UNIVERSITÀ DEGLI STUDI DI MILANO – BICOCCA
Facoltà di Scienze Matematiche, Fisiche e Naturali
Dipartimento di Biotecnologie e Bioscienze

Comparative systems-level analysis of the G1/S transition in yeast and higher eukaryotes: focusing on the Whi5/Rb network and initiation of DNA replication

Md Mehedi Hasan
Tutor: Prof. Marco Vanoni

Anno Accademico 2011-2012
Dottorato in Biotecnologie Industriali, XXV ciclo

Comparative systems-level analysis of the G1/S transition in yeast and higher eukaryotes: focusing on the Whi5/Rb network and initiation of DNA replication

By
Md Mehedi Hasan

A dissertation submitted to
Department of Biotechnology and Biosciences
of
The University of Milan-Bicocca
for the degree of
Doctor of Philosophy

Md Mehedi Hasan
Matricola: 734705
January 2013

Abstract

Molecular systems biology, holds that biological processes are the result of complex, coordinated, dynamical, non-linear interactions that generate the corresponding function as an emergent property of the system, that therefore is not found in individual components, but only in their networking. Conversely, it has been shown that even a single, multi-domain protein may present a system-level behavior that can be described by adapting the formalism used to describe inter-molecular networks. Accordingly, a multi-scale approach is required to fully understand any given biological function.

The G1/S transition and the initiation of DNA replication are major regulatory events in the eukaryotic cell cycle. In this thesis I have provided a comparative survey of the molecular events leading to the initiation of DNA replication and analyzed some structural features and the interactome profile of *Saccharomyces cerevisiae* Whi5, a regulator of G1/S-specific transcription in budding yeast in comparison with those with those of Rb – and members of the pocket family - that perform equivalent function(s) in mammalian cells.

In eukaryotes DNA replication initiates from hundreds to thousands of replication origins in a coordinated manner, in order to efficiently duplicate the genome. The sequence of events leading to the onset of DNA replication is conventionally divided in two interdependent processes: licensing—a process during which replication origins acquire replication competence but are kept inactive— and firing—a process during which licensed origins are activated but not re-licensed. In this thesis we investigate the evolutionary conservation of the molecular machinery orchestrating DNA replication initiation both in yeast and in mammalian cells, highlighting a remarkable conservation of the general architecture of this central biological mechanism. Many steps are

conserved down to molecular details and are performed by orthologous proteins with high sequence conservation, while differences in molecular structure of the performing proteins and their interactions are apparent in other steps. Tight regulation of initiation of DNA replication is achieved through protein phosphorylation, exerted mostly by Cyclin-dependent kinases in order to ensure that each chromosome is fully replicated once, and only once, during each cycle, and to avoid the formation of aberrant DNA structures and incorrect chromosomal duplication, that in mammalian cells are a prerequisite for genome instability and tumorigenesis.

Among many regulatory proteins, Retinoblastoma (Rb) of higher eukaryotes and Whi5 of yeasts play a key role in the regulation of gene expression of proteins involved in DNA replication including Cyclin-dependent kinases and its partners. Rb is part of a protein family encompassing three members. Rb plays many cellular roles, mediated by several downstream effectors and transcriptional targets. The best known role of Rb is the control of cell cycle progression at the G1/S transition, mediated through its interaction with E2F transcription factors required for entrance into S phase. Rb is also involved in regulation of cellular differentiation during embryogenesis and in adult tissue, regulation of p53-dependent apoptosis and DNA repair, maintenance of permanent cell cycle arrest, as quiescence in stem cell promotion of cell cycle exit through inhibition of the E2F transcription factors and the transcriptional repression of genes encoding cell-cycle regulators.

While proteins related to the retinoblastoma tumor suppressor Rb and the E2F transcription factor are conserved in most eukaryotic lineages, in yeast no hortologues of Rb or E2F have been found. Both Rb and Whi5 are disordered proteins. In Rb, that is a much larger protein than Whi5, structured domains alternate with disordered regions, whereas in Whi5

only a single structured domain is likely to exist. The family of Whi5-related proteins is only present in *Saccharomycetales*. The higher-than-expected conservation of sequence in disordered regions, correlates with abundance of phosphorylation sites and allows to predict conservation of a similar phosphorylation rhythm with a potentially functionally relevant role. Notably, conserved motifs 1 and 3 may act as phosphorylation-dependent seeds in Whi5 folding/unfolding.

The (partial) disordered nature of both proteins allows them to act as protein hubs, able to interact with many partners. The interaction landscape of Rb and Whi5 is quite large, with more than one hundred proteins interacting either genetically or physically with either protein. Interestingly, the smaller Whi5 acts as a hierarchical hub. Comparison of Rb and Whi5 interactors (both physical and genetical ones) allows to highlight a significant core of conserved common functionalities associated with the interactors indicating that network structure and function – rather than individual proteins, are conserved during evolution

By step-wise adding interactors to existing models for Whi5 function, an improved "concept map" of Whi5 function is being constructed. Ultimately, such a map will allow to construct dynamic mathematical model(s) of increasing granularity and to design experiments to proof novel regulatory links within the Whi5 network.

Acknowledgements

The author expresses his gratitude to almighty Creator, the most gracious and the most merciful, who made this universe and arranges all things as necessary and directs all of them for the betterment of human beings.

The author is most grateful to Professor Marco Vanoni, Department of Biotechnology and Biosciences, University of Milan-Bicocca for his guidance, constant encouragement, close supervision, monitoring and supreme ability for solving all type of problems arising during the research work.

I am indebted to Professor Lilia Alberghina and Professor Marco Vanoni for the financial support provided by Department of Biotechnology and Biosciences during the whole time of my research and the writing-up period, without which I could not have completed my study at Milan. I would also like thank them for their interest in my work.

The author expresses his sincere gratitude to Professor Marco Vanoni for his supervision and close monitoring during the course of study and research and in preparation of the manuscript. I would like to thank him for his knowledge, direction and perceptiveness. His noteworthy comments and constructive criticism on my work have been invaluable.

This thesis would not have been possible without the help and support of many people.

It is also great pleasure for the author to express his gratitude and regard to her supervisor Elena Sacco for her encouragement, suggestions and supervision during the course and research work.

The author also expresses his heartfelt gratitude and regards-----to all lab mates in the Biotechnology Laboratory of the Department of Biotechnology and Biosciences, University of Milan-Bicocca especially to, Stefania Broccca, Stefano Busti and Laura Gotti. Special thanks to Stefania for guiding Whi5-Rb related work.

The author expresses his immense gratitude to his wife, son, brothers, sisters and relatives for their blessings, encouragement and constant inspiration in the pursuit of his higher study. Finally, the deepest gratitude is due to his father and mother for their sacrifices and inspiration.

Author

January 2013

Declaration

This thesis has not been previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree. All the work reported in this thesis has been performed by myself and my colleagues, except where otherwise stated. Other sources are acknowledged explicitly in the text. A bibliography is appended.

Author

January 2013

Contents

List of Figures

List of Tables

CHAPTER

1

Introduction

1.1 Scope of the Thesis

Cell cycle regulation is major process in all living eukaryotes. From yeast to human same core cell cycle control points are observed though higher eukaryotes equipped with more diverse control points as they need diversity not only in cell types but also in protein level to cope their multicellular environment. To understand complex biological systems, comparative analysis of between lower eukaryotes like yeast and higher eukaryotes like human is better way to evaluate complexity in higher eukaryotes on the basis of lower eukaryotes. As key regulatory proteins are conserved across very diverse eukaryotes, it is interesting to investigate whether key regulatory pathways are conserved in what extent. The emerging field of Systems Biology provides a powerful foundation and established scientific methods to enable the understanding of biological regulatory pathways at the system level. By applying the holistic approach of systems biology, it is worthwhile to compare pathway/regulatory model among very distantly related species.

In this thesis, the challenge of integrating and comparing of complex biological systems is investigated, when the prior knowledge about the system is incomplete and the available experimental data is sparse at least in one comparing species. We have made an attempt to compare regulatory proteins and regulatory steps involved in the G1/S transition in yeast and higher eukaryotes and specially focused on the Whi5/Rb network and initiation of DNA replication. This comparative analysis takes advantage of most recent published high quality data supported by strong publication records and state of the art computational tools and methodologies. To represent a simple and understandable portrait of complex process involved in DNA replication initiation and Whi5/Rb network, cautious steps were always followed.

The following sections give an overall introduction to the relevant background of the research subject.

1.2 Cell Cycle

All living organisms are product of reiteration of cell growth and division. An elaborate series of events including growth and division are known as cell cycle events in which genetic materials are duplicated and then distributed into two daughter cells. In eukaryotes cell divides through cell cycle which is an orderly sequence of events of four phases such as *G1 phase*-first gap phase, *S phase*- Synthetic phase, *G2 phase*-second gap phase and *M phase*- Mitotic phase (Fig. 1.1). Alternatively, cell cycle can be divided briefly in three phases such as interphase (G1, S, and G2 phases), mitosis (M phase) and cytokinesis (C) (Fig. 1.1). Cell cycle is also equipped with a control system or checkpoints that ensure faithful completion of each of the phases (Morgan 2007) . The checkpoints are G1/S checkpoint, G2/M checkpoint and metaphase-anaphase (M checkpoint) (Fig. 1.1). The following paragraphs will give a brief description of cell cycle phases and their checkpoints.

G1 phase

After cell division, new daughter cells allow more time for growth, a gap phase known as a G1 phase between M phase and S phase is required to grow and to prepare for S phase. G1 is an important regulatory period for taking decision of cell. In this phase, most cells become committed to either continued division or exit from the cell cycle. In higher eukaryotes cells in G1 phase can be in a dividing proliferative state or they can enter non-dividing states which can be divided four possible states such as quiescence (G0), senescence, differentiation, and apoptosis. Most of the terminally differentiated cells and the senescent cells have permanently withdrawn the cell cycle though can survive for long periods of time.

Upon appropriate stimuli quiescent cells can reversibly re-enter the cell cycle. Prolonged non dividing state, sometimes called G_0 (G zero) can occur in the presence of unfavorable growth conditions or inhibitory signals from other cells (Morgan 2007).

S phase

In this the phase, cell synthesizes a replica of the genome which is also known as chromosome duplication more specifically DNA replication of the genome. Chromosome duplication not only duplicates DNA content but also duplicates chromatin structure. DNA-protein complex of chromosome is known as chromatin. Chromatin is extensively packaged with histone proteins and various regulatory proteins involved in the control of gene expression. Highly condensed chromatin is called heterochromatin, whereas in more open structure is called euchromatin. Euchromatin is more accessible to regulatory factors and actively participate in gene expression. The major unit of DNA packaging is the nucleosome which is the histone octamer---composed of eight histone proteins such as histones H2A. H2B, H3 and H4 (two copies of each protein). The making of these histone proteins increases during S phase due to Cdks stimulation to provide the raw materials needed to package the newly synthesized DNA (Groth *et al.* 2007, Onn *et al.* 2008).

DNA replication can be divided into three continues steps: initiation, elongation and termination. The whole replication process is tightly regulated that requires coordinated assembly various replicative proteins or factors. DNA replication occur at remarkable speed and accuracy. The regulation of DNA replication ensures to replicate every piece of DNA strictly once per cell cycle. DNA replication start from a site known as origin of replication recognized by origin replication complex proteins or ORC proteins. DNA replication initiation can be divided into sequential steps such as licensing and firing. Cdks mediated

assembly of various protein at origin of replication help to license and then to fire the origin. The basic steps are conserved from yeast to human. See chapter 2 to have detail description on DNA replication initiation event and comparative analysis of these process between yeast and mammalian cells counterpart (Diffley 2004, Masai *et al.* 2010).

After DNA replication initiation from thousands of replication origin, elongation phase starts. The elongation phase of replication includes two distinct but related process such as leading strand synthesis and lagging strand synthesis. Leading strand synthesis is straight forward, whereas lagging strand synthesis is accomplished in short Okazaki fragments. Various DNA polymerases and other replication factors have crucial role in performing DNA replication elongation and termination faithfully (Diffley 2004, Masai et al. 2010).

G2 phase

G2 phase is gap between S phase and mitosis. In this phase, cell prepares for entering M phase. It arranges for genomic separation. During this time, mitochondria and other organelles replicate, chromosomes condense, and microtubules begin to assemble at a spindle. Duration of this greatly vary depending on cell type. After the chromosome duplication, chromosomes remain fully extended and uncoiled. In this phase, chromosomes start slowly the long process of condensation. Early in mitosis the quick final condensation occurs due to involvement of some motor proteins. In this time, all eukaryotic cells start an extensive synthesis of tubulin which is required for microtubules formation. In higher eukaryote, centrioles- a pair of microtubule organizing centers are formed (Morgan 2007).

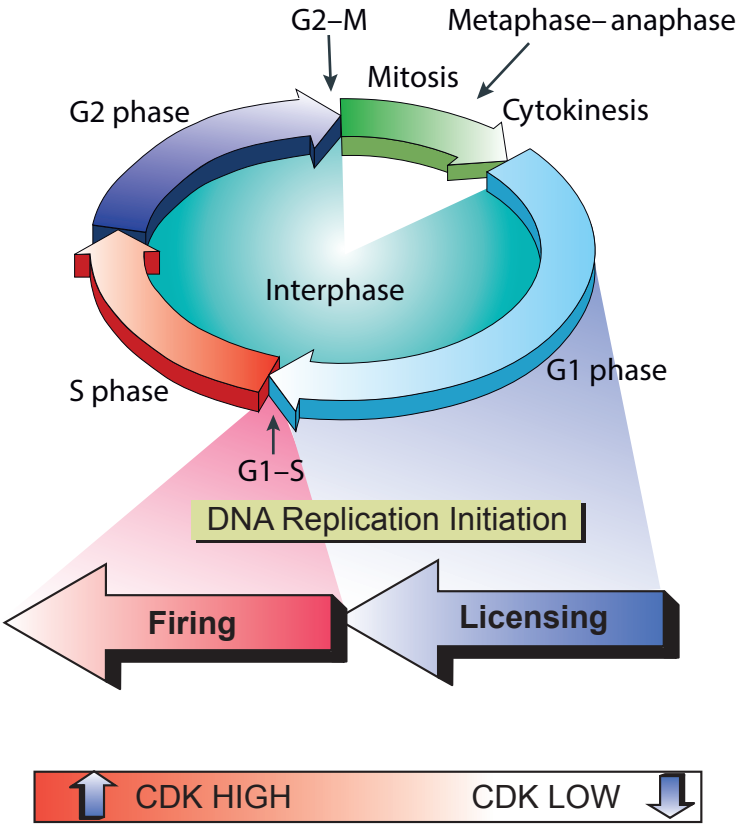


Fig. 1.1 The different phases and the control of the cell cycle. In most cells, G1 is the gap between M phase and S phase, while G2 is the gap between S phase and M phase. A cell-cycle control system triggers the essential processes of the cell cycle such as DNA replication, mitosis, and cytokinesis. The control system is represented here as G1/S checkpoint, G2/M checkpoint and metaphase-anaphase. The control system may arrest the cycle at these checkpoints depending on information about the completion of cell-cycle events and/or signals from the environment. The steps of DNA replication initiation such as licensing and firing occur at G1/S transition and activation of these steps relays on activation of Cdks.

M phase

After the completion of S phase and G2 phase, the cell undergoes the steps of M phase. The M phase begins with mitosis when the sister chromatids are separated and distributed and is ended up with cytokinesis. Traditionally mitosis can be divided into five stages- prophase, prometaphase, metaphase, anaphase, and telophase. Alternatively, mitosis can be divided into two main steps/checkpoints such as Cdk activity at the G2/M checkpoint which triggers the events of early mitosis and APC/C activity at metaphase-toanaphase transition which triggers the destruction of securin, liberating a protease that cleaves cohesin and thereby initiates separation of the sister chromatids.

Checkpoints

All eukaryotic cells use similar molecular machines and regulatory mechanisms to control the events of cell cycle. The cell cycle progresses through the regulatory transitions known as checkpoints. It has three major checkpoints such as G1/S checkpoint, G2/M checkpoint and metaphase-anaphase (M checkpoint) (fig. 1.1).

The first checkpoint is called Start (yeast) or restriction point (mammalian cells) or the G1/S checkpoint. When conditions are ideal for cell division, cyclin-Cdk complexes responsible for promoting G1/S and S-phase are activated. These kinases phosphorylate wide array of proteins that promote chromosome duplication events including DNA replication, centrosome duplication and others.

Cyclin-Cdk complexes responsible for promoting M phase, which drive cell cycle progression through the second major checkpoint at the entry into mitosis which is also known as G2/M checkpoint. The G2/M checkpoints essentially represent multiple checkpoints such as regulating cellular entry into mitosis, transiting through mitosis etc. This

checkpoint ensures the cells cycle arrest in late G2 if DNA is damaged during the late S phase. There are some known proteins and kinases play role in these checkpoint such as ATM, ATR, Chek1, Chek2, polo kinase, aurora kinase etc. In case of DNA damage some of these key proteins help to block of cell cycle progression from G2 into mitosis. G2/M checkpoint promotes spindle assembly and bringing the cell to metaphase.

The third notable checkpoint is the metaphase-to-anaphase transition also known as spindle assembly checkpoint, which leads to sister chromatid segregation, completion of mitosis and cytokinesis. This checkpoint ensues when M-phase specific cyclin-Cdk complex stimulate or inhibit the anaphase promoting complex (APC) that promotes the proteolytic destruction of cyclins and of proteins involved in cohesion of sister chromatids (Peters 2006). When APC is activated, it triggers sister-chromatid separation and segregation through destruction of key proteins and allows phosphatases to dephosphorylate Cdk substrates as Cdk is inactive due to destruction of cyclins. Dephosphorylation events promoted by APC is necessary for spindle disassembly and for cytokinesis.

1.3 Cell Cycle Control and Cyclin-Dependent Kinases

Cyclin-dependent kinases (Cdks), members of a family of protein kinases, are fundamental components of the cell-cycle control system. The most important of binding partner of these kinases as their name implies are proteins known as cyclins. In simple cell cycle such as budding yeast cell cycle Cyclins undergo a cycle of synthesis and degradation, by contrast budding yeast Cdk1 (also known as Cdc28) remain constant. On the other hand, in complex cell cycle like mammalian cell cycle both Cyclins and Cdks undergo synthesis and

degradation. Cdks are dependent on cyclins for their kinase activity. Due to their cyclical degradation and synthesis, different Cdks and Cyclins combination controls different phases and check points of cell cycle. By thus they can phosphorylate wide range of intracellular proteins that initiate or regulate the major events of the cell cycle (Malumbres and Barbacid 2009).

In general Cyclins can be divided into four classes such as G1 –cyclins, G1/S –cyclins, S –cyclins and M –cyclins. As their names imply, these cyclins are function in specific cell cycle. In budding yeast Cdk1 or Cdc28 bind all of these cyclins. But in higher eukaryotes for example mammalian cells, by contrast, there are at least 13 CDKs and 29 cyclins. Among all Cdks, Cdk1–Cdk6, Cdk10, Cdk11, and Cdk7 (Cdk-activating kinase), are all involved in cell cycle control. Cdk 4/6 is involved in G1 phase, Cdk2 is involved in G1/S transition and S phase and Cdk1 is involved in S and M phase. Other Cdks such as CDK8, and CDK9 are involved in transcriptional regulation through phosphorylating RNA polymerase and Cdk12 and Cdk13 regulate alternative RNA splicing. (Liu and Kipreos 2000, Loyer *et al.* 2005, Malumbres and Barbacid 2009, Phatnani and Greenleaf 2006). In yeast Cln3 is G1 –cyclins, Cln1,2 are G1/S –cyclins, Clb5,6 are S –cyclins and Clb 1,2,3,4 are M –cyclins. In higher eukaryotes, by contrast, Cyclin D (multiple isoform) is G1 –cyclins, Cyclin E (multiple isoform) are G1/S –cyclins, Cyclin A (multiple isoform) are S –cyclins and Cyclin B (multiple isoform) are M –cyclins (Pavletich 1999) .

