia coli." Mol Microbiol **48**(3): 657-70.
Weinert, T. (1997). "A DNA damage checkpoint meets the cell cycle engine." Science **277**(5331): 1450-1.
Weinert, T. and L. Hartwell (1989). "Control of G2 delay by the rad9 gene of Saccharomyces cerevisiae." J Cell Sci Suppl **12**: 145-8.
Werner, A. and A. Berdal (2005). "Natural antisense transcripts: sound or silence?" Physiol Genomics **23**(2): 125-31.
Wesley, S. V., C. A. Helliwell, N. A. Smith*, et al.* (2001). "Construct design for efficient, effective and high-throughput gene silencing in plants." Plant J **27**(6): 581-90.
Whitehouse, I., O. J. Rando, J. Delrow and T. Tsukiyama (2007). "Chromatin remodelling at promoters suppresses antisense transcription." Nature **450**(7172): 1031-5.
Wichert, S., K. Fokianos and K. Strimmer (2004). "Identifying periodically expressed transcripts in microarray time series data." Bioinformatics **20**(1): 5-20.
Wilkie, G. S., K. S. Dickson and N. K. Gray (2003). "Regulation of mRNA translation by 5'- and 3'-UTR-binding factors." Trends Biochem Sci **28**(4): 182-8.
Williams, B. R. (1999). "PKR; a sentinel kinase for cellular stress." Oncogene **18**(45): 6112-20.
Williams, T. and M. Fried (1986). "A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3' ends." Nature **322**(6076): 275-9.

Willkomm, D. K., J. Minnerup, A. Huttenhofer and R. K. Hartmann (2005). "Experimental RNomics in Aquifex aeolicus: identification of small non-coding RNAs and the putative 6S RNA homolog." Nucleic Acids Res **33**(6): 1949-60.

Winzeler, E. A., B. Lee, J. H. McCusker and R. W. Davis (1999a). "Whole genome genetic-typing in yeast using high-density oligonucleotide arrays." Parasitology **118 Suppl**: S73-80.

Winzeler, E. A., D. R. Richards, A. R. Conway*, et al.* (1998). "Direct allelic variation scanning of the yeast genome." Science **281**(5380): 1194-7.

Winzeler, E. A., D. D. Shoemaker, A. Astromoff*, et al.* (1999b). "Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis." Science **285**(5429): 901-6.

Withey, J. H. and D. I. Friedman (2003). "A salvage pathway for protein structures: tmRNA and trans-translation." Annu Rev Microbiol **57**: 101-23.

Wittenberg, C., K. Sugimoto and S. I. Reed (1990). "G1-specific cyclins of S. cerevisiae: cell cycle periodicity, regulation by mating pheromone, and association with the p34CDC28 protein kinase." Cell **62**(2): 225-37.

Wodicka, L., H. Dong, M. Mittmann*, et al.* (1997). "Genome-wide expression monitoring in Saccharomyces cerevisiae." Nat Biotechnol **15**(13): 1359-67.

Workman, C., L. J. Jensen, H. Jarmer*, et al.* (2002). "A new non-linear normalization method for reducing variability in DNA microarray experiments." Genome Biol **3**(9): research0048.

Wu, L. F., T. R. Hughes, A. P. Davierwala*, et al.* (2002). "Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters." Nat Genet **31**(3): 255-65.

Wu, Z. and R. A. Irizarry (2005). "Stochastic models inspired by hybridization theory for short oligonucleotide arrays." J Comput Biol **12**(6): 882-93.

Wyers, F., M. Rougemaille, G. Badis*, et al.* (2005). "Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase." Cell **121**(5): 725-37.

Xie, Z., L. K. Johansen, A. M. Gustafson*, et al.* (2004). "Genetic and functional diversification of small RNA pathways in plants." PLoS Biol **2**(5): E104.

Xie, Z., K. D. Kasschau and J. C. Carrington (2003). "Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation." Curr Biol **13**(9): 784-9.

Yamada, K., J. Lim, J. M. Dale*, et al.* (2003). "Empirical analysis of transcriptional activity in the Arabidopsis genome." Science **302**(5646): 842-6.

Yamaguchi, M., K. Yoshida and N. Yanagishima (1994). "Sexual agglutination substances require a 'carrier' glycoprotein for integration into the cell wall of Saccharomyces cerevisiae." Microbiology **140 (Pt 9)**: 2217-23.

Yang, Z., Q. Zhu, K. Luo and Q. Zhou (2001). "The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription." Nature **414**(6861): 317-22.

Yelin, R., D. Dahary, R. Sorek*, et al.* (2003). "Widespread occurrence of antisense transcription in the human genome." Nat Biotechnol **21**(4): 379-86.

Yeong, F. M., H. H. Lim, Y. Wang and U. Surana (2001). "Early expressed Clb proteins allow accumulation of mitotic cyclin by inactivating proteolytic machinery during S phase." Mol Cell Biol **21**(15): 5071-81.

Young, R. A. (2000). "Biomedical discovery with DNA arrays." Cell **102**(1): 9-15.

Yuan, G., C. Klambt, J. P. Bachellerie*, et al.* (2003). "RNomics in Drosophila melanogaster: identification of 66 candidates for novel non-messenger RNAs." Nucleic Acids Res **31**(10): 2495-507.

Yusupov, M. M., G. Z. Yusupova, A. Baucom*, et al.* (2001). "Crystal structure of the ribosome at 5.5 A resolution." Science **292**(5518): 883-96.

Zalfa, F., M. Giorgi, B. Primerano*, et al.* (2003). "The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses." Cell **112**(3): 317-27.

Zamore, P. D. (2002). "Ancient pathways programmed by small RNAs." Science **296**(5571): 1265-9.
Zamore, P. D. and B. Haley (2005). "Ribo-gnome: the big world of small RNAs." Science **309**(5740): 1519-24.
Zamore, P. D., T. Tuschl, P. A. Sharp and D. P. Bartel (2000). "RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals." Cell **101**(1): 25-33.
Zanolari, B. and H. Riezman (1991). "Quantitation of alpha-factor internalization and response during the Saccharomyces cerevisiae cell cycle." Mol Cell Biol **11**(10): 5251-8.
Zeng, Y. (2006). "Principles of micro-RNA production and maturation." Oncogene **25**(46): 6156-62.
Zhang, A., K. M. Wassarman, J. Ortega*, et al.* (2002). "The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs." Mol Cell **9**(1): 11-22.
Zhang, A., K. M. Wassarman, C. Rosenow*, et al.* (2003). "Global analysis of small RNA and mRNA targets of Hfq." Mol Microbiol **50**(4): 1111-24.
Zhang, Z. and G. G. Carmichael (2001). "The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs." Cell **106**(4): 465-75.
Zhao, L. P., R. Prentice and L. Breeden (2001). "Statistical modeling of large microarray data sets to identify stimulus-response profiles." Proc Natl Acad Sci U S A **98**(10): 5631-6.
Zhao, Y., J. F. Ransom, A. Li*, et al.* (2007). "Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2." Cell **129**(2): 303-17.
Zhu, H., M. Bilgin and M. Snyder (2003). "Proteomics." Annu Rev Biochem **72**: 783-812.
Zhu, Z., Y. Pilpel and G. M. Church (2002). "Computational identification of transcription factor binding sites via a transcription-factor-centric clustering (TFCC) algorithm." J Mol Biol **318**(1): 71-81.
Zilberman, D., X. Cao and S. E. Jacobsen (2003). "ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation." Science **299**(5607): 716-9.

7. APPENDIX

Table A1.

Table A2.