Normally Cdks phosphorylate at serine or threonine residues of target substrate or proteins and the active site of these Cdks recognize a specific sequence combination that is present target proteins. Usually, the target sequence contains a serine (S) or threonine (T) residue is followed by a proline (P). The typical phosphorylation sequence motif for Cdks is [S/T]PX[K/R], where X represents any amino acid and K/R

represents lysine (K) or arginine (R) residues. There are many available computational tools that recognize this motif and predict possible phosphorylatable sequence/sites.

Activation and inhibition of Cyclin-Cdk complex are also regulated by cell cycle dependent manner. Cdk-activating kinases (CAKs) help to fully activate the cyclin-Cdk complex by phosphorylating an amino acid near the entrance of the Cdk active site. Cdk inhibitor proteins (CKIs) inhibit the function of Cdks by binding. In budding yeast Sic1 is such inhibitor which suppresses Cdk1 activity in G1. CIP and KIP proteins are such inhibitors in mammalian cells.

Cdk activity oscillates during the cell cycle and this oscillation helps to regulate transcription of hundreds of regulatory genes which synthesizes essential proteins and enzymes that regulate and control cell cycle. The main transcriptional control points are G1/S or Start or restriction point and G2/M. Cdks control timing of activation/inactivation of different set of regulatory proteins by phosphorylation to control these checkpoints (Morgan 1997).

In budding yeast, SCB-binding factor (SBF) and MCB-binding factor (MBF) are two gene regulatory complexes, control gene expression of hundreds of proteins that are essential for cell cycle progression (Wittenberg and Reed 2005). Gene expression of these genes are inhibited by some key proteins including Whi5. In G1/S transition, Cdks phosphorylate these key proteins and alleviate their repression. In higher eukaryotes, E2F family protein control gene expression of hundreds of proteins that are essential for cell cycle progression. Gene expression of these genes are inhibited by some key proteins including Rb and Cdks here also play same role like yeast.

At the G2/M checkpoint, Cdk activity increases the phosphorylation of proteins both in yeast and higher eukaryotes that control nuclear

envelope breakdown, spindle assembly, chromosome condensation and other events that occur at the onset of mitosis.

1.4 Eukaryotic DNA replication initiation and Rb and Whi5 proteins

The most primitive function of cell cycle is to duplicate the chromosome through DNA replication and then segregate them precisely. DNA replication event can be divided into three sequential steps: initiation, elongation and termination. Initiation event takes place from late G1 to early S phase when all replication proteins are assembled onto the origin of replication. The initiation event is composed of two sequential process, *licensing* when origin is licensed through recruitment of licensing proteins and *firing* when origin is fired through recruitment of firing proteins (Arias and Walter 2007, Bell and Dutta 2002, Masai et al. 2010). To maintain genomic integrity, DNA replication initiation event must be strictly regulated to prevent re-initiation of replication and re-firing of an origin in the same S phase. CDKs are key players to control and regulate the whole process of DNA replication specially the initiation event. The failure of this control involves the re-replication of DNA that induces genomic instability, which in turn correlates with higher eukaryotes cancer. Among the many proteins involved in regulation of upstream of DNA replication initiation event, Rb of higher eukaryotes and Whi5 in yeasts play a key role in the regulation of gene expression of proteins involved in DNA replication including CDK partners (see chapter 2 for detail description on DNA replication initiation).

Among many cell cycle regulator proteins Rb (human homolog) is one the best studied proteins since its discovery in 25 years ago as first tumor suppressor. Besides its central regulatory function of cell cycle

progression which was derived from its capability to arrest cells in G1, the other functions of Rb includes control of cellular differentiation, regulation of apoptosis dependent cell death, maintenance of senescence or permanent cell cycle arrest and protection of genomic and chromosomal stability (Chicas *et al.* 2010). The Rb counterpart in budding yeast as proposed is Whi5 derived from its deletion that not only influence cell size but also the subset of these mutants that affect the critical cell size threshold which is required for passage through START by accelerating the G1/S phase transition while overexpression of Whi5 causes G1 delay and an increase in cell size in wild-type cells (Jorgensen *et al.* 2002, Zhang *et al.* 2002). Beyond its regulatory function of cell cycle progression in budding yeast very little is known about other function of this protein (see chapter 3 for detail description on Whi5 and Rb proteins).

.

1.5 Systems Biology and Comparative Interactomics

According to (Kitano 2002) Systems Biology is to understand biology at the system level. It is an interdisciplinary field provides a vibrant interface among biologists, computer scientists, applied mathematicians and statisticians to support the development of a unified understanding of the biological phenomenon. Alternatively "it is to understand in quantitative, predictable ways the regulation of complex cellular pathways and of intercellular communication so as to shed light on complex biological functions (e.g. metabolism, cell signaling, cell cycle, apoptosis, differentiation, and transformation)" (adapted from Westerhoff and Alberghina 2005).

Now a days systems Biology is becoming very widespread as it is generally recognized that, in biology, dynamic behavior of the whole system may not be easily assumed from descriptions of individual parts,

and can only be achieved by a combined description of all known parts and applying systematic approach with assistance of advanced computing technology to predict unknown phenomenon. The exponential growth of biological knowledge offers the possibility to perform various computational analyses on one organism or across different organisms.

Interactomics is one of the fastest moving fields in molecular biology due to availability of high throughput methodologies to investigate systematically all the possible interactions in a variety of model organisms. This huge data provides an opportunity to compare interaction networks among various species and finding their conservations (Kiemer and Cesareni 2007). Comparing interaction network of proteins of different species can provide information about the protein network evolution, insight of molecular mechanisms leading to organism diversity, and shed light on unknown part of networks. At the intersection of comparative interactomics and systems biology lies great possibility for discovery, analysis and prediction (Lin and Qian 2007).

1.6 Thesis Statement

Comparative network analysis not only helps to understand complex networks of biological pathways but also it sheds light on unknown pathways. Coupled with experimental data, comparative network analysis aims to enable the construction and interpretation of complex systems in a sound and integrated environment. The thesis focused on is to analyze the network and regulation of DNA replication initiation in yeast and mammalian system. The basic mechanisms of DNA replication are conserved across eukaryotic species despite some differences between simple and complex eukaryotic organisms. Timing of both processes is strictly regulated by CDK activity. In contrast to lower

eukaryotes like yeast, complex eukaryotes like mammalians are more complex and regulated in more plastic manner, even though both systems contain similar conserved protein machinery for replication purpose. To understand deregulation and defects in the process of DNA replication initiation in mammalian cells that could lead to genome instability, carcinogenesis or chromosome instability syndrome which ultimately predispose in cancer, this comparative analysis sheds some light on complex and unknown steps of DNA replication initiation of mammalian cells based on known steps involved in yeast DNA replication initiation.

The another goal of this thesis is to explore similarities and differences between Rb and Whi5 specially their human and budding yeast counterpart proteins and covers not only their sequence and structure level architecture but also their involved interaction pathway and controlled target genes to visualize the system level similarity between these two proteins. Insight of these analyses eventually show the path to build model(s) of Whi5 and Rb regulation capturing essential features of the process will allow probing evolutionary conservativity of the function of Whi5 and Rb and will give testable predictions to be used to design experiments aimed at defining the main regulatory events of the circuit.

1.7 Thesis Contributions

A list of contributions of the thesis is as follows:

- The comprehensive analysis of DNA replication initiation in yeast and mammalian cells is presented in Chapter 2. An outline followed by step by step detail analysis of DNA replication initiation is discussed for both yeast and mammalian cells. The

presented model of DNA replication initiation in yeast and mammalian cells not only gives a global view of overall process but also pinpoints regulatory steps involved in overall process.

- A comparative view of the events involved in DNA replication initiation in yeast and mammalian cells is depicted in Chapter 2. Figures and tables represented in Chapter give detailed step by step comparison.
- Multisite protein phosphorylation is central regulation event that control step by step regulation of DNA replication initiation. Phosphorylation exerted by Cdks are discussed thoroughly in Chapter 2 and 3.
- Though some reviews were focused on functional similarities between these two proteins (Cooper 2006), a detail comparative study of Whi5 and Rb protein is not present in literature. Chapter 3 represents detail analysis of Whi5 and Rb at the level of sequence and evolutionary perspectives. Interaction network of both proteins are also analyzed which provides essential understanding of their function in detail.

1.8 Publications

Publications during the period of this research include

1. Sacco, E, Hasan, MM, Alberghina, L, and Vanoni, M (2012). Comparative analysis of the molecular mechanisms controlling the initiation of chromosomal DNA replication in yeast and in mammalian cells. Biotechnol Adv 30, 73-98.

2. Hasan MM, Sacco E, Papaleo E, Brocca S, de Gioia L, Alberghina L, Vanoni M (2012) Comparative analysis of Whi5 and Rb network: towards a system level understanding ECCB'12, Basel, Switzerland (poster)

CHAPTER

2

Comparative analysis of the molecular mechanisms controlling the initiation of chromosomal DNA replication in yeast and in mammalian cells

2.1 Introduction

In eukaryotic cells proliferation shows two major features: faithful chromosomal DNA replication and coordination of growth with the discontinuous events of DNA replication and of mitosis/cell division in order to maintain a fairly constant cell size. One of the more exciting areas of research aims to discover the molecular mechanisms that orchestrate this complex, integrated regulatory program.

As is common in biology, lower eukaryotes are often used as model systems, with the aim of progressing to mammalian cells, given that the derangement of regulation of cell proliferation is a hallmark of cancer, a human disease still in strong medical need.

DNA replication takes place in the S phase of the cell cycle, starting from a large number of origins of replication (OR), some of which fire early and some late in the S phase, moving in both directions on the DNA molecule (Araki 2010a).

The rate-limiting step of DNA replication in budding yeast is the initiation phase, during which, following a complex sequence of events, a number of OR are activated in a more or less synchronous way. While the coherence of the onset of DNA replication on hundreds of OR is mandatory for a faithful DNA synthesis in budding yeast (Remus and Diffley 2009), the stringency of synchrony is less evident in mammalian cells, as shown in Table 2.1 which reports several relevant features of the onset of chromosomal DNA replication in both type of cells.

The sequence of events leading to the onset of DNA replication is conventionally divided in two interdependent processes: *licensing* and *firing*. *Licensing* is a process during which replication origins acquire replication competence but are kept inactive, and *firing* is a process during which licensed origins are activated but not re-licensed (Arias and Walter 2007, Bell and Dutta 2002, Masai, *et al.* 2010). Licensing and firing are complex processes involving approximately 20 to 30 proteins:

with so many players on stage it is not surprising that there is not complete agreement in the literature on the detailed sequence of assembly of the various interactors (when comparing one species with the other or comparing different studies within the same organism), although there is a large consensus on the basic organization (see section 2.**2**).

Several regulatory steps control DNA replication, in order to ensure that each chromosome is fully replicated once, and only once, during each cycle. This tight regulation of the initiation of DNA replication is achieved through protein phosphorylation, exerted mostly by Cyclin-dependent kinases (Cdks), whose activity is dependent upon cellular growth. Cdks control onset of S phase with a complex regulatory mechanism so that a number of replication origins - sufficient to allow complete genome duplication in the time window of the S phase - are licensed during G1 phase, so to prevent loss of genetic material that could cause lethality in progeny. In addition, a second layer of control entails inhibition of licensing, in order to prevent re-licensing (and re-replication) of origins which have already fired, effectively hindering DNA amplification and chromosomal aberration. Both insufficient licensing during G1 phase and re-firing during S and G2/M phase can originate aberrant DNA structures and incorrect chromosomal duplication, that in mammalian cells are a prerequisite for genome instability and tumorigenesis (Lau*, et al.* 2007, Truong and Wu 2011). Furthermore, alternance between S phase and mitosis is required to preserve the maintenance of genome dosage. An uncoupling between DNA duplication and mitosis, due to failure in cell division, will cause endo-reduplication leading to polyploidy, that only occasionally is included in normal developmental programs (Lee*, et al.* 2009, Ullah*, et al.* 2009).

In this thesis we investigate the evolutionary conservation of the molecular machinery orchestrating DNA replication initiation both in

yeast and in mammalian cells, considering the budding yeast Saccharomyces cerevisiae as a paradigm for eukaryotic DNA replication. We then consider a molecular mathematical model of DNA replication, recently proposed by our group in a collaborative project (Brummer, *et al.* 2010), as a frame of reference to discuss similarities and differences observed in the regulatory program controlling DNA replication initiation in yeast and in mammalian cells and discuss whether they may be dependent upon different functional constraints.

Table 2.1: DNA replication in yeast and mammalian cells

	Genome size (base pairs)	Duration of S phase (min)	Early replication origins	Replication sites	Rate of DNA replication	Onset of DNA Replication
Saccharomyces cerevisiae	12×10^6 haploid genome	20-30	143	576 (potential) 351 (confirmed)	50 bp/sec	Coherent
Mammalian Cells	3200×10^6 diploid genome	300-600	~ 1700 regions	~ 50.000	30 bp/sec	Not Coherent

2.2 Comparative analysis of the molecular events involved in the onset of DNA replication

2.2.1 Replication origins and origin selection

In eukaryotes DNA replication initiates from hundreds to thousands of replication origins, in order to efficiently duplicate the genome (Table 2.1).

In *S. cerevisiae*, a replication origin consists of a short, genetically defined DNA element (100-200 bp), named Autonomously Replicating Sequence (ARS) (Mechali 2010). Each ARS contains several specific elements such as the A element, an 11 bp consensus sequence rich in adenines and thymines that is essential for initiation (Gilbert 2001). In the genome of budding yeast over 700 potential ARS have been

identified and classified in the *OriDB* database (www.oridb.org) (Nieduszynski, *et al.* 2007). Of these potential replication origins only approximately half constitutes replication initiation sites (Wyrick, *et al.* 2001) that in a defined timing program of activation of early and late origins allows duplication of the entire genome (12 million of base pairs) in a short time, in the order of 20-30 min under most growth conditions (Lord and Wheals 1980, Rivin and Fangman 1980, Sclafani and Holzen 2007, Vanoni, *et al.* 1983) (Table 2.1).

In mammalian cells, replication origins are genomic locations defined by a combination of epigenetic features, including local chromatin structures and complex architecture of chromosome folding in the nuclear space (Karnani, *et al.* 2010, Mechali 2010). In particular, initiation regions contain AT rich elements and CpG islands and are in proximity to DNAse I-hypersensitive sites and regions enriched in H3K4-(di- and tri)-methylation and H3 acetylation modifications on histones typical of open chromatin. Furthermore early initiation regions generally map in proximity of transcription start sites containing RNA polymerase II binding sites, and to high density transcription factor binding region clusters that characterize a nucleosome-free local environment.

In metazoans replication initiation sites are not uniformly scattered throughout the genome (Gilbert 2001): at some loci, such as the human gene encoding Lamin B2 (Todorovic, *et al.* 1999), initiation sites are localized within a few kbp, while at other loci, such as the Chinese hamster ovary (CHO) dihydrofolate reductase (DHFR) locus, multiple dispersed origins can be identified throughout "initiation zones" of 10 to 50 kbp (Dijkwel, *et al.* 2002).

It has been demonstrated that the topology around the most well-characterized mammalian replication origin mapped to a ~500 bp region 3' of the gene Lamin B2 (Giacca, *et al.* 1994) - determined by the competitive cell-cycle regulated binding to the two topoisomerases I and II - affects its activation. In particular topoisomerase II seems to

contribute to origin licensing during middle G1, while topoisomerase I to initiation of DNA replication (Abdurashidova, *et al.* 2007). Accordingly inhibition of topoisomerase I activity abolishes origin firing (Abdurashidova, *et al.* 2007).

Starting from approximately 30,000 to 50,000 replication origins, mammalian cells (Table 2.1) replicate their chromosomes in a variable time period, from 5 to 10 h, depending on the cell type and the differentiation/developmental status (Mechali 2010). Details on the timing program of origins activation are given in section 2.2.4. In the lower metazoan *Xenopus*, the genome in fertilized eggs is duplicated within 15-30 min probably by activating simultaneously all or almost all the replication origins, while in somatic/differentiated cells is duplicated in 5-10 h following a specific program of activation of different clustered origins (Callan 1973, Newport and Kirschner 1982).

A trend for an increase in S phase length during development is apparent also in mammalian cells, although it is less dramatic with a 20% increase from the first to the second cycle after fertilization (from 5 to 6 h) arriving at a maximum of about 10 h in somatic cells (Ferreira and Carmo-Fonseca 1997).

In conclusion while the S phase of yeast cells seems to be optimized for coherence of firing of origins resulting in a S phase length that is almost uncompressible, metazoans may have less evolutionary pressure on coherence and hence on a fast S phase.

2.2.2. Licensing

In all eukaryotes, licensing consists in the assembly of the pre-Replication Complexes (pre-RCs) at the replication origins through chromatin loading of several conserved replication factors: the Origin Recognition Complex ORC, Cdc6, Cdt1 and the Mini-Chromosome Maintenance complex MCM2-7. Proteins involved in licensing in yeast and mammals are compared in Table 2.2, while Fig. 2.1 provides a graphical time-course outline of the process in both organisms.

The first licensing event is the recruitment to the replication origins of the ORC complex (Fig. 2.1A,B) which in turn recruits Cdc6 (Fig. 2.1C,D). Then, Cdt1 facilitates clamping around the DNA of the ring-shaped MCM2-7 complex, a catalytic core of replicative helicase that unwinds DNA at the replication forks (Fig. 2.1E,F). The pre-RCs are kept inactive by Cdc6 and Cdt1, that will then be released following a ATPase hydrolysis-dependent mechanisms (Tsakraklides and Bell 2010) (Fig. 2.1G,H).

While these events are common to both organisms, in mammalian cells additional factors contribute to regulate pre-RC assembly: HMGA1, MCM8, HOXD13, Geminin and MCM9 (Fig. 2.1B, D, F, H and table 2.2). Also other Hox proteins, such as HOXC13 participate in the mammalian licensing process (Marchetti, *et al.* 2010), even if their role has yet to be fully clarified.

Both in yeast and in mammalian cells several replication factors promoting pre-RCs assembly are regulated by Cdk-dependent phosphorylation that induce either degradation, nuclear export or both (see section 2.4), thus avoiding the potential assembly of pre-RCs during S phase on newly synthesized DNA in order to prevent re-replication events (Blow and Dutta 2005).

In summary the logic of the simultaneous assembly of pre-RCs on the potential replication initiation sites in yeast and mammals is quite similar although higher eukaryotes have evolved additional controllers to more tightly regulate the licensing process.

2.2.3. Firing

Proteins involved in firing in yeast and mammals are compared in Table 2.3, while Fig. 2.2 provides a graphical time-course outline of the process in the two organisms. The first event of firing process in all eukaryotes is the activation of chromatin-bound MCM2-7 obtained by phosphorylation mediated by S-phase Cdks and Ddk (activation of the pre-RCs, Fig. 2.2A,B). This event is followed by the Cdk-regulated

recruitment onto the activated pre-RCs of essential initiation factors, such as Cdc45 (Fig. 2.2 C,D), GINS complex (*i.e.,* Sld5, Psf1, Psf2, and Psf3) and DNA polymerases (Fig. 2.2E-H), that can promote DNA unwinding, establishment and progression of the replication forks.

In particular Cdc45, Ddk-phosphorylated MCM2-7 and GINS constitute the CMG complex which provides the helicase activity ahead at the replication fork in both yeast and mammalian cells (Aparicio, *et al.* 2006, Kanemaki and Labib 2006, Moyer, *et al.* 2006, Pacek and Walter 2004), while the replication of unwound DNA is essentially performed by the combined action of three multi-subunit DNA polymerases [as reviewed in (Johansson and Macneill, 2010)]. The hetero-tetrameric DNA polymerase alpha, containing the DNA primase activity, initiates the replication on both the leading and lagging strands, and promotes the recruitment of factors, such as RFC and PCNA, that tether to the DNA template – and stimulate the polymerase activity and processivity of – the DNA polymerase epsilon and delta that participate in the synthesis of the leading and lagging strand respectively.

In budding yeast Cdc45 is loaded onto the activated pre-RCs by Sld3 in a S-Cdk-dependent manner (Muramatsu, *et al.* 2010) (Fig. 2.2C), and such recruitment is facilitated by the interaction between Cdc45 and MCM10 (Sawyer, *et al.* 2004) (Fig. 2.2C).

The yeast next event, *i.e.* the recruitment of GINS and DNA polymerase, in yeast is a two-step process that involves the formation of a chromatin-unbound multi-molecular complex, named pre-Loading complex (pre-LC), constituted by S-Cdk-phosphorylated Sld2, Dpb11, GINS and DNA polymerase epsilon (Fig. 2.2E), followed by the recruitment of such a complex onto the chromatin by the Cdk-phosphorylated Sld3 (Fig. 2.2G).

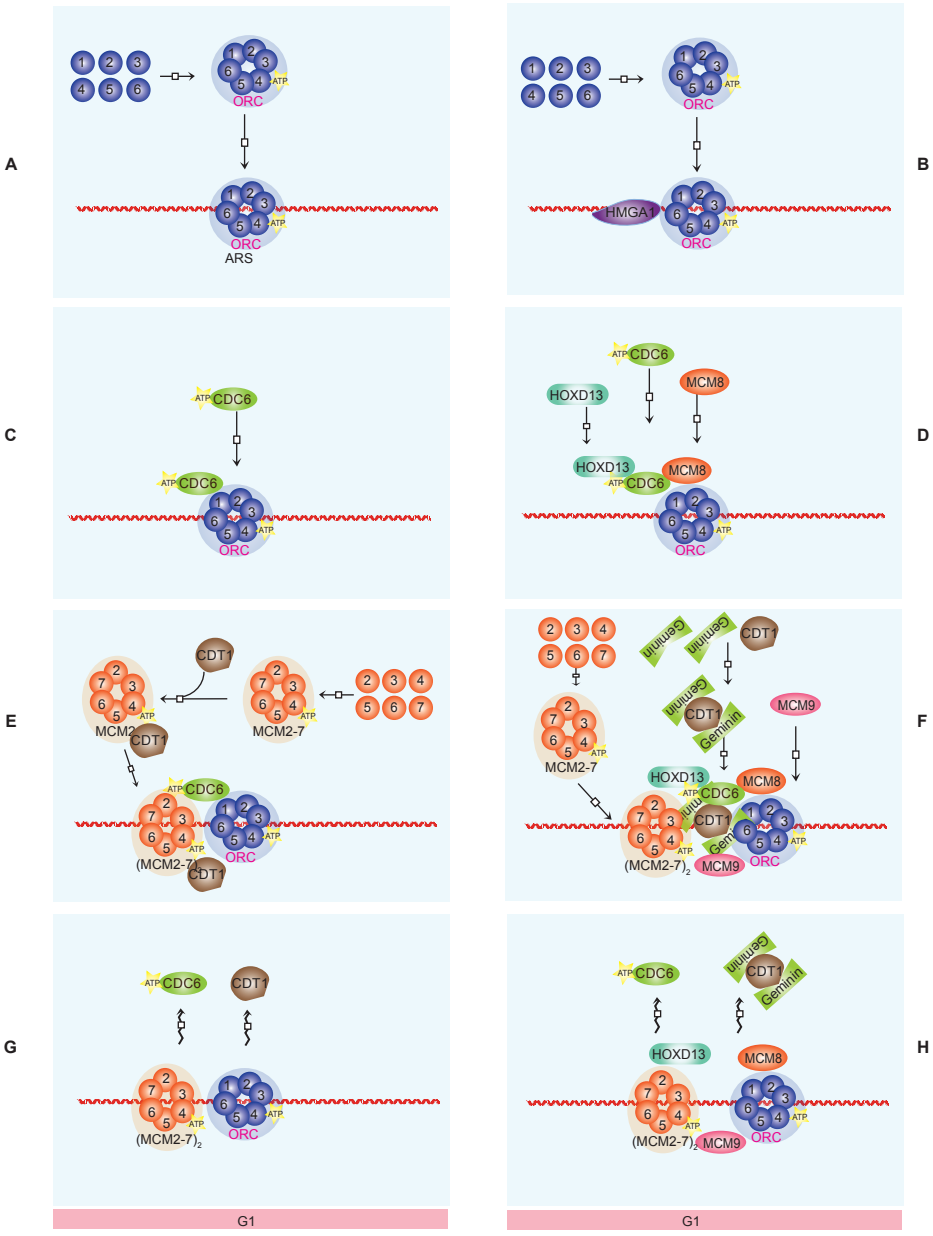


Fig. 2.1 Schematic map of *Licensing* events in budding yeast and in mammalian cells. See text for details. Arrows indicate association and dissociation events. Cdk-mediated phosphorylation events regulating level and activity of Licensing factors have been omitted and are shown in Fig. 2.7.

Table 2.2

Comparative analysis of licensing factors in budding yeast and in mammalian cells

	Proteins/ factors	Homology	Function	Known interactors involved in DNA replication initiation	Regulation	N. molecules per cell	References
Budding yeast	ORC	Origin recognition complex (ORC) composed of 6 polypeptides (Orc1-6) conserved in all eukaryotes belonging to the ORC1-6 family. ATPase AAA+ proteins (Orc 1-5)	binds to genetically defined replication origins, ARS, in ATP dependent manner creates a scaffold for recruitment of other replication factors	Cdc6; Cdt1 (via Orc6); MCM2 (via Orc2); MCM10 (via Orc1,2,6); Clb5 (via Orc6)	is stable bound to chromatin binding of Orc6 to Clb5 and its phosphorylation by Clb5-Cdc28 prevents its binding to other licensing factors	3930±530 Orc1 ; 1700±82.5 Orc2; 2170±90.9 Orc4; 768±163 Orc5; 2970±143 Orc6	(DePamphilis, 2005, Duncker et al. , 2009)
Mammalian cells			binds to replication origins defined by a combination of epigenetic features in an ATP-dependent manner creates a scaffold for recruitment of other replication factors	CDC6 (via Orc1, Orc6); MYST2/HBO1 (via Orc1); Dbf4 of DDK (via Orc2,4,6); MCM10 (via Orc2); HMGA1 (via Orc6)	is bound to chromatin during M/G1 transition and G1 phase phosphorylation of Orc1 by CycA-Cdk1 causes its export from the nucleus in some cells Orc1-CycA complex is degraded through SCF-Skp2-dependent mechanism	N.D.	
Budding yeast	Cdc6	Cell division control protein 6 is conserved in all eukaryotes belongs to the Cdc6/Cdc18 family. shows sequence similarity to Orc1 ATPase AAA+ protein	binds to ORC:DNA:chromatin for constituting preRC mediates chromatin recruitment of MCM2-7 participates in Chk1-mediated checkpoint controls that ensure DNA replication is completed before mitosis is initiated.	Orc1; Cdc28; MCM2 (may not be direct)	upon S-Cdk-dependent phosphorylation is degraded through Skp1-Cullin-Fboxprotein dependent mechanism Cdk-phosphorylated Cdc6 bound to Clb2 is not able to bind to ORC	N.D.	(Tsakraklides and Bell, 2010)
Mammalian cells				Orc1; PCNA (indirect); cyclin-CDK; HOXD13; MCM8	phosphorylation by CycE-Cdk2 prevents its degradation upon phosphorylation by CycA-Cdk2 is exported from the nucleus	N.D.	(Paolinelli et al. , 2009)
Budding yeast	Cdt1	Conserved in all eukaryotes	binds to ORC:DNA:chromatin and cooperates with Cdc6 to promote chromatin loading of MCM2-7 to form preRC	MCM2-7; Topoisomerase	upon S-Cdk-phosphorylation is exported from the nucleus	2190	(Devault et al. , 2002, Tanaka and Diffley, 2002)

Mammalian cells				Orc2; Cdc6; MCM2-7; Geminin; MCM9; Ddk; CycA-Cdk; Hbo1; Topoisomerases I and II; PCNA	during S and G2 phase is degraded upon ubiquitination by three different ubiquitin ligases: 1. SCF-Skp2, targeting Cdt1 proteins phosphorylated by CycA-Cdk2 2. Cul4–DDB1–CDT2, requiring the presence of PCNA 3. APC	N.D.	(Cook et al. , 2004, De Marco et al. , 2009, Wohlschlegel et al. , 2000)
Budding yeast	MCM2-7	Minichromosome maintenance complex composed by six polypeptides MCM 2-7 conserved in all eukaryotes ATPase AAA+ proteins	provides essential helicase function in eukaryotes at replication forks.	Ddk; Cdt1; Cdc6; GINS; Hbo1; cdc45	phosphorylation of the N-terminal tails of Mcm2–4–6 by Ddk allows the stable formation of the Cdc45–MCM 2-7 complex and promotes the initiation of DNA replication phosphorylation of chromatin-bound MCM2 by Ddk during G1/S transition promotes the initiation of DNA replication.	MCM2 1690 MCM3 35100 MCM4 8800 MCM5 10300 MCM6 13400 MCM7 16600	(Bochman and Schwacha, 2009)
Mammalian cells				Cdt1; RPA; p27Kip1;		N.D.	
Mammalian cells	MCM8	Minichromosome maintenance 8, belongs to MCM family No known homologues in yeast	promotes preRCassembl,likely facilitating chromatin loading of Cdc6	Cdc6; Orc2; MCM4, 6 and 7	N.D.	N.D.	(Kinoshita et al. , 2008, Volkening and Hoffmann, 2005)
Mammalian cells	HOXD13	Homeobox protein Hox-D13 Belongs to the Abd-B homeobox family.	promotes pre-RC assembly likely facilitating chromatin loading of Cdc6 antagonist of Geminin	Cdc6; Geminin	N.D.	N.D	(Salsi et al. , 2009)
Mammalian cells	MCM9	Minichromosome maintenance 9, belongs to MCM family No known homologues in yeast	forms a stable complex with the licensing factor Cdt1, preventing an excess of geminin on chromatin during the licensing reaction also essential activating linker between Cdt1 and the MCM2-7 complex	Cdt1; Geminin	N.D.	N.D.	(Lutzmann and Mechali, 2008)
Mammalian cells	Geminin	belongs to Geminin family No known homologues in yeast Homotetramer	interacts with CDT1 and inhibits its ability to recruit MCM2-7 to origins prevent Cdt1 degradation	Cdt1; HoxD13; Cyclins	is degraded by APC/C dependent mechanism during G1 phase	N.D.	(De Marco, Gillespie, 2009, Wohlschlegel, Dwyer, 2000)
Mammalian cells	HMGA1	Belongs to the HMGA family	preferentially binds to the minor groove of A+T-rich regions in double-stranded DNA. helps to ORC to recognize origins	ORC	N.D.	N.D.	(Thomae et al. , 2008)

[a]Besides the literature cited in the text, the reader is also directed to http://www.dnareplication.net, a freely available database on DNA replication in eukaryotes (Cotterill and Kearsey, 2009)

In mammalian cells, Cdc45 is loaded onto the activated preRCs in a S-Cdk-dependent manner by Treslin—a protein bearing some sequence conservation with yeast Sld3 (Sanchez-Pulido, *et al.* 2010, Kumagai, *et al.* 2010) (Fig. 2.2D). Other proteins contributing to Cdc45 loading in metazoans include MCM10 (Fig. 2.2B) and TopBP1/Cut5. While a role in Cdc45 loading has been supported by several congruent data for *Xenopus* Cut5, the role of its ortholog TopBP1 in human cells remains less defined. Other proteins involved in Cdc45 loading, apparently not present in yeast and whose function seems to be regulated by S-Cdk include DUE-B (Chowdhury, *et al.* 2010) and GEMC1 (Balestrini, *et al.* 2010, Piergiovanni and Costanzo 2010) (Fig. 2.2D).

Another critical step of firing in mammalian cells is the MCM10 dependent recruitment to origins of the essential initiation factor RecQL4, a protein that shares with Sld2 a region of homology (Fig. 2.2F). Loading of GINS, in mammalian cells, is not achieved by the formation of a pre-LC - as indeed has also been proposed in yeast (Gambus, *et al.* 2006) - , but is mediated by direct interaction with RecQL4 (Fig. 2.2H). DNA polymerase epsilon is then recruited to the origins by GINS, while the recruitment of DNA polymerase alpha is mediated by And-1, in a RecQL4/MCM10 dependent mechanism (Fig. 2.2H).

2.2.4. Timing program of origin activation

Both in yeast (Raghuraman, *et al.* 2001, Yabuki, *et al.* 2002) and in metazoans (Desprat, *et al.* 2009, Drouin, *et al.* 1990, Natsume and Tanaka 2010, Woodfine, *et al.* 2004) DNA replication initiation is temporally and spatially regulated. While licensing occurs simultaneously at different replication origins during G1 phase, firing occurs following a specific timing program of initiation on the different licensed origins, and the origins that fire simultaneously (during a defined interval of S phase) are clustered in similar regions of the chromosome or sub-chromosomal domains, in mammalian cells called replication foci (Jackson and Pombo 1998, Lebofsky, *et al.* 2006, Ma, *et al.* 1998, Maya-Mendoza and Jackson

2009, Sadoni, *et al.* 2004, Taddei, *et al.* 2004, Takebayashi, *et al.* 2001). Replication foci are structural units of chromosomes stable during cell cycle that contain groups of adjacent replicons that fire quite synchronously (coherently), according to a tightly controlled temporal program during S phase. The size and number of replicons per focus is heterogeneous: most foci contain small replicons that are replicated in less than an hour (on ten hours of S-phase), others contain very large replicons, in some cases larger than 1Mbp, that require most of S-phase length to be replicated (Berezney, *et al.* 2000). Therefore in all eukaryotes DNA replication proceeds through the coordinated firing of a large number of spatially clustered origins of replication. In general foci in transcriptionally active chromatin with high gene density are engaged in replication during early S-phase, while foci in transcriptionally inactive heterochromatin replicate later (Cadoret, *et al.* 2008, Desprat, *et al.* 2009, Hansen, *et al.* 2010, Hiratani, *et al.* 2008, Sequeira-Mendes, *et al.* 2009). Therefore there is a spatial compartimentalization of early and late replicating DNA sequences that likely is defined during each G1 phase, at the timing decision point, with the anchorage of chromosomal segments at specific locations within the nucleus (Dimitrova and Gilbert 1999, Li, *et al.* 2001). Foci that are replicated during consecutive intervals of S-phase generally are adjacent (Sporbert, *et al.* 2002) and the spatial continuity within individual chromosomes seems to determine the sequential activation of adjacent replicon clusters, whose timing program defines the progression of the S-phase (Maya-Mendoza, *et al.* 2010).

In metazoans replication timing domains are reorganized during development and differentiation (Hiratani, *et al.* 2010, Hiratani, *et al.* 2008), as described in section 2.2.1.

2.2.5. Coherence of origin activation

A recently described model of replication in yeast, possibly extensible to mammalian cells and validated by genome-wide replication time course

data (Yang, *et al.* 2010) describes origin firing as a stochastic process in which the regulation of the timing of origin firing can be explained by the number of initiators loaded at each origin. In particular, earlier-firing origins should have more initiator complexes (authors refer to MCM2-7) loaded, and a more accessible chromatin environment. Accordingly, it has been reported that on average there are six times more MCM2-7 loaded on efficient origins than on inefficient origins (Wyrick, *et al.* 2001). Hence the firing probability of all origins should increase during S-phase to ensure that origins with relatively low firing probability become likely to fire in late S phase, as discussed in (Rhind, *et al.* 2010).

In yeast large groups of neighboring origins and origins in similar regions of the chromosomes (e.g., centromeric origins) fire coherently: of 250 initiation sites 143 origins are reported to fire early during S-phase (Yabuki, *et al.* 2002), and they are rapidly activated within a ca. 10 min interval (Raghuraman, *et al.* 2001, Yabuki, *et al.* 2002). The molecular (Muramatsu, *et al.* 2010, Tanaka, *et al.* 2007, Zegerman and Diffley 2007) and kinetic (Brummer, *et al.* 2010, Salazar, *et al.* 2010) basis of such coherent firing have been elucidated, as described in detail in section 2.5 and depend upon the coordinated activation by S-Cdk of the initiation factors Sld2 and Sld3.

Furthermore the number of origins coherently activated in early S-phase reflects- and likely depends upon - the availability in each cell of the molecules of the limiting essential initiation factor Sld3 [reported to be around 125 per cell (Ghaemmaghami, *et al.* 2003)], which is involved in the recruitment to the origins of the essential initiation/elongation factor Cdc45 (see section 2.2.3 and 2.3.2.1.). Other initiation factors appear to be present in the cell in no limiting amount, as reported by (Ghaemmaghami, *et al.* 2003). Therefore in order to activate all initiation sites each Sld3 molecule should associate first with an early origin thereby promoting firing, then dissociate and associate again with a late-firing origin. This indicates a catalytic role of Sld3 on different origins

so to maintain the coherence of firing of replication origins during S-phase, as described by a computational kinetic model (Brummer, *et al.* 2010) and as discussed in (Araki 2010a).

Similarly in metazoans a large number of spatially clustered origins of replication fire coherently, in particular in early embryonic cell cycles where thousands of replication origins are activated in a short time interval (10-20 min) (Goldar, *et al.* 2008). In differentiated cells such coherence characterizes a limited number of origins (early origin), possibly due to a shortage of a key component of the replication machinery.

The molecular and kinetic basis of the coherence of the replication initiation in metazoans needs to be elucidated.

2.3 Comparative analysis of proteins involved in the onset of DNA replication

In this section we report and compare molecular details on proteins involved in licensing and firing in yeast and in mammalian cells (see Tables 2.2 and 2.3). For a few poorly conserved proteins that play a critical role in regulation of DNA replication initiation in yeast, a more in-depth analysis is presented.

2.3.1 Licensing

2.3.1.1 ORC binding to DNA replication origins

Both in yeast and mammals the first step in pre-RC assembly involves recruitment onto the replication origins of the Origin Recognition Complex (ORC) composed of six polypeptides (Orc1-6) conserved in all eukaryotes (Bell and Stillman 1992, Duncker, *et al.* 2009, Gavin, *et al.* 1995) (Fig. 2.1A,B). While yeast ORC recognizes and binds to the short genetically defined DNA elements ARS (Mechali 2010), mammalian ORC binds to genomic locations defined by a combination of epigenetic features (Karnani, *et al.* 2010, Mechali 2010) (see 2.1). In mammalian cells, the metazoan-specific HMGA1 protein helps ORC to recognize AT rich DNA at the origin (Thomae, *et al.* 2008). HMGA1 is a DNA-binding

High-Mobility Group non-histone Protein involved in many cellular processes, by regulating the transcription of many genes (Resar 2010).

In yeast the six ORC subunits form a stable complex, whose 3D structure has been recently solved (Chen, *et al.* 2008), while mammalian ORC consists of a stable core of Orc2 to Orc5 subunits, weakly bound to Orc1 and Orc6 (Ghosh, *et al.* 2011, Giordano-Coltart, *et al.* 2005). Orc1-5 belong to the AAA+ (ATPases associated with a variety of cellular activities) family, that contains a region of sequence similarity including the Walker A and B motifs common to many nucleotide binding proteins (Patel and Latterich 1998). In all eukaryotes specific subunits of ORC can bind ATP: Orc1 and 5 in yeast (Klemm and Bell 2001) and Orc1,4 and 5 (Siddiqui and Stillman 2007) in human cells, and ATP binding is involved in ORC complex assembly. However, only Orc1 possesses ATPase activity, in yeast stimulated by ss-DNA (Lee, *et al.* 2000). It has been proposed that in yeast the ATPase activity of ORC may be involved in chromatin loading of the MCM2-7 helicase (Bowers, *et al.* 2004, Randell, *et al.* 2006).

Yeast ORC is constitutively bound to the origins (DePamphilis 2005). Mammalian ORC is bound to chromatin only during M/G1 transition and during G1 phase. In mammalian cells during M/G1 transition and G1 phase, Orc1 is unphosphorylated, its affinity for chromatin is high and the ORC complex is tightly bound to chromatin (DePamphilis 2005). With the onset of S-phase upon S-Cdk phosphorylation, Orc1 is targeted for nuclear export and in some cases for protein degradation by SCF^{Skp2} (Mendez, *et al.* 2002, Ohta, *et al.* 2003, Tatsumi, *et al.* 2003) (see section 2.4).

Table 2.3
Comparative analysis of Firing Factors in budding yeast and in mammalian cells

	Protein/Factor	Homology	Function	Known interactors involved in DNA replication initiation	Regulation	N. molecules per cell	References
Budding yeast	MCM10	• Minichromosome maintenance 10 • belongs to MCM 10 family	• regulates chromatin association of Pol alpha.	Orc; Mcm2-7; Cdc45; Pol alpha; PCNA	N.D.	1860	(Eisenberg et al. , 2009)
Mammalian cells			• regulates chromatin association of Pol alpha • promotes recruitment of RecQ4 and regulates its DNA unwinding activity. • likely promotes chromatin loading of CDC45.	Orc2; Mcm2/6; Pol alpha; AND1/Ctf4; RecQ4; Cdc45	• is temporarily recruited to the replication sites 30–60 min before replication • dissociates from chromatin after the activation of the Pre-RC	N.D.	(Xu et al. , 2009)
Budding yeast	Ctf4/ And-1	• DNA polymerase alpha-binding protein • Contains several WD repeats	• not essential for yeast viability • is involved in chromatin recruitment of Pol alpha	Pol alpha; CDC45; GINS; DIA-2 F-box protein component of a SCF	• is degraded through SCF-Dia2 dependent mechanism	3280	(Gambus et al. , 2009)
Mammalian cells			• is required for CMG assembly • is required for chromatin recruitment of Pol alpha • is involved in tethering MCM2-7-RecQ4-Pol alpha at replication forks	Pol alpha; MCM10	N.D.	N.D.	(Im et al., 2009)
Budding yeast	Sld2	• (Synthetically lethal with Dpb11-1) factor 2 • Belongs to Sld2 family	• when Cdk phosphorylated promotes the Pre-LC assembly and initiation of DNA replication • required for the proper activation of RAD53 in response to DNA damage and replication blocks.	Dpb11	• is phosphorylated by S-CDK at the onset of S-phase. • phosphorylation of Thr-84 promotes interaction with Dpb11 leading to pre-LC assembly	656	(Muramatsu et al. , 2010, Tanaka et al. , 2007, Zegerman and Diffley, 2007)
Mammalian cells	RecQ4/RecQL4	• ATP-dependent DNA helicase Q4 • belongs to RecQ sub-family • shares sequence similarity in N terminal region with Sld2 • mutants correlated with Rothmund-Thompson-RAPADILINO and Baller Gerold Syndromes	• mediates chromatin recruitment of GINS complex • is likely involved together with MCM 2-7 in DNA unwinding at replication fork	MCM10; MCM2-7; CDC45; GINS	• may be multi phosphorylated by CDK • its helicase activity is regulated by MCM10	N.D.	(Im et al. , 2009, Thangavel et al. , 2010, Xu, Rochette, 2009)
Budding yeast	Sld3	• (Synthetically lethal with Dpb11-1) factor 3	• mediates CDK dependent chromatin recruitment of Cdc45 • mediates chromatin recruitment of pre-LC	Dbp11; Cdc45; Mcm2-7 in competition with GINS	• is phosphorylated on multisite by S-CDK • its phosphorylation modulates Dpb11 binding	125	(Araki, 2010, Muramatsu, Hirai, 2010)
Mammalian cells	Treslin	• belongs to Treslin family • shows sequence homology with Sld3	• mediates CDK-dependent chromatin loading of Cdc45 onto replication origins, likely collaborating with TopBP1	TopBP1; CDC45;Cyclin E-CDK2	• its phosphorylation by Cyc E-CDK2 modulates TopBP1 binding and chromatin loading of CDC45	N.D.	(Kumagai et al. , 2010, Sanchez-Pulido et al. , 2010)
Budding yeast	Dpb11	• DNA polymerase binding protein 11	• interacts with CDK phosphorylated Sld2 to form Pre-LC	Pol epsilon; Mcm10; Sld2 and Sld3 (via	N.D.	540	(Muramatsu et al. , 2010)

		contains 4 BRCT domains	mediates chromatin recruitment of Pre-LC complex	BRCT domains);psf1 of GINS			
Mammalian cells	TopBP1	DNA topoisomerase 2-binding protein II Contains 8 BRCT domains putative homologue of Dpb11	is not essential for CMG assembly is likely involved in chromatin recruitment of Pol epsilon.	Pol epsilon; Treslin; DUE-B; GEMC1; CDC45	is degraded through HectH9 ubiquitin dependent mechanism	N.D.	(Jeon et al. , 2007, Sokka et al. , 2010)
Mammalian cells	GEMC1	Geminin coiled-coil domain-containing protein 1 No known homologues in yeast	promotes initiation of DNA replication in multicellular organisms by mediating TopBP1- and Cdk2-dependent recruitment of Cdc45 onto replication origins	TopBP1; CDC45; Cyclin E-CDK2	its phosphorylation by Cyc E-CDK2 modulates TopBP1 binding and chromatin loading of CDC45	N.D.	(Balestrini et al. , 2010, Piergiovanni and Costanzo, 2010)
Mammalian cells	DUE-B	DNA unwinding element binding protein No known homologues in yeast	DUE-B interacts with TopBP1 and CDC45 Cdk2-dependent manner.	TopBP1; CDC45; Cyclin E-CDK2	its phosphorylation by Cyc E-CDK2 modulates TopBP1 binding and chromatin loading of CDC45	N.D.	(Chowdhury et al. , 2010)
Budding yeast	CDC45	Cell division control protein 45 homolog Conserved in all eukaryotes	forms CMG complex with CDC45 and Ddk-phosphorylated MCM 2-7, that provides helicase activity at replication forks moves with forks during replication elongation.	MCM 2,3.and 5; Pol epsilon; RPA; Dbp11; Sld3; GINS and MCM 2-7	N.D.	1730	(Aparicio et al. , 2009, Ballabeni et al. , 2009, Broderick and Nasheuer, 2009, Ilves et al. , 2010, Im, Ki, 2009)
Mammalian cells				Orc2, MCM7, TopBP1, Pol alpha, delta epsilon	is degraded through ubiquitin dependent mechanism	N.D.	Pollok 2007
Budding yeast	GINS	GINS from the Japanese 'Go -Ichi-Ni-San', which, in English, means 'five–one–two–three'. It composed by Sld5–Psf1–Psf2–Psf3 subunits conserved in all eukaryotes	forms pre-LC with CDK phosphorylated Sld2, Dpb11and Pol epsilon, that is recruited to origins by CDK phosphorylated Sld3 forms CMG complex together with CDC45 and Ddk phosphorylated MCM 2-7 moves with forks in replication elongation.	Dbp11; Pol epsilon (via Psf1); Mcm2-7 and Cdc45 (via Sld5) in competition with Sld3; Ctf4 (via Sld5); Mcm10 and TopoI;	N.D.	Psf1 1430 Psf3 2210	(Aparicio, Guillou, 2009, Bruck and Kaplan, 2011, MacNeill, 2010) (Muramatsu et al. , 2010)
Mammalian cells			is recruited to chromatin by RecQ4 forms CMG complex together with CDC45 and the MCM 2-7 move with forks.	Cdc45; MCM2-7; Pol alpha; RecQ4	N.D.	N.D.	
Budding Yeast	Sld7	(Synthetically lethal with Dpb11-1) factor 7 Conserved in fungi strictly correlated with budding yeast	likely stabilizes Sld3 protein reduces the affinity Sld3 for CDC45	Sld3	N.D.	N.D.	(Tanaka et al. , 2011)
Budding yeast	DNA Polymerase alpha	it is composed of 4 subunits: DNA pol catalytic subunit POL1	initiates the DNA replication on both leading and lagging strand	Mcm10	Phosphorylated in cell cycle dependent manner	N.D.	(Johansson and Macneill, 2010)

		structural subunit POL12 2 primase subunits PRI1 and PRI2				
Mammalian cells		it is composed of 4 subunits: catalytic subunit POLA1/p180 regulatory subunit POLA2/p70 2 primase subunits PRIM1/p49 and PRIM2/p58		Rb; RPA; Mcm10; And-1; GINS	Phosphorylated in cell cycle dependent manner	N.D.	(Burgers, 2009, Johansson and Macneill, 2010)
Budding yeast		it is composed of 3 subunits: catalytic subunit POL3 processivity factor binding subunit POL32 structural subunit POL31		PCNA	N.D.	N.D.	(Johansson and Macneill, 2010)
Mammalian cells	DNA Polymerase delta	it is composed of 4 subunits: pol catalytic subunit POLD1 processivity factor binding subunit POLD3 2 structural subunits POLD2 and POLD4	participates in the synthesis of the lagging strand	RPA; RFC/PCNA; Rb; GINS	N.D.	N.D.	(Johansson and Macneill, 2010)
Budding yeast	DNA Polymerase epsilon	it is composed of 4 subunits: catalytic subunit POL2 DPB2 DPB3 DPB4	participates in the synthesis of the leading strand forms pre-LC with CDK phosphorylated Sld2, Dpb11and GINS	Sld2; Dpb11; Psf1; Cdc45	N.D.	N.D.	(Johansson and Macneill, 2010) (Muramatsu et al. , 2010)
Mammalian cells		It is composed of 4 subunits: pol catalytic subunit POLE1 processivity factor binding POLE2 2 structural subunits POLE3 and POLE 4	participates in the synthesis of the leading strand	RFC/PCNA; GINS; TopBP1	N.D.	N.D.	(Johansson and Macneill, 2010) Bermudez 2011
Budding yeast	PCNA	proliferating cell nuclear antigen (PCNA) belongs to the PCNA family	functions as the sliding clamp for DNA polymerase delta	MCM10; Pol delta	N.D.	N.D.	(Motegi et al. , 2008)
Mammalian cells			increases the processivity of DNA polymerase delta and epsilon	Cdt1; RFC; Pol delta; Pol epsilon	N.D.	N.D.	
Budding yeast	RFC	Replication factor-C it is heteropentamer of RFC1--5	PCNA clamp loading	PCNA; RPA; Ctf4	N.D.	N.D.	Yao 2003 jbc Mayer 2001 Molcel
Mammalian cells		Replication factor-C it is heterotetramer of RFC2-5	PCNA clamp loading	PCNA; Pol epsilon; Pol delta	N.D.	N.D.	Gupte 2005 cell cycle Mossi 1997 jbc Zhang 1999
Budding yeast	RPA	Replication protein A it is composed of 3 subunits RPA1-3	Single-stranded DNA binding it plays a key role in DNA replication, recombination, transcription, checkpoints, telomere maintenance, and DNA repair	Cdc45; MCM2; Pol epsilon; MCM10	N.D.	N.D.	(Brush and Kelly, 2000)Longhese 1994)
Mammalian cells				MCM2-4; PCNA; Pol delta	N.D.	N.D.	(Weisshart 2004, Zou 2006)

[a] Besides the literature cited in the text, the reader is also directed to http://www.dnareplication.net, a freely available database on DNA replication in eukaryotes (Cotterill and Kearsey, 2009)

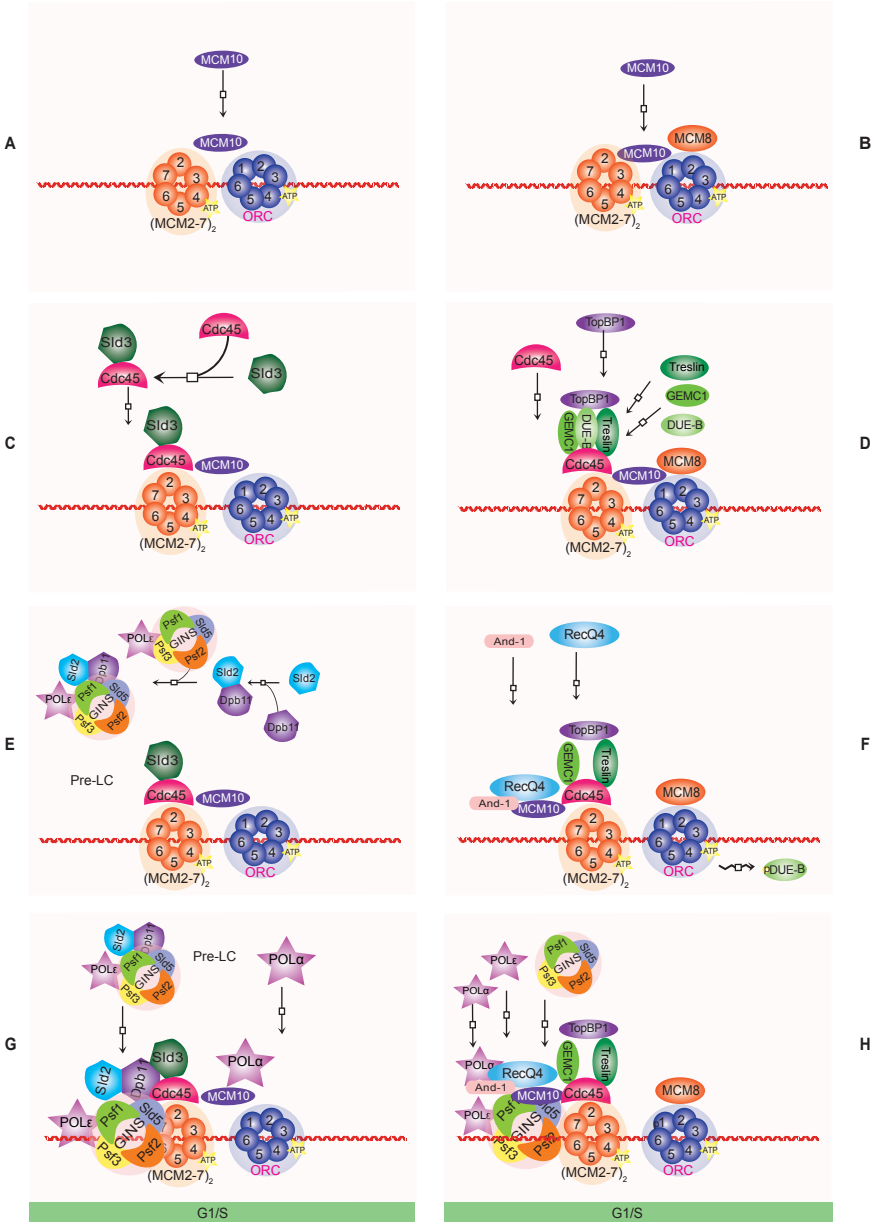


Fig. 2.2 Schematic map of *Firing* events in budding yeast and in mammalian cells. See text for details. Arrows indicate association and dissociation events. Ddk- and Cdk-mediated phosphorylation events regulating level and activity of Firing factors have been omitted and are shown in Fig. 2.7. For sake of clarity the loading of only one replication complex is shown, but DNA replication proceeds bidirectionally from each origin of replication.

2.3.1.2 Cdc6 and Cdt1 loading on ORC-chromatin

In all eukaryotes the next step of the pre-RCs assembly is the recruitment of the conserved replication factor Cdc6 – an AAA+ ATPase that shares sequence similarity with Orc1 - to ORC-chromatin sites (Borlado and Mendez 2008, Tsakraklides and Bell 2010) (Fig.2.1C,D). Chromatin recruitment of Cdc6 both in yeast (Semple, *et al.* 2006, Wang, *et al.* 1999) and in mammalian cells (Saha, *et al.* 1998) is mediated by Orc1. Recently it has been shown that human Cdc6 also binds to Orc6 (Thomae, *et al.* 2011). Recruiting of mammalian Cdc6 to chromatin also involves MCM8, a metazoan-specific protein (Kinoshita, *et al.* 2008, Volkening and Hoffmann 2005), that is also involved in replication elongation (Maiorano, *et al.* 2005). In human cells, phosphorylation of Cdc6 by CycE-Cdk2 promotes pre-RCs assembly, while Cdc6 acetylation by GCN5 promotes its chromatin dissociation (Paolinelli, *et al.* 2009).

Two Cdc6 functions rely on ATP hydrolysis: the ability of Cdc6 to influence the specificity of ORC association to DNA, so far demonstrated only in yeast (Mizushima, *et al.* 2000), and the actual loading of MCM proteins to OR (Borlado and Mendez 2008).

The conserved replication factor Cdt1 cooperates with Cdc6 in recruiting the mini-chromosome maintenance complex MCM2-7 (Fig. 2.1E,F) to the replication origins (Cook, *et al.* 2004, Devault, *et al.* 2002). While in yeast Cdt1 forms a stable complex with MCM2-7, whose nuclear localization is cell cycle regulated (Tanaka and Diffley 2002), in mammalian cells the ability of Cdt1 to recruit MCM2-7 onto the chromatin is modulated by the antagonist action of several metazoan-specific proteins, including MCM9 (Petropoulou, *et al.* 2008) and Geminin (McGarry and Kirschner 1998, Montanari, *et al.* 2006, Saxena and Dutta 2005, Wohlschlegel, *et al.* 2000).

MCM9 is required for the loading of the MCM2-7 helicase onto DNA replication origins (Lutzmann and Mechali 2008). Geminin binds to Cdt1

to give rise to a complex that can exist in two distinct forms, a "permissive" heterotrimer and an "inhibitory" heterohexamer, in which residues of Cdt1 important for licensing are buried. It has been postulated that the transition between the heterotrimer and the heterohexamer may represent a molecular switch between licensing-competent and licensing-defective states (De Marco, *et al.* 2009). Furthermore, Geminin binding to HOXD13 - that belongs to the family of homeodomain-containing DNA binding proteins regulating processes dealing with the acquisition and maintenance of cell identity (Duboule 1995) and that binds to Cdc6 promoting pre-RC assembly (Salsi, *et al.* 2009) - antagonizes HOXD13 licensing action (Salsi, *et al.* 2009). In mammalian cells, loading of MCM2-7 is also facilitated by the histone acetyltransferase HBO1 (Iizuka, *et al.* 2006) that can directly interact with Cdt1 (Miotto and Struhl 2008) whose acetylation protects it from ubiquitination and subsequent proteasomal degradation (Glozak and Seto 2009) and promotes licensing (Miotto and Struhl 2010).

Geminin levels are cell-cycle regulated: Geminin is active during S and G2 and early M phases, and is inactivated by APC-dependent ubiquitination during late M and G1 phase.

2.3.1.3 Chromatin loading and activation of MCM2-7

MCM2-7 is the eukaryotic replicative helicase core that during firing unwinds DNA ahead of the replication forks, either alone or helped by additional initiation factors - Cdc45 and GINS - loaded onto replication origins (Bochman and Schwacha 2008, Labib and Diffley 2001, Moyer, *et al.* 2006). It is therefore an essential replication factor for both initiation and elongation events of DNA replication.

MCM2-7 is an heterohexamer composed by the six proteins MCM2, MCM3, MCM4, MCM5, MCM6 and MCM7 conserved in all eukaryotes (Mendez and Stillman 2000) belonging to the ATPase AAA^+ family. The MCM proteins oligomerize into a ring-shaped complex and the ATPase active sites contributing to DNA unwinding are formed at dimer

interfaces, as reviewed in (Bochman and Schwacha 2009). Before loading, MCM2-7 is a single hexamer in solution that is converted into a double-hexamer during pre-RC formation (Evrin, *et al.* 2009).

Loading of MCM2-7 onto chromatin is promoted by several replication factors (Fig. 2.1E,F), as described in 3.1.2. and once helicase is loaded onto chromatin the other components of pre-RC are dispensable for replication initiation (Maiorano, *et al.* 2000). Loading of yeast MCM2-7 is accompanied by ATP hydrolysis-dependent dissociation of ORC, Cdc6, and Cdt1 from origin DNA (Tsakraklides and Bell 2010) (Fig. 2.1G,H).

The helicase MCM2-7 complex is loaded at origins in an inactive form, during G1 phase and is activated at G1/S transition and throughout S-phase by Cdk- and Ddk-phosphorylation (Bochman and Schwacha 2009, Labib 2010, Remus and Diffley 2009) (Fig. 2.7D). Such phosphorylation promotes in yeast the stable association between MCM2-7 and the essential initiation protein Cdc45 (Francis, *et al.* 2009, Masai, *et al.* 2006) and in mammalian cells the formation of the CMG helicase complex (Im, *et al.* 2009). Recently, the structural basis for MCM2-7 helicase activation by GINS and Cdc45 has been clarified (Costa, *et al.* 2010).

Both in yeast and in mammalian cells there is a large excess of MCM2-7-loaded licensed origins over the number of origins effectively selected for the replication (fired) during S-phase (Ge, *et al.* 2007, Hyrien, *et al.* 2003, Ibarra, *et al.* 2008, MacAlpine and Bell 2005, Rhind 2006, Woodward, *et al.* 2006). This excess of licensed origins allows cells to promptly respond to replicative stresses that cause fork failure, by using dormant origins (Doksani, *et al.* 2009). If chromatin loading of MCM2-7 is reduced, fewer replication origins are used resulting in DNA strand breaks, checkpoint activation and genomic instability (Chuang, *et al.* 2010). In normal conditions dormant origins are passively replicated by forks initiated at adjacent origins (Mechali 2010).

2.3.2 Firing

Once pre-RCs are assembled onto the replication origins, S-phase specific Cdk (Cyclin- dependent kinases) and Ddk (Dbf4 dependent Kinase, whose catalytic and regulatory subunits are, respectively, Cdc7 and Dbf4) activities are required in order to activate those complexes from which bidirectional replication forks will start following activation by specific initiation factors.

While the assembly of the pre-RCs occurs simultaneously onto the different replication origins, the firing process occurs on several licensed origins at distinct but reproducible times during S-phase (Donaldson 2005, Goren and Cedar 2003, Zink 2006), as discussed in section 2.2.4. Therefore the assembly of the replication factors promoting the replication initiation (firing) is temporally regulated, *i.e.*, the initiation factors bind first to early firing origins.

2.3.2.1 Chromatin loading of MCM10 and Cdc45

In all eukaryotes MCM10 is the earliest initiator factor recruited onto the pre-RCs (Fig. 2.2A,B). In mammalian cells the cellular level and chromatin association of MCM10 are cell cycle regulated, being higher in S-phase, similarly to ORC (Izumi, *et al.* 2000, 2001).

Both in yeast and in human cells MCM10 interacts with ORC and MCM2-7 (Izumi, *et al.* 2000, Kawasaki, *et al.* 2000, Lee, *et al.* 2003, Merchant, *et al.* 1997, Warren, *et al.* 2009, Xu, *et al.* 2009) and plays a key role in regulating chromatin association and stability of DNA Polymerase alpha that contains primase activity involved in initiation of DNA replication on both leading and lagging strands (Chattopadhyay and Bielinsky 2007, Lee, *et al.* 2010, Ricke and Bielinsky 2004).

Cdc45 - a conserved, essential protein for both initiation and elongation of DNA replication, (Pospiech, *et al.* 2010) - can be considered a marker for origin activation, since its pattern of tight association with origins is coincident with their temporal order of firing (Aparicio, *et al.* 1999, Zou and Stillman 2000). Cdc45 can bind directly MCM10 at the initiation sites in yeast (Fig. 2.2C) (Gregan, *et al.* 2003, Sawyer, *et al.* 2004), *Xenopus*

(Wohlschlegel, *et al.* 2002) and likely in mammals (Fig. 2.2D), since *Xenopus* and human proteins share 78% of identity on all amino acid sequence. Human Cdc45 is also able to directly bind MCM2 and MCM6 of the activated pre-RCs (Ilves, *et al.* 2010)

Additional factors are nevertheless involved in chromatin loading of Cdc45 in both lower and higher eukaryotes. In yeast, Sld3 mediates loading of Cdc45 upon phosphorylation by S-Cdk, (Tanaka, *et al.* 2007, Zegerman and Diffley 2007). Then by binding to the two BRCT domain of Dpb11, plays a key role in recruitment of GINS complex, as described in section 2.*3.2.2*. In mammalian cells Treslin, that shares partial sequence conservation with yeast Sld3, can promote Cdk-dependent chromatin loading of Cdc45, and bind in a S-Cdk (Cdk2)-dependent manner to the first two BRCT-repeats of TopBP1, that shares partial sequence conservation with yeast Dpb11 (Kumagai, *et al.* 2010) (Fig. 2.2D). Treslin and TopBP1 are independently loaded onto the chromatin (Kumagai, *et al.* 2010). A comparative analysis of the role of metazoan Treslin yeast Sld3 in DNA replication initiation is presented in section 2.*3.2.4*.

Furthermore, in mammalian cells, loading of Cdc45 onto the origins can occurs only after the assembly onto activated pre-RCs of other proteins apparently not conserved in yeast, DUE-B and GEMC1 (Fig. 2.2D). DUE-B is an ATP-dependent D-tyrosyl-tRNA (Tyr) deacylase necessary for DNA replication initiation, and its immunodepletion in *Xenopus* eggs blocks DNA replication and loading of Cdc45 and a fraction of TopBP1 (Chowdhury, *et al.* 2010). DUE-B can bind to TopBP1 and phosphorylation of its carboxy-terminal tail by Cdk2 enables transfer and tight binding of Cdc45 to the pre-RC (Chowdhury, *et al.* 2010).

GEMC1 is another necessary component of the pre-Initiation complex (preIC) conserved in metazoan cells. Its depletion in *Xenopus* prevents the onset of DNA replication owing to the impairment of Cdc45 loading onto chromatin (Balestrini, *et al.* 2010, Piergiovanni and Costanzo 2010).

GEMC1 from *Xenopus* binds Cdc45, TopBP1, and CyclinE-Cdk2, and promotes DNA replication initiation by mediating TopBP1- and Cdk2-dependent recruitment of Cdc45 onto replication origins (Balestrini, *et al.* 2010). GEMC1 from *Xenopus* and human share 54% of identity over all the amino acid sequence, and so they are likely to play a similar role in firing.

2.3.2.2 Chromatin loading of GINS

The next step of firing is the loading onto the activated origins of the GINS complex that consists of four proteins (Sld5, Pfsf1, Psf2 and Psf3) conserved in all eukaryotes, and constitutes together with Cd45 and Ddk-phosphorylated MCM2-7 the essential CMG initiation complex, described in section 2.*2.3*.

In yeast chromatin loading of GINS to origins is preceded by the S-Cdk activity-dependent formation of the so called pre-Loading Complex (pre-LC), not associated with chromatin and constituted by Sld2, Dpb11, GINS, and DNA polymerase epsilon (Muramatsu, *et al.* 2010). When Sld2 is phosphorylated by S-Cdk on Thr 84, it binds to Dpb11 and can assemble pre-LCs. Through its BRCT-repeats, Dpb11 binds also phospho-Sld3. Sld2 is phosphorylated on multiple-sites by S-Cdk through a distributive mechanism that involves repeated steps of phosphorylation and dissociation of the kinase and that represents a major regulatory mechanism of the initiation of DNA replication, as described in detail in section 2.*5*. Since Sld3 and GINS compete for the same binding sites on MCM2-7 and Cdc45 (Bruck and Kaplan 2011), when GINS within pre-LC is recruited to chromatin, by binding directly to MCM2-7 and Cdc45, it induces chromatin displacement of Sld3 and its Sld2 and Dpb11 partners (11-2-3 complex) (Muramatsu, *et al.* 2010).

In mammalian cells loading of GINS does not depend on the formation of a pre-LC, but is directly mediated by RecQ4 - a protein containing a yeast Sld2-like domain - that in turn is recruited onto the chromatin at the initiation sites from MCM10 (Xu, *et al.* 2009). A comparative analysis

of the role of RecQ4 and its yeast counterpart Sld2 in DNA replication initiation is reported in section *2.3.2.3*.

In human cells the assembly of a stable CMG complex requires the simultaneous presence of all the components of the CMG complex (Cdc45, Ddk-phosphorylated MCM2-7, Sld5, Psf1, 2 and 3), MCM10, RecQ4 and And1/Ctf4, but not of Dpb11 (Im, *et al.* 2009). A comparative analysis of the role of TopBP1 and Dpb11 in DNA replication initiation is reported in section *2.3.2.5*.

It has been demonstrated that human GINS complex binds to - and specifically stimulates - human DNA polymerase alpha (primase) (De Falco, *et al.* 2007) that is recruited to the origins by MCM10 and Ctf4/And1 (Gambus, *et al.* 2009, Zhu, *et al.* 2007).

Ctf4 is not essential for budding yeast viability, while its hortologs in fission yeast (*Schizosaccharomyces pombe*) (Tsutsui, *et al.* 2005) and in mammalian cells (Bermudez, *et al.* 2010, Chattopadhyay and Bielinsky 2007, Zhu, *et al.* 2007) are essential for the establishment of sister chromatid cohesion (paring) during S phase and initiation of the DNA replication by recruiting and likely activating DNA polymerase alpha. In particular in human cells Ctf4/And1, recruited onto chromatin by MCM10 (Zhu, *et al.* 2007), plays a key role in tethering the replicative helicase MCM2-7/RecQ4 complex and DNA polymerase alpha/primase to the replication fork (Bermudez, *et al.* 2010).

In conclusion, the comparative analysis of the proteins involved in DNA replication in budding yeast and in mammalian cells reported in section 2.3 and in Tables 2.2 and 2.3 clearly shows the general substantial degree of conservation during evolution of the DNA replication machinery. On the other hand one has to stress that the set of regulatory proteins in mammalian cells is more abundant than in yeast (for instance geminin, Hox proteins, MCM9, MCM8, GEMC1, DUE-B) and that three yeast proteins (Sld2, Sld3 and Dbp11) - that have pivotal role in yeast onset of DNA replication – are found to have counterparts (RecQ4, Treslin and

TopBP1/Cut5) with only partial sequence conservation and additional metazoan-specific domains and functions in higher eukaryotes. How these difference may impact on the molecular mechanism of the initiation of DNA replication in mammalian cells will be discussed in *2.3.2.3* to *2.3.2.5* after a careful analysis of the evolutionary conservation of these cited proteins.

2.3.2.3. Comparative analysis of yeast Sld2 and Xenopus and human RecQ4

In metazoans the well conserved multi-domain proteins RecQ4/RecQL4 belonging to the RecQ helicase family show variable sequence similarity to the yeast essential initiation factor Sld2 in their N-terminal region (Xu and Liu 2009). Fig. 2.3 shows the domain structure of the yeast, *Xenopus* and human proteins. Location and extension of the Sld2 homology domain reported in the figure has been identified through a pfam server (http://pfam.sanger.ac.uk/) analysis (Finn, *et al.* 2010). The split Sld2 homology regions shown are derived from a sequence alignment presented by (Matsuno, *et al.* 2006).

RecQ4 contains a helicase domain (residues 450-830 in hRecQ4) that together with the C-terminal region is not necessary for cell viability and is involved in DNA repair (Abe, *et al.* 2011). The separate helicase ATP-binding and C-terminal domains of the *Xenopus* and human proteins as identified through a pfam server analysis are reported in Fig. 2.3. RecQ4 contains a second ATP-dependent DNA unwinding activity also in the N-terminal region (Xu and Liu 2009). Residues 1-496 of human RecQL4 are essential and sufficient for cell viability (Abe, *et al.* 2011). Consistently, N-terminus deletion of mRecQ4 causes embryonic lethality in mice (Ichikawa, *et al.* 2002).

A

B

Proteins	Cdks-phosphorylation sites
S.c. Sld2	**T84**, **S100**, **S128**, S138, S164, S165, S166, **T168**, **S172**, S188, **S208**, **T241**, S254, S256, S413, T418
X.l. RecQ4	T62, S72, T101, T143, S151, S162, T168, S171, T291, S297, T306, S310, T321, S357, S377, S391,S473, S477, S486, S490, S518, T694, T714, S743, S780, S829, S913, S924, T945, T977, T1171, T1182, S1237, S1383, S1465
H.s. RecQ4	S61, T87, **S89**, **T93**, S135, **T139**, T145, S240, S251, S256, S271, S312, S313, S323, S326, S421, S436, T425, S436, T465, S523, S531, T580, T672, T690, T739, S862, T940, S1039, S1169

C

S.c. Sld2 Cdks-phosphorylation sites		
T84	VVEIGP**T**PQVYGK	Tak et al., 2006
S100	IFDMN.**S**PIKPIY	Masumoto et al., 2002
S128	TISNES**S**PRKTIL	Masumoto et al., 2002
T168	NASSTR**T**PTS**S**PCK	Masumoto et al., 2002
S172	TR**T**PTS**S**PCKNRN	Masumoto et al., 2002
S208	GYYGPN**S**PLKLDE	Masumoto et al., 2002
T241	YPSLQK**T**PSKDQA	Masumoto et al., 2002

H.s. RecQ4 Cdks-phosphorylation sites		
S89	LNRAATK**S**PQPTPGR	Xu et al., 2009
T93	ATKSPQP**T**PGRSRQG	Xu et al., 2009
T139	RASSKAS**T**PKPPGTG	Xu et al., 2009

Fig. 2.3 Comparative map of Sld2 from *Saccharomyces cerevisiae*, RecQ4 from *Xenopus leavis* and *Homo sapiens*. **A** domain architecture. **B** Cdk-phosphorylation sites of proteins reported in panel **A**. Sites were identified with the GPS2.1 software. Sites reported in bold have been experimentally validated, or by direct analysis with antibody directed against the phosphorylated peptides or mass spectrometry analysis, or through mutational analysis. **C** Sequence of phosphorylation sites

Notably, mutations of hRecQ4 correlate with human autosomal recessive disorders, such as Rothmund-Thomson (RTS), Baller-Gerold (BGS) and RAPADILINO syndromes (Liu 2010), characterized by chromosome fragility, developmental abnormalities and in some cases predisposition to cancer, such as osteosarcoma (Larizza, *et al.* 2010, Liu 2010, Maire, *et al.* 2009).

Even if recombinant human RecQ4 complements the depletion of *X*RecQ4 in *Xenopus* egg extracts (Sangrithi, *et al.* 2005), suggesting a conservation of the role of RecQ4 proteins through the evolution of metazoans, human and *Xenopus* RecQ4 proteins have different binding partners and regulate the onset of DNA replication through different mechanisms. While hRecQ4 plays a key role in chromatin loading of GINS complex, (see section *2.3.2.2*), XRecQ4 does not.

Notably while Sld2 promotes chromatin loading of GINS by binding to Dpb11, hRecQ4 does not bind to hTopBP1, (Xu, *et al.* 2009) and promotes chromatin loading of GINS by directly interacting with this complex (see section *2.3.2.2*). Differently, *X*RecQ4 can bind to Cut5, the *Xenopus* hortolog of Dpb11 (Matsuno, *et al.* 2006). *X*RecQ4 can be phosphorylated *in vitro* by S-Cdk (CycA/Cdk2), as Sld2, but such phosphorylation does not affect binding to Cut5 (Matsuno, *et al.* 2006), while in yeast Cdk phosphorylation of Sld2 stimulates binding to Dbp11 (Tak, *et al.* 2006). The interaction domains of the yeast and *Xenopus* protein are both located within the N-terminal region of the proteins.

Both interaction domains contain several *consensus* Cdk phosphorylation sites (Fig. 2.3B), but display no obvious sequence homology or common motif. Cdk activity necessary to transform pre-RC into active replication fork potentiates chromatin binding of *X*RecQ4 (Matsuno, *et al.* 2006), even if the loading of Cut5, GINS, Cdc45 and DNA Polymerase epsilon on origin does not seem to be dependent upon *X*RecQ4. Instead, *X*RecQ4 mediates chromatin loading of DNA Polymerase alpha, and

could be involved in DNA unwinding, since its helicase activity is required for chromatin association of *X*RPA, a protein that binds ss-DNA of unwound origins (Matsuno, *et al.* 2006).

Human RecQ4 interacts with pre-RC containing MCM2-7 by binding MCM10, and then mediates the recruitment of Sld5 belonging to the GINS complex (Xu, *et al.* 2009). In particular it has been demonstrated that MCM10 interacts with the first 200 residues of RecQ4 containing the Sld2 homology region, even if the Sld2-like domain alone is not sufficient for the interaction (Xu, *et al.* 2009) (Fig. 2.3). MCM10 may regulate the intrinsic ATP-dependent helicase activity of hRecQ4 to prevent unlicensed replication initiation and acts as a bridge connecting hRecQ4 and MCM7 (Thangavel, *et al.* 2010). No direct interaction between yeast MCM10 and Sld2 has been reported so far.

A very relevant question for the initiation of DNA replication is that of the involvement of S-Cdk dependent phosphorylation of firing proteins that in yeast has been shown to connect onset of DNA replication with cell cycle events (activation of S-Cdk, following degradation of the Cdk inhibitor Sic1). Distributive phosphorylation of yeast Sld2 by S-Cdk has been well documented (Salazar and Hofer 2009, Tak, *et al.* 2006, Zegerman and Diffley 2007) and its role in imposing coherent firing explored by simulation (Brummer, *et al.* 2010, Salazar, *et al.* 2010). No data are instead presently available on the potential regulation of hRecQ4 activity by Cdk-dependent phosphorylation, although the protein contains multiple potential phosphorylation sites for Cdks. The position of *consensus* Cdk phosphorylation sites for the yeast, *Xenopus* and human Sld2-like proteins are shown in Fig. 2.3A,B. Fig. 2.3C reports the sequence around the phosphorylatable serine or threonine for those sites whose phosphorylation has been experimentally demonstrated or whose role has been probed by site-directed mutagenesis. These residues are shown by arrows or bold character in panels A and B, respectively. Taken together the findings reported so far in the literature

do not support the notion that multisite distributive phosphorylation of RecQ4 by S-Cdk, having a regulatory role for onset of DNA replication, occurs in metazoans, despite the fact that consensus Cdk phosphorylation sites can be phosphorylated *in vitro* by incubating purified hRecQ4 with human cell extracts (Xu, *et al.* 2009).

2.3.2.4. Comparative analysis of yeast Sld3 and Xenopus and human treslin

Sld3 is a 668 residue protein. Its functional homolog in metazoan is Treslin that is almost three times as large. Only recently a sequence homology between Sld3 and Treslin has been ascertained by means of an alignment of animal and plant Treslin proteins against non-redundant protein sequence database followed by remote protein homology detection using the HHpred method (Sanchez-Pulido, *et al.* 2010) (Fig. 2.4, shaded area). The Dbp11 binding site is located in the C-terminus of the protein (Fig. 2.4, yellow box). Recently a novel Sld3 partner, Sld7, has been identified (Tanaka, *et al.* 2011). This protein can bind to the non-essential N-terminal region of Sld3, likely playing a role in stabilizing the essential initiation factor. Sld7 can reduce the affinity of Sld3 for Cdc45, and affects the association of GINS with replication origins. Hortologs of Sld7 are only found in a limited range of yeast closely related to *Saccharomyces cerevisiae* (Tanaka, *et al.* 2011).

The position of *consensus* Cdk phosphorylation sites for the yeast, *Xenopus* and human proteins are shown in Fig. 2.4 A,B. Fig. 2.4C reports the sequence around the phosphorylatable serine or threonine for those sites whose phosphorylation has been experimentally demonstrated or whose role has been probed by site-directed mutagenesis. These residues are shown by arrows or bold character in panels A and B, respectively. S-Cdk-phosphorylation (Fig. 2.4 B and C) makes Sld3 and Treslin able to bind Cdc45 and the BRCT-repeats of Dpb11/TopBP1 (Kumagai, *et al.* 2010, Tanaka, *et al.* 2007, Zegerman and Diffley 2007). However while binding of phospho-Sld3 to Dpb11 is a key event for chromatin recruitment of GINS and determines, together with the

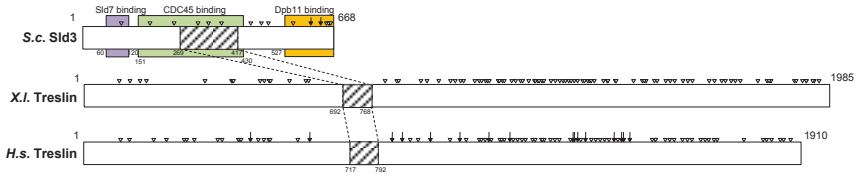
binding of phosphoSld2 to Dpb11, the coherence of origin firing in yeast, binding of Treslin to TopBP1 plays a different role, since TopBP1 is not required for loading of GINS, as described in detail in 2.3.2.5..

2.3.2.5. Comparative analysis of yeast Dpb11, Xenopus Cut5 and human TopBP1

Dpb11 and its homologues in other species (TopBP1 - DNA topoisomerase 2-binding protein 1 - in mammals, Cut5 in *Xenopus*) (Garcia, *et al.* 2005) contain multiple pairs of BRCT [BRCA1 C-terminal] domains, initially identified in the C-terminal region of the *BRCA1* (breast cancer susceptibility gene 1) gene product. BRCT domains recognize phosphopeptide targets (Caldecott 2003, Yu, *et al.* 2003) and are commonly found in proteins involved in DNA repair or cell cycle checkpoint (Bork, *et al.* 1997, Callebaut and Mornon 1997, Zhang, *et al.* 1998). Accordingly all protein with partial sequence conservation to Dpb11 play a key role in DNA damage checkpoint response (Araki, *et al.* 1995, Garcia, *et al.* 2005, Greer, *et al.* 2003, Kumagai, *et al.* 2006, Navadgi-Patil and Burgers 2008, Yamane, *et al.* 2003, Yamane, *et al.* 2002).

Fig. 2.5 shows the domain structure of the *S. cerevisiae*, *Xenopus* and human proteins. Location and extension of the BRCT homology domains reported as pink boxes in the figure has been identified through the Pfam server (http://pfam.sanger.ac.uk/) (Finn, *et al.* 2010). BRCT domains reported as fucsia boxes in the figure have been identified through different alignment tools, such as BLAST, as regions similar to regions 1 and 2 of Rad4 protein (Araki, *et al.* 1995, Yamane, *et al.* 1997).

A



B

Proteins	Cdks-phosphorylation sites
***S.c.* Sld3**	T93, T174, S233, T323, T361, S439, S467, S485, T487, T553, T609, **T600**, **S622**, T638, S639, T642, S643, T644, T646
***X.l.* Treslin**	S93, S122, S147, S165, T324, S392, S395, S473, S483, T549, S589, S596, T801, S802, T831, S838, S899, S914, T944, S976, S989, S1003, S1008, T1041, S1055, T1058, T1075, T1079, T1083, S1102, T1105, T1108, T1111, T1115, T1121, S1141, S1149, T1165, S1171, T1186, T1200, T1204, T1206, S1208, T1212, T1223, S1227, T1247, T1259, T1264, T1276, T1281, T1293, T1298, T1310, T1315, T1327, T1332, T1344, T1349, S1353, T1361, T1361, S1363, T1383, T1397, S1400, T1416, S1420, T1465, S1487,T1493, S1498, T1583, T1593, T1601, T1615,T1624, T1669, S1681, S1701, T1711, T1726, S1736, T1747, S1749, S1804, S1828, S1831, T1896, S1899, S1917, T1925, S1928, T1947, S1963, T1985
***H.s.* Treslin**	T98, S120, T184, S222, S268, S273, S292, S354, S366, S410, S413, **S441**, S463, S472, T486, S489, T566, **S599**, **S820**, **S838**, S865, T886, **S923**, T969, **S1001**, S1013, S1051, S1057, S1064, **S1078**, T1105, S1109, S1117, T1120, T1127, T1130, **T1134**, S1141, T1146, S1162, S1167, T1174, T1192, T1195, T1201, S1215, S1218, S1219, S1219, S1222, S1224, T1230, S1231, T1242, T1250, T1260, T1291, T1301, **S1303**, **T1307**, **S1314**, S1323, **S1334**, S1346, S1348, T1360, S1365, T1372, S1379, S1389, S1403, T1405, **S1413**, **S1430**, **S1433**, S1452, **S1453**, S1509, S1512, S1517, S1546, S1555, S1568, S1582, S1590, S1606, S1615, T1620, S1623, T1633, T1648, S1655, T1656, S1658, T1668, S1670, T1678, S1750, S1806, T1817, T1828, S1850, S1850, T1858, S1861, S1881

C

***S.c.* Sld3 Cdks-phosphorylation sites**		
T600	ENVQVQAT**T**PAVKKRT	Tanaka et al., 2007; Zegerman and Diffley, 2007
S622	QLQSIIE**S**PLNFKDD	Tanaka et al., 2007; Zegerman and Diffley, 2007

***H.s.* Treslin Cdks-phosphorylation sites**		
S441	LQTAVAD**S**PRDTASL	Olsen et al., 2010; Cantin et al., 2008
S599	QKLHPDG**S**PDVAGEK	Olsen et al., 2010; Olsen et al., 2006; Dephoure et al., 2008
S820	SMTQENK**S**PLLSVPF	Olsen et al., 2010
S838	RRSVSG**S**PESDEL	Dephoure et al., 2008; Olsen et al., 2010
S923	LFNQELL**S**PSKRSLK	Olsen et al., 2010; Brill et al., 2009
S1001	DIGVVEE**S**PEKGDEI	Mayya et al. , 2009; Olsen et al., 2010
S1078	GAMSEMI**S**PSEKGSA	Olsen et al., 2010
T1134	TPQTPLY**T**PERLQKS	Cantin et al. , 2008
S1303	GNSTVTS**S**PPVTPKK	Dephoure et al., 2008; Gauci et al. , 2009; Mayya et al. , 2009
T1307	VTSSPPV**T**PKKLFTS	Dephoure et al., 2008; Gauci et al. , 2009; Mayya et al., 2009; Olsen et al., 2010
S1314	TPKKLFT**S**PLCDVSK	Brill et al., 2009
S1334	KSKIECP**S**PGELDQK	Brill et al., 2009
S1413	PGVGTAD**S**PAAPTDS	Dephoure et al., 2008
S1430	DQKGLSL**S**PQSPPER	Olsen et al., 2010; Bennetzen et al., 2010
S1433	GLSLSPQ**S**PPERRGY	Olsen et al., 2010; Bennetzen et al., 2010
S1453	RSDWHAS**S**PLLITSD	Chen et al., 2009

Fig. 2.4 Comparative map of Sld3 from *Saccharomyces cerevisiae*, Treslin from *Xenopus leavis* and *Homo sapiens*. **A** domain architecture. **B** Cdk-phosphorylation sites of proteins reported in panel **A**. Sites were identified with the GPS2.1 software. Sites reported in bold have been experimentally validated, or by direct analysis with antibody directed against the phosphorylated peptides or mass spectrometry analysis, or through mutational analysis. **C** Sequence of phosphorylation sites reported in bold in panel **B**.

A

B

Proteins	Cdks-phosphorylation sites
***S.c.* Dpb11**	T267, S308, S423, S442, S443, S553, T558
***X.l.* Cut5**	S16, S20, T42, S142, S287, T288, T295, T297, T450, S454, T623, T632, T771, T796, S801, T814, T840, S852, T853, S856, T874, S880, S1095, T1098, T1142, S1159, T1179, S1202, T1206, S1215, T1217, T1223, T1225, S1230, S1253, S1422
***H.s.* TopBP1**	**S142**, S296, S297, T298, S301, S312, S455, S460, T506, T635, T644, T724, **T779**, **S805**, T822, **T848**, **S860**, **T861**, **S864**, **S888**, S987, S1016, T1092, S1101, S1104, T1105, S1108, S1183, **T1221**, T1231, S1477

C

	***H.s.* TopBP1 Cdks-phosphorylation sites**	
S142	VYRDLNV**S**VTHLIAG	Olsen et al., 2010
T779	THRKTVV**T**PLDMNRF	Dephoure et al. , 2008
S805	HARQVAA**S**PAVGQPL	Beausoleil et al., 2006; Wang et al., 2008
T848	DALAALE**T**PGRPSQQ	Dephoure et al. , 2008
S860	SQQKRKP**S**TPLSEVI	Dephoure et al. , 2008
T861	QQKRKPS**T**PLSEVIV	Dephoure et al. , 2008
S864	RKPSTPL**S**EVIVKNL	Dephoure et al. , 2008
S888	NAVALSA**S**PQLKEAQ	Beausoleil et al., 2006; Wang et al., 2008; Dephoure et al. , 2008; Mayya et al., 2009
T1221	PISEELE**T**PIKDSHL	Olsen et al., 2010

Fig. 2.5 Comparative map of Dpp11 from *Saccharomyces cerevisiae*, Cut5 from *Xenopus leavis* and TopBP1 from *Homo sapiens*. **A** domain architecture. **B** Cdk-phosphorylation sites of proteins reported in panel A. Sites were identified with the GPS2.1 software. Sites reported in bold have been experimentally validated, or by direct analysis with antibody directed against the phosphorylated peptides or mass spectrometry analysis, or through mutational analysis. **C** Sequence of phosphorylation sites reported in bold in panel **B**.

The position of *consensus* Cdk phosphorylation sites for the yeast, *Xenopus* and human proteins are shown in Fig. 2.5 A,B. Fig. 2.5C reports the sequence around the phosphorylatable serine or threonine for those sites whose phosphorylation has been experimentally demonstrated or whose role has been probed by site-directed mutagenesis. These residues are shown by arrows or bold character in panels A and B, respectively.

The *S. cerevisiae*, *Xenopus* and human proteins do not share all binding partners. Dpb11 binds through distinct basis pairs of BRCT repeats with Cdk-phosphorylated Sld2 and Sld3 (Masumoto, *et al.* 2002, Tanaka, *et al.* 2007, Zegerman and Diffley 2007).

Both human TopBP1 and *Xenopus* Cut5 are able to bind in a Cdk-dependent manner to initiation factors such as Treslin (Kumagai, *et al.* 2010), and other metazoan specific initiation factors such as DUE-B (Chowdhury, *et al.* 2010) and GEMC1 (Balestrini, *et al.* 2010). Cut5 also binds *X*RecQ4 (Matsuno, *et al.* 2006), while hTopBP1 is not able to bind to hRecQ4 (Matsuno, *et al.* 2006). Furthermore differently from Dpb11, both human TopBP1 and *Xenopus* Cut5 directly bind to Cdc45 (Schmidt, *et al.* 2008). Fig. 2.5 shows the known binding sites for the above described interactors: for Cdk-phosphorylated Sld2 and Sld3 onto Dpb11, for Cdc45 and Cdk-phosphorylated Treslin onto hTopB1. Notably, yeast Dpp11 and human TopBP1 bind Sld3 and Treslin through a very similar region, involving the first two NH_2-terminal BRCT domains (Fig. 2.5, yellow boxes).

Dbp11, Cut5 and TopBP1 are all essential in DNA replication initiation, but their role seems to be quite different. In yeast, Dpb11 is necessary for chromatin recruitment of GINS complex (Muramatsu, *et al.* 2010) by acting as a bridge between phosphorylated Sld3, bound to Cdc45 and loaded onto chromatin, and phosphorylated Sld2 bound to GINS, promoting the recruitment of pre-LC at the initiation sites (Muramatsu,

et al. 2010) and the assembly, with additional factors, of the replisome progression complex (RPC) [for review see (Labib and Gambus 2007, MacNeill 2010)]. Fission yeast (Noguchi, *et al.* 2002) and *Xenopus leavis* Cut5 (Hashimoto and Takisawa 2003, Van Hatten, *et al.* 2002), are required for chromatin loading of Cdc45, while the role of hTopBP1 is controversial. Depletion of Cut5 in *Xenopus* eggs suppresses chromatin loading of Cdc45 and DNA polymerases (Hashimoto and Takisawa 2003). It has also been suggested that Cut5-dependent Cdc45 chromatin loading requires S-Cdk activity (Hashimoto and Takisawa 2003).

Similarly to Cut5, hTopBP1 is able to interact with hCdc45 both *in vitro* and in Hela-S3 cells exclusively at G1/S boundary, through its first, second and sixth BRCT-domains and the overexpression of such six BRCT-domain affected loading of Cdc45 onto the chromatin (Schmidt, *et al.* 2008) (Fig. 2.5). While some authors suggest a role for hTopBP1 in chromatin loading of Cdc45 (Schmidt, *et al.* 2008, Van Hatten, *et al.* 2002), other ones demonstrate that the depletion of TopBP1 in Hela cells does not affect assembly of the CMG complex (Im, *et al.* 2009), suggesting that the essential role played by hTopBP1 in DNA replication (Jeon, *et al.* 2007, Makiniemi, *et al.* 2001) likely takes place after that CMG complex is assembled, possibly by recruiting DNA polymerase epsilon or stabilizing its binding at the initiation sites (Makiniemi, *et al.* 2001).

2.4 Role of Cdks in controlling the onset of DNA replication in yeast and in mammalian cells

Regulation of DNA replication in eukaryotes is tight, in order to avoid that the genome is replicated unequally (i.e., some parts are under-replicated and/or others are over-replicated) or more than once during each cell cycle. Regulation is effected mainly through the action of Cdks, their cyclin regulators and Cdk Inhibitors (Ckis).

In budding yeast, the critical cyclin-dependent protein kinase involved in cell cycle regulation is Cdk1 (Cdc28) whose function is regulated by

interaction with G1 cyclins (Cln1 and Cln2,3), S cyclins (Clb5,6), G2 cyclins (Clb3,4) and mitotic cyclins (Clb1,2). Since Clb1-4 cyclins are not involved in regulation of the G1/S transition or the initiation of DNA replication, they will not be further considered in this thesis.

At least one among *CLN1*, *CLN2* and *CLN3*-encoding genes is required for viability of otherwise wild type yeast cells (Levine, *et al.* 1996), indicating that the encoded proteins have a partially redundant function. In wild type cells, however, Cln3 is required for transcriptional activation of the *CLN1* and *CLN2* gene [(Stuart and Wittenberg 1995, Tyers, *et al.* 1993), see also 2.4.1.] and Cln1 and 2 have distinct subcellular localization and functional roles (Miller and Cross 2000). Several evidences indicate that Clb5 is the major S phase cyclin in cycling cells (Schwob and Nasmyth 1993).

In mammalian cells several Cdks involved in the mitotic cycle have been described: Cdk4/6 (whose function is regulated by interaction with Cyclin D1,3), Cdk2 (whose function is regulated by interaction with Cyclin E and Cyclin A) and Cdk1 (whose function is regulated by interaction with Cyclin A and Cyclin B). Since cyclin B is not involved in regulation of the G1/S transition or of the initiation of DNA replication, they will not be further considered in this thesis (Malumbres and Barbacid 2009, Malumbres and Barbacid 2005, Sanchez and Dynlacht 2005).

Cdk/cyclin complexes affect initiation of DNA replication, both by promoting transcription of genes encoding proteins involved in the G1/S transition and by affecting activity, interaction, subcellular localization or stability of proteins involved in promotion of licensing or firing or in inhibition of re-licensing. The major phosphorylation events that promote DNA replication and prevent re-licensing are summarized in Fig.s 2.6 and 2.7. Besides Cdks, also the Ddk protein kinase is involved in phosphorylation events promoting DNA replication.

2.4.1 Role of Cdks in promoting G1/S-specific transcription

In budding yeast, active Cdk1-Cln3 activates the transcription factors SBF and MBF, whose basal activity is kept low by Whi5 (Costanzo, *et al.* 2004, de Bruin, *et al*. 2004). Cdk1-Cln3 phosphorylates Whi5, promoting its dissociation from SBF and, possibly, its nuclear export. After Cln3-Cdk1 phosphorylation, SBF and MBF start to activate transcription of a number of genes essential for G1/S transition and DNA replication, including *CLN1,2*, *CLB5,6*, *SLD2*, *CDC45*, all subunits of heterotrimeric Replication Protein A (RPA), a highly conserved single-stranded DNA binding protein involved in DNA replication, repair, and recombination, catalytic subunit of DNA Polymerase alpha/Primase and DNA Polymerase delta, and the gene encoding the processivity factor PCNA (Ferrezuelo, *et al.* 2010). The Cln1,2-Cdk1 complex then further phosphorylates Whi5 (Charvin, *et al.* 2010), allowing full transcriptional derepression. Cytoplasmic Clb5 and Clb6 bind to Cdk1, and the newly formed complex Cdk1-Clb5,6 is inhibited by the cytoplasmic Cki Sic1. Therefore, active Cdk1-Clb5,6 complexes are released only after Sic1 degradation that is primed by multisite phosphorylation by Cdk1-Cln1,2 (Nash, *et al.* 2001). These regulatory events are outlined in Fig. 2.6, top panel.

In mammalian cells the transition of cells through the early G1 stage of the cell cycle is coordinated by the activities of Cdk4 and Cdk6 in complex with the mitogen-dependent expressed D-type cyclins (Matsushime, *et al.* 1992, Meyerson and Harlow 1994). These complexes perform initial phosphorylation of Retinoblastoma protein (Rb), that destabilizes the complex Rb-E2F so to initially release the transcription factor E2F which bring to a partial stimulation of the expression of E2F target genes involved in G1 progression, including CycE (Bracken, *et al.* 2004). CycD-Cdk4/6 complexes act also as a sink for Ckis of the Cip/Kip family, potent inhibitors of Cdk2 (Sherr and Roberts 2004). As the CycE-Cdk2 complex become active in late G1, it promotes phosphorylation of key target proteins, including Rb which in its hyperphosphorylated form

fully releases E2F transcription factor and allows the transcription of several genes essential for G1/S transition, including CycA, and DNA replication, including ORC1, Cdc6, Cdt1, MCM2-7, Cdc45, PCNA, POLA (Bracken, *et al.* 2004) and MCM9 (Yoshida 2005). These regulatory events are outlined in Fig. 2.6, bottom panel. In conclusion, as shown in Fig. 2.6, there is a general similarity between yeast and mammalian cells in the regulatory steps controlling ttranscription of genes involved in the G1 to S transition.

2.4.2 Role of Cdks in promoting DNA replication initiation

In both yeast and mammalian cells Cdc6 is stabilized by interaction with a cyclin-cdk complex (Fig. 2.7B): binding of Clb2-Cdk1 protects yeast Cdc6 from degradation at the G2/M transition of the previous cycle (Mimura, *et al.* 2004), while CycE-Cdk2 phosphorylates mammalian Cdc6, preventing its degradation through APC/C-dependent proteolysis (Mailand and Diffley 2005). Similarly, in CycD-dependent transformed cells, CycD-Cdk4,6 activity, by inducing Cul4 repression, promotes Cdt1 stabilization (Aggarwal, *et al.* 2010). During late M and early G1 phase Cdk1 phosphorylates mammalian MCM3 promoting its incorporation into the heterohexamer MCM2-7 and thereby the assembly of the helicase complex [Fig. 2.7D; (Lin, *et al.* 2008)].

After the MCM2-7 complex has been loaded at origins in an inactive form during G1 phase, it is activated at G1/S transition and throughout S-phase by Cdk- and Ddk-phosphorylation [Fig. 2.7D; (Bochman and Schwacha 2009, Labib 2010, Remus and Diffley 2009)]. The S-phase kinase Ddk is a conserved serine/threonine kinase composed of the catalytic subunit Cdc7 and the regulatory subunit Dbf4 (Jares, *et al.* 2000, Masai and Arai 2002, Sclafani 2000). MCM2, 4 and 6 in all eukaryotes possess amino-terminal unstructured extensions of about 200 residues, named N-terminal serine/threonine-rich domain (NSD), that likely represent the major substrate of Ddk, besides being substrate of other kinases including Cdk, as reviewed in (Labib 2010). It has been

demonstrated that Cdk-phosphorylation on specific residues of this region helps Cdc7 to phosphorylate other residues in the same region (Devault, *et al.* 2008, Masai and Arai 2000, Montagnoli, *et al.* 2006).

In yeast at the G1/S transition Ddk is recruited to origins likely through interaction with MCM4, thereby hyperphosphorylating the N-terminal region of several MCM subunits, including MCM4 (Sheu and Stillman 2006), and relieving the inhibitory action exerted by this region of MCM4 (Sheu and Stillman 2010). Such phosphorylation promotes also the stable association between MCM2-7 and the essential initiation protein Cdc45, that can more loosely associate with pre-RCs in the absence of Cdk phosphorylation (Francis, *et al.* 2009, Masai, *et al.* 2006).

Loading of the GINS complex and DNA polymerases requires prior activation - through multiple S-Cdk-mediated phosphorylations – of Sld2 and Sld3 [Fig. 2.7E; (Araki 2010a, Tanaka, *et al.* 2007, Zegerman and Diffley 2007)]. Distributive phosphorylation of Sld2 and Sld3 is crucial in originating switch-like, coherent activation of replication origins (Brummer, *et al.* 2010) and will be discussed in more detail in Section *2.5*.

In mammalian cells the regulation of MCM2-7 by Ddk phosphorylation is not yet well understood. It has been demonstrated that chromatin-bound hMCM2 is phosphorylated by Cdc7/Dbf4 during G1/S, and this event coincides with the initiation of DNA replication (Tsuji, *et al.* 2006). In particular in human cells Cdc7 (and Cdk) is required for formation of the CMG helicase complex (Im, *et al.* 2009). However, recently it has also been reported that phosphorylations of MCM2 by Ddk promote loading of MCM helicase on the chromatin during re-entry into the cell cycle from quiescence (Chuang, *et al.* 2009).

Also in mammalian cells the activity of S-Cdks orchestrates the assembly of the preICs, allowing chromatin binding of essential initiation factor, such as Cdc45. In particular, chromatin loading of Cdc45 is mediated by the S-Cdk dependent activity of the metazoan specific initiation factors DUE-B (Chowdhury, *et al.* 2010), Treslin (Kumagai, *et al.* 2010), and

GEMC1 (Balestrini, *et al.* 2010), as shown in Fig. 2.7E. GEMC1 from *Xenopus* can bind Cdc45, TopBP1, and CyclinE-Cdk2, and promotes DNA replication initiation by mediating TopBP1- and Cdk2-dependent recruitment of Cdc45 onto replication origins (Balestrini, *et al.* 2010).

In mammalian cells Cdk activity is also required for promoting the additional metazoan-specific helicase activity exerted by RecQ4 [Fig. 2.7E; (Xu, *et al.* 2009)]. In particular Cdk phosphorylation of RecQ4 releases the inhibitory activity exerted by MCM10 on the helicase domain of RecQ4 (Xu, *et al.* 2009). Also the initiation factor TopBP1 is a substrate of Cdks (Fig. 2.5), but the role of its phosphorylation has yet to be elucidated.

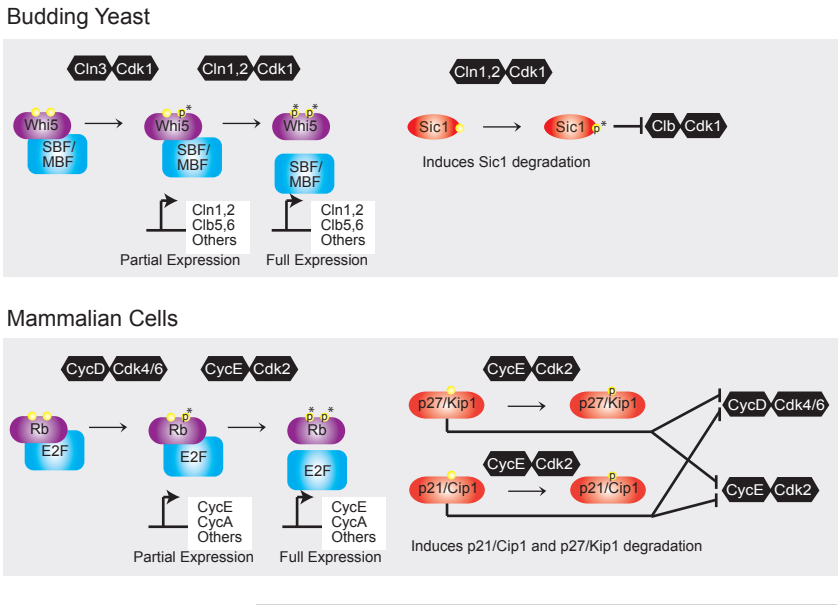


Fig. 2.6 Schematic map of the known Cyclin-Cdk mediated phosphorylation events regulating positively the expression and activity of other Cyclin-Cdk complexes promoting the expression of molecular elements involved in cell cycle progression and DNA replication, both in yeast and in mammalian cells. Cdk binding partners not known are not reported. The legend put in the upper part of the figure is referred to regulatory events of both yeast and mammalian cells.

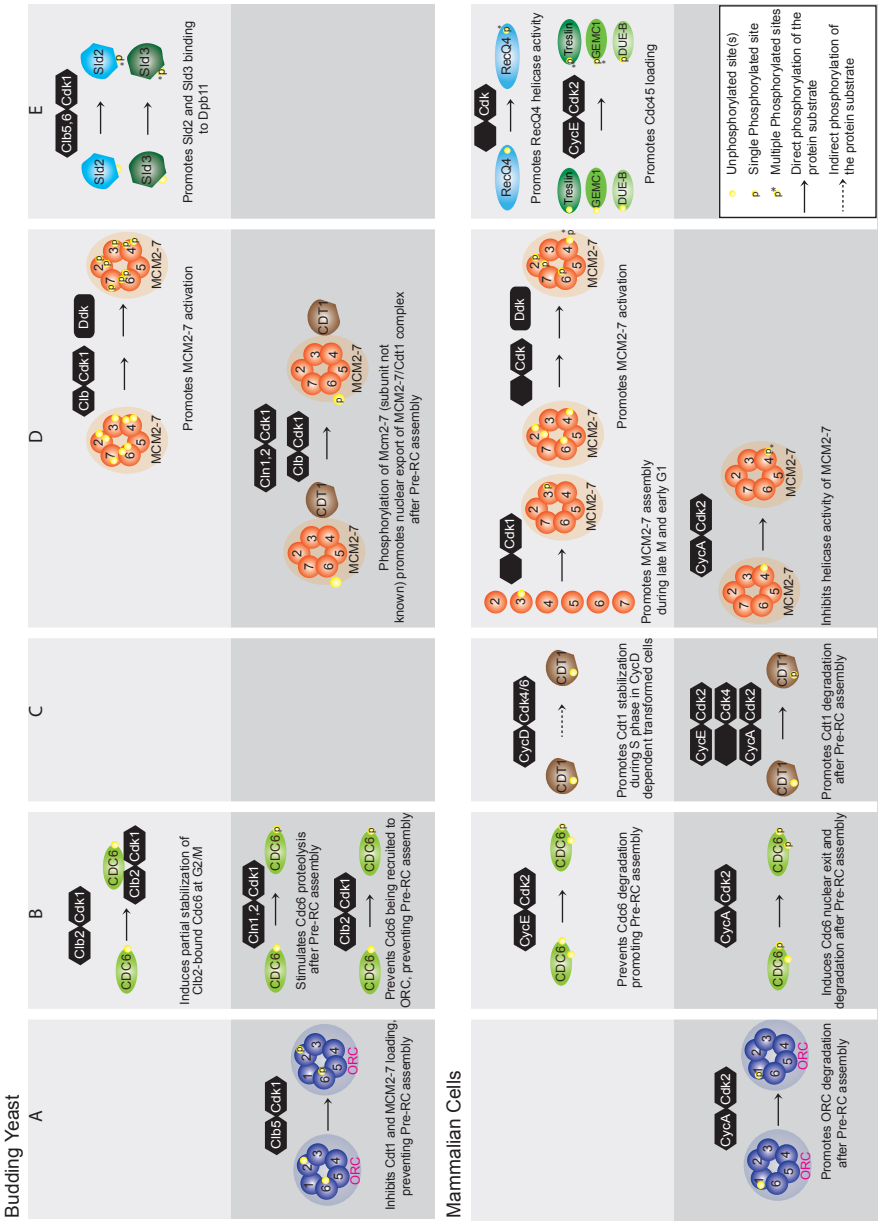


Fig. 2.7 Schematic map of the known Cyclin-Cdk mediated phosphorylation events regulating both positively and negatively the level and the activity of licensing and firing factors involved in the initiation of DNA replication in yeast and in mammalian cells. Columns shows the regulation of the licensing factors ORC (A), Cdc6 (B), Cdt1 (C), MCM2-7 (D) and the firing factors (E). Regulatory events shown in the light and dark grey boxes are involved in promoting DNA replication and in inhibiting re-replication respectively. See text and Fig. 2.1 and 2.2 for the detailed role of each phosphorylated protein in the licensing or firing process respectively. Cdk binding partners not known are not reported. The legend put in the lower part of the figure is referred to regulatory events of both yeast and mammalian cells.

2.4.3 Role of Cdks in inhibiting re-licensing

2.4.3.1 Regulation of ORC activity by Cdk-dependent phosphorylation

In yeast, binding of the Orc6 subunit to S-phase cyclin Clb5, and its phosphorylation by the Clb5-Cdc28 complex prevents its binding to other licensing factors [Fig. 2.7A; (Chen and Bell 2011)]. Clb5-Orc6 interaction requires replication initiation, and is maintained throughout the remainder of S phase and into M phase. Eliminating the Clb5-Orc6 interaction does not affect initiation of replication but sensitizes cells to over-replication (Wilmes, *et al.* 2004).

As mammalian cells enter S-phase, cyclin A binds Orc1 and recruits Cdk2. The CycA-Cdk2 complex phosphorylates Orc1 causing its export from the nucleus and preventing assembly of pre-RCs during S-phase (Fig. 2.7A). Such mechanism contributes to avoid re-licensing origins during S phase. Furthermore in some cells Orc1 bound to CycA can be selectively degraded through the 26S proteasome by SCF^{Skp2}, a ubiquitin-dependent mechanism (DePamphilis 2005, Mendez, *et al.* 2002, Tatsumi, *et al.* 2003).

2.4.3.2 Regulation of Cdc6 activity by Cdk-dependent phosphorylation

In budding yeast cellular level of Cdc6 is cell cycle-regulated: G1-Cdk (Cln1-Cln2/Cdc28)-dependent phosphorylation of Cdc6 promotes its degradation mediated by Skp1-Cullin-F-boxprotein (SCF/Cdc4) complex, and this phosphorylation occurs after pre-RCs assembly [Fig. 2.7B; (Drury, *et al.* 2000)]. The direct interaction between the Cdk-phosphorylated NH2-terminal domain of Cdc6 and Clb2 prevents it being recruited to ORC at G2/M, and also results in a partial stabilization of Cdc6 during mitosis (Elsasser, *et al.* 1996, Mimura, *et al.* 2004). During mitotic exit Cdc6 cooperates with Sic1 to directly inactivate Cdk1 (Calzada, *et al.* 2001).

In mammalian cells phosphorylation of Cdc6 by Cyclin A/Cdk2 causes its export from the nuclei after initiation of DNA replication [Fig. 2.7B; (Jiang, *et al.* 1999, Petersen, *et al.* 1999)] and Cdc6 remains mainly

cytoplasmic until it is targeted for proteolysis by the anaphase promoting complex/cyclosome (APC/CDH1) in early G1 (Petersen, *et al.* 2000). The phosphorylation of Cdc6 by Cyclin A/CDK2 is a negative regulatory event likely implicated in preventing re-replication during S and G2 phases (Petersen, *et al.* 1999). The ectopic expression of Cyclin A, but not of Cyclin E, leads to rapid relocalization of Cdc6 from the nucleus to the cytoplasm. Besides being required in the initiation of DNA replication, Cdc6 is involved in the Chk1-mediated checkpoint pathway coupling the completion of the S-phase with the onset of mitosis and ensuring that all DNA is replicated before mitotic entry (Clay-Farrace, *et al.* 2003, Murakami, *et al.* 2002, Oehlmann, *et al.* 2004).

2.4.3.3 Regulation of Cdt1 activity and its modulators by Cdk-dependent phosphorylation

In yeast Cdt1 forms a stable complex with MCM2-7, whose nuclear localization is cell cycle regulated (Tanaka and Diffley 2002). Such complex accumulates in the nucleus during G1-phase and is exported from the nucleus in a Cdk-dependent manner upon phosphorylation mediated by Cdk1 associated to either Cln1,2 or Clb of an unknown subunit of MCM2-7 complex after Pre-Rc assembly (Fig. 2.7D).

In mammalian cells, Cdt1 cellular levels are subjected to cell cycle-dependent regulation. During S and G2 phase Cdt1 is degraded upon ubiquitination by three different ubiquitin ligases: (i) SCF^{Skp2}, targeting Cdt1 proteins phosphorylated by CycA-Cdk2 (Li, *et al.* 2003, Liu, *et al.* 2004, Sugimoto, *et al.* 2004, Takeda, *et al.* 2005) or CycE-Cdk2 or Cdk4 (Liu, *et al.* 2004) (Fig. 2.7C); (ii) the cullin-dependent ubiquitin ligase Cul4–DDB1–CDT2 (also known as DTl), requiring the presence of the processivity factor for replicative DNA polymerase, PCNA (Arias and Walter 2006, Arias and Walter 2005, Lovejoy, *et al.* 2006, Senga, *et al.* 2006) and (iii) Anaphase promoting complex APC ubiquitin ligase (Sugimoto, *et al.* 2008). The first two described ubiquitylation mechanisms of Cdt1

couples the degradation of Cdt1 to ongoing DNA synthesis, limiting re-replication events.

In mammalian cells CycA-Cdk2 phosphorylates MCM4 by inhibiting the helicase activity of the MCM2-7 complex (Ishimi and Komamura-Kohno 2001) (Fig. 2.7D).

In mammalian cells Geminin – whose activity is cell cycle-regulated - plays a key role in timely inhibition of Cdt1. Geminin is active during S and G2 and early M phases, and it is inactivated by APC-dependent ubiquitination during late M and G1 phase.

In conclusion, it is clear that while an acceptable general consensus is available on the organization of the molecular events involved in the initiation of DNA replication in budding yeast, the situation is much more complex and much less defined for mammalian cells, although several indications appear to suggest some sort of evolutionary conservation in the logic of control of S-phase initiation. Given that a better understanding of complex biological processes may be achieved by constructing mathematical models and by their simulation analysis (Alberghina, *et al.* 2009, Meng, *et al.* 2004, Resat, *et al.* 2011), in the following we are going to utilize a mathematical model of the initiation of DNA replication in budding yeast, developed by our laboratory in a collaborative project (Brummer, *et al.* 2010), as a framework to answer the more critical questions on the organization of molecular events involved in the onset of S-phase in mammalian cells.

2.5 Regulatory events in the initiation of DNA replication in budding yeast and in mammalian cells

2.5.1 A kinetic model of the onset of DNA replication in budding yeast

The model of DNA replication initiation presented by Brummer *et al.* (2010) describes the events of licensing and firing as a sequence of molecular interactions triggered by the formation - dependent on multi-site protein phosphorylation - of the activator complex containing Dpb11

and Sld2. The major insights offered by the computational analysis of the model is that distributive multisite phosphorylation of S-Cdk targets Sld2 and Sld3 can generate both a robust time delay and a coherent, switch-like, activation of replication origins. In addition a time delay between the activation of S-Cdk and the initiation of DNA replication is needed to prevent DNA re-replication. Although the sequence of molecular interactions considered by the model is slightly different from that reported in Fig. 2.3 (phosphorylated Sld2 in the model binds to Dpb11 and phosphorylated Sld3 in a sequence different from that reported in Fig. 2.3), the crucial regulatory event is in both cases the multisite phosphorylation of Sld2 that allows coherent binding of GINS and DNA polymerases.

The model has been shown to link the dynamics of the network to its performance in terms of rate and coherence of origin activation events, number of activated origins, the resulting distribution of replicon sizes and robustness against DNA re-replication. Computational analysis showed that optimized kinetic parameters are required for the biological functionality of the network, and such parameters indeed account for the measured kinetics of replication initiation (Brümmer *et al.* 2010).

2.5.2 Is there a role for multisite protein phosphorylation in the control of DNA replication initiation in mammalian cells?

The role of distributive multisite protein phosphorylation in controlling DNA replication initiation in budding yeast has been shown to be twofold: to assure coherence in the activation of DNA replication origins, in order to allow fast complete replication of DNA (see Table 2.1) and to provide robustness against re-replication (Salazar, *et al.* 2010). The critical target molecules for these regulatory phosphorylations are Sic1, Sld2 and Sld3.

While Sic1 degradation requires multisite phosphorylation by Cln1,2-Cdk1, degradation of the mammalian Ckis $p21^{Cip1}$ and $p27^{Kip1}$ is regulated by single phosphorylation events (Lu and Hunter 2010). On

the other hand, in mammalian cells more kinases than in yeast are involved in regulating stability of the Ckis (Sherr and Roberts 1999).

Together with metazoan specific proteins such as GEMC1 and DUE-B, RecQ4 and Treslin - that share sequence conservation with yeast Sld2 and Sld3, respectively - are potential Cdk substrates in mammalian cells. Both RecQ4 and Treslin are much larger proteins than yeast Sld2 and Sld3 and their role in initiation of DNA replication appears different than that of of the yeast proteins. For RecQ4 only three putative Cdk-phosphorylation sites have been experimentally validated and they all map to the Sld2 homology region. Data reported by Xu *et al.*, (2009) suggest that Cdk phosphorylation at the Sld2-like domain of RECQ4 may serve as a switch to activate RECQ4 helicase activity (i.e., an activity that is specific for metazoan protein) during replication by allowing transient dissociation of RECQ4 from MCM10. Since the experimental approach used involved phosphorylation of a recombinant Sld2-homologous fragment by whole cell extracts, the responsible kinase is unknown and actual phosphorylation of other potential phosphorylation sites cannot be excluded. A sizable number of Cdk phosphorylation sites have been reported in Treslin (Fig. 2.4 B and C). Treslin binding to the origin appears dependent upon CycE-Cdk2 activity and the binding of human TopBP1 to Treslin is Cdk-dependent (Kumagai, *et al.* 2010).

The second role of multisite protein phosphorylation in budding yeast is that of preventing DNA re-replication. DNA re-replication is a common feature found in tumor cells (Cook 2009, Hook, *et al.* 2007, Truong and Wu 2011). Modeling and simulation experiments (Brummer, *et al.* 2010) underline the importance of appropriate spacing between G1-Cdk (i.e., Cln1,2-Cdk1) and S-Cdk (ie., Clb5,6-Cdk1) activity in preventing re-replication in yeast. Alteration of the ratio between Sic1 and Clb5,6 or conditions that make DNA replication initiation independent of S-Cdk lead to sparse firing and re-replication (Brummer, *et al.* 2010).

Could a similar mechanism be operative in mammalian cells as well? Figure 2.6 and 2.7 provide a temporal time-line of phosphorylations regulating DNA replication initiation in yeast and mammalian cells. Involvement of a given cyclin-Cdk complex in any of the phosphorylation events depends on timing of syntesis and degradation of the components of each complex. In several cases the actual cyclin and/or Cdk subunit involved is not known, so that a precise correspondence between the functional role of yeast and mammalian cyclin-Cdk complexes in each regulated step cannot be drawn. Considering mammalian cells, with their very large nucleus, one should also wonder whether it is correct to assume a homogeneous distribution of key molecules (as carried out in the yeast model), or is it necessary to consider possible waves of replicative factors moving over replicative origins (Mechali 2010).

2.6 Conclusions and perspectives

Initiation of eukaryotic DNA replication is a very complex sequence of events. Results summarized in this thesis highlight a remarkable conservation of the general architecture of this central biological mechanism. Many steps are conserved down to molecular details and are performed by hortologous proteins with high sequence conservation, while differences in molecular structure of the performing proteins and their interactions are apparent in other steps. Notably, several metazoan-specific proteins exist that make initiation of DNA replication in these organisms more complex and subject to a larger number of potential regulatory mechanisms.

Tight regulation of DNA replication initiation is achieved through protein phosphorylation, exerted mostly by Cyclin-dependent kinases. Differently from yeast, in somatic mammalian cells no convincing evidence to support a role of distributive multisite protein

phosphorylation in the initiation of DNA replication is so far available. However, we cannot exclude that the mechanism orchestrating the coordination of replication initiation in yeast is present also in mammals. Initiator factors that can be subjected to multiple Cdks phosphorylation with distributive mechanism could be potential candidates of such regulatory mechanism and could be involved in the coordination between the activation of the pre-Replication Complex (due to phosphorylation of MCM2-7) and the assembly of the replication factors at the initiation sites. On the other hand, one has to recall that the switch-like onset of DNA replication provided by this regulatory device is instrumental, in budding yeast, in ensuring a short S phase period. As discussed in section *2.2*, in somatic mammalian cells the onset of DNA replication is instead not synchronous for a large number of origins (as occurs in budding yeast) and the S phase duration is much longer, while the rate of fork progression is not very different. As shown by studies in *Xenopus* a synchronous onset of DNA replication takes place in embryos, where it would be interesting to analyse the pattern of phosphorylation of the proteins discussed in 2.**4** and 2.**5**. It should also be remembered that multisite phosphorylation of Rb - that control transcription of genes essential for G1/S transition and DNA replication might contribute to impart a switch-like response to DNA replication (see *2.4.1* and references therein).

A system-level device that appears required to avoid DNA re-replication in yeast is an appropriate temporal spacing between G1-Cdk and S-Cdk activities. Even in the absence of distributive multisite phosphorylation, the requirement for time spacing may be achieved by sequential chromatin loading of multiple replication factors, each of which may be regulated by various Cdks phosphorylation. This alternative mechanism could allow the introduction of more control points and to maintain the temporal delay between the pre-RCs activation and the initiation event, in a system where more Cdk-activities are present. A systematic

examination of kinetics of cyclins, Cdks, their complex formation, activity, substrate phosphorylation and their correlation with DNA re-replication in normal and transformed cells would be highly informative and provide clues on causative dynamics of DNA re-replication and genomic instability that might prove valuable in cancer biology.

Taken together the observations reported previously, it is clear that drawing accurate parallelisms between mammalian and yeast cells not only requires much more data than those currently available but needs to take into account differences in regulatory circuits as well as different evolutive strategies. For instance yeast uses both Cln1,2-Cdk1 and Clb2-Cdk (i.e., a G1-Cdk and a G2/M-Cdk) to promote Cdc6 degradation, while mammalian cells use only CycA-Cdk2 (i.e., a S-Cdk). CycE-Cdk2 that at least in activation of the G1/S transcriptional program plays a role parallel to that of Cln1,2, plays here an opposite role by preventing Cdc6 degradation (and hence promoting pre-RC assembly). A similar role in yeast is played by mitotic cyclins that stabilize Cdc6 by binding it at G2/M. While these differences undoubtedly reflect differences in synthesis of Cdc6 during the cell cycle in yeast and mammalian cells, they underlie the fact that different strategies may have originated during evolution to handle the same problem, possibly because of the different constraints imposed on DNA replication in different organisms. A systems biology approach, conjugating molecular analysis with modelling and computational investigations, possibly at a single cell level, appears therefore to be the best choice to investigate the control of DNA replication in mammalian cells, in order to reach a better understanding of this central cellular function.

CHAPTER

3

Structural and system-level analysis of regulation of Rb and Whi5 proteins: role of phosphorylation and comparative analysis of intracellular networks

3.1 Introduction

Among many cell cycle regulator proteins Rb is one the best studied proteins since its discovery in 25 years ago as first tumor suppressor. Besides its central regulatory function of cell cycle progression which was derived from its capability to arrest cells in G1, the other functions of Rb includes control of cellular differentiation, regulation of apoptosis dependent cell death, maintenance of senescence or permanent cell cycle arrest and protection of genomic and chromosomal stability (Chicas *et al.* 2010). The Rb counterpart in budding yeast as proposed is Whi5 derived from its deletion that not only influence cell size but also the subset of these mutants that affect the critical cell size threshold which is required for passage through START by accelerating the G1/S phase transition while overexpression of Whi5 causes G1 delay and an increase in cell size in wild-type cells (Jorgensen *et al.* 2002, Zhang *et al.* 2002). Beyond its regulatory function of cell cycle progression in budding yeast very little is known about other function of this protein.

Some studies are published regarding interaction map of Rb/E2F protein such as a curated interaction map contains 78 proteins (Calzone *et al.* 2008). Another attempt was made to construct of a large-scale molecular interaction map of cell-cycle process for budding yeast which also include a brief interaction map of Whi5 (Kaizu *et al.* 2010). Though some reviews were focused on functional similarities between these two proteins (Cooper 2006), a detail comparative study of Whi5 and Rb protein and their interaction network is essential for understanding their function in detail.

The goal of this analysis is to explore similarities and differences between Rb and Whi5 specially their human and budding yeast counterpart proteins and will cover not only their sequence and

structure level architecture but also their involved interaction pathway and controlled target genes to visualize the system level similarity between these two proteins. Insight of these analyses will eventually show the path to build model(s) of Whi5 and Rb regulation capturing essential features of the process will allow probing evolutionary conservativity of the function of Whi5 and Rb and will give testable predictions to be used to design experiments aimed at defining the main regulatory events of the circuit.

3.2 Materials and Methods

All the bioinformatics tools used are readily accessible through the relevant websites (see Appendix E); for each predictor, the default settings were used, except where otherwise stated.

3.2.1 Bioinformatic analyses of Whi5 and Rb

Secondary structure of Whi5 was assessed by the suite of algorithms offered by the server Proteus, that uses a "Jury of Experts" approach involving predictions from PSIPRED (Jones 1999), JNET (Cuff and Barton 2000), TRANSSEC (a locally developed tool) and structural alignment (XALIGN). The sequences of Rb-related proteins from different *Saccharomycetales* were analyzed by Pfam (Finn *et al.* 2010, Sonnhammer *et al.* 1997). The relative disorder was analyzed by the Composition Profiler (Vacic *et al.* 2007), taking advantage of the standard amino acid data sets, Disprot 3.4 and Swissprot 51, provided by the program. The plot shows the fractional difference in amino acid composition of Whi5 and of a set of intrinsically disordered proteins from the DisProt database (Sickmeier *et al.* 2007)(grey bars) relative to a reference set of ordered, globular proteins. The fractional difference is calculated as $(C_X\text{-}C_{order})/C_{order}$, where C_X is the content in a given amino acid of Whi5 (or of the set of intrinsically disordered proteins) and C_{order}

is the corresponding value in the set of ordered proteins. Negative fractional difference indicates depletion, while positive difference indicates enrichment, in the corresponding amino acid. Amino acids are arranged on the x axis from the most rigid to the most flexible according to the Vihinen's flexibility scale (Vihinen 1987). The error bars correspond to the confidence intervals evaluated by the 10,000 bootstrap iterations in the definition of the reference protein sets. Amino acids were ordered by Vihinen flexibility scale (Vihinen *et al.* 1994). Algorithms used to characterize the physiochemical properties of amino acids were the charge−hydropathy (CH) plots from Predictor on Natural Disordered Regions, PONDR (Uversky *et al.* 2000) and FoldIndex (Prilusky *et al.* 2005) with a sliding window of 31 amino acids. CH is a binary predictor that classifies proteins as "mostly" disordered or ordered on the basis of the mean net charge and hydropathy. In a two-dimensional space defined by mean hydrophobicity 〈H〉 and mean net charge 〈q〉, a single line 〈q〉 = 2.785 〈H〉 -1.151 separates IDPs from those with well-defined folds (Uversky *et al.* 2000).

The prediction of intrinsic structural disorder by PONDR (Oldfield *et al.* 2005) was done with the algorithms VL-XT, suitable for the prediction of regions locally ordered, containing short motifs that serve as binding site, and VSL1, suitable to predict both long and short disordered regions. Similar results are obtained with the algorithm VSL2 (Obradovic *et al.* 2005, Peng *et al.* 2006); that combines two predictors optimized for the recognition of short and long disordered regions. VSL2 can be considered one of most advanced predictors based on the concept that short disordered regions are context dependent, while long disordered regions are entirely defined by their own amino acid composition. The analysis was carried out with the VSL2B algorithm. Analysis of isoelectric point and of charged amino acid has been carried out by ProtParam (Wilkins *et al.* 1999). The prediction of intrinsic

structural disorder of human Rb protein was also calculated by PONDR algorithm.

The prediction of phosphorylation sites of Rb and its homologs in other species was carried out by the program GPS 2.1 (Group-based Prediction System, version 2.1) (Xue *et al.* 2008), combining the results obtained with high threshold for the three kinases Cdk2, Cdk4 and Cdk6. This procedure gave for the analysis of human Rb the phosphorylation pattern most similar to the experimental one. The prediction of Whi5 phosphorylation sites was carried out by the program PPSP (Prediction of PK-specific Phosphorylation site), (Xue *et al.* 2006) based on the Bayesian decision theory and trained with an unambiguous experimentally verified data set of >2,000 non-redundant positive data. This algorithm was chosen because its predicted Cdk-phosphorylation map for *S. cerevisiae* fully coincides with the experimental one (Wagner *et al.* 2009). To assign a relative score of probability to be phosphorylated by a Cdk to each site, we applied the algorithm GPS 2.1, setting a medium threshold for the recognition of a generic Cdk. The prediction of phosphorylation sites of Whi5 and its homologs in other species was also carried out by the program GPS 2.1.

3.2.2 Constructing Interaction map

All genetic and physical protein interaction datasets of Whi5 of *Saccharomyces cerevisiae* and pocket proteins of *Homo sapiens* were primarily downloaded from the BioGRID database v3.2 (Stark *et al.* 2006). In this analysis all genetic and physical interactors of Whi5 are taken from mainly BioGRID which reports the most comprehensiveness for yeast protein interactors (Cusick *et al.* 2009). To make Rb dataset comprehensive, interaction datasets were also collected from iRefWeb (Turner *et al.* 2010) which is a web interface to a broad landscape of data on protein-protein interactions (PPI) consolidated from databases: BIND,

BioGRID, CORUM, DIP, IntAct, HPRD, MINT, MPact, MPPI and OPHID. Some of Rb interactors were also collected from literature search. The collected datasets used in this analysis are provided in Appendix B. Interaction datasets of Whi5 and Rb were provided as input data for Cytoscape 2.8 which is a tool for visualizing and integrating complex networks (Smoot *et al.* 2011). In case of Whi5 three interaction maps were built: a) physical interaction map, b) genetic interaction map and c) multilevel interaction map of physical and genetic interactors of Whi5 (Figure 3.7). In case of Rb interactors a physical interaction map was built. All maps also characterized in groups and differentiated by colors based on their functions and then exported to graphic software to enhance graphic view (Figure 3.6 and 3.7). The function of each interactor was collected from literature search (Costanzo *et al.* 2010).

3.2.3 GO Enrichment analysis

The Gene Ontology (GO) database (Harris *et al.* 2004) allocates biological descriptors also known as GO terms to genes on the basis of the properties of their encoded products. GO terms can be three types: cellular component, biological process, and molecular function. Genes allocated the same GO term regarded as same GO category of genes that are more closely related some aspect of their biological roles.

GoBean is a comprehensive and flexible GUI tool for GO term enrichment analysis, combining the merits of other programs and incorporating extensive graphic exploration of enrichment results (Lee *et al.* 2012) was used to enrich GO terms for Rb and Whi5 interactors. This tool also used to enrich GO terms for p107 and 130 interactors. This tool calculates p-values using the upper tail of a hypergeometric distribution probability. The implementation was fast enough for most interactive analyses. The Term-for-Term (classic) enrichment algorithm used in many popular tools was implemented to obtain GO data. For multiple testing corrections, Bonferroni-Holm method was applied.

GO term enrichment data was used as input of Revigo (Supek *et al.* 2011) web server which generates a representative subset of the GO terms enriched in a given list through a clustering algorithm relying on semantic similarity measures. GO term enrichment data for biological process of Rb, p107, p130 and Whi5 interactors were obtained from GoBean and common GO terms of Rb, p107 and p130 interactors and common GO terms of Rb and Whi5 were calculated (see Appendix B). These GO terms are used as input for Revigo program. Revigo generates treemap and scatter plot based GO terms sorted by absolute log10 value. If log value was not provided, map was calculated using uniqueness values. R packages were used to refine the tree maps obtained from Revigo.

3.2.4 Pathway Model Building

Pathway model construction was based on functionally characterized data of interactors of Rb and Whi5 as well as supported by published literature data. These pathways were built based on briefly analyzed through literature search. Though initially currently available various model and pathway building tools and web service such as CellDesigner 4.2, Cell Illustrator, rxncon etc (URL in Appendix E). were used to build a better understandable version Whi5 regulated pathway model, finally a customized graphical version of the model was considered due to reducing complexity and better understandability.

3.3 Results and discussion

3.3.1 Analysis of disorder and phosphorylation sites of Rb proteins

3.3.1.1. Domain organization, disorder distribution and phosphorylation pattern of Rb proteins

Human Rb is a 928-amino acid protein whose 3D structure has been extensively studied. It consists of three major domains (Fig. 3.1): an N-terminal domain (residues 52-355, RbN), a pocket domain (residues 380-787) and a C-terminal domain (residues 787-928, RbC) (Rubin *et al.* 2005). Although most of the Rb structure has been determined (PDB accession numbers are reported in appendix B Table 3), there are regions that escaped X-ray diffractometry analysis and that are possibly structurally disordered. Indeed, only a tiny portion of the RbC domains has been reported. To assess the extent of structural disorder in Rb, we collected literature data on experimentally determined structure and combined this information with prediction of structural disorder. The analysis of structural disorder was carried out by VL3, predicting the occurrence of large disordered sequences (see material and methods), and by VL-XT, an algorithm suitable for the prediction of locally disordered regions, and presented in Figure 3.1, while the figure resumes VL-XT profiles, the position of experimental phosphorylation sites and available structural data.

We found a good correspondence between regions predicted as disordered and structurally undetermined regions or experimentally determined as coiled coil. Overall, the longer disordered regions, in the following given according to VL3 prediction, were found at the N-terminus (residues 1-60), inside the RbN domain (residues 246-267), at the linker positions - interconnecting RbN and the pocket domain (residues 339-379 of RbIDL) or within the bipartite pocket domain (residues 580-635 of RbPL) - and in the whole RbC domain (780-928). The correlation between the map of structural disorder and that of

phosphorylatable residues was, as expected (Iakoucheva *et al.* 2004), very good, with the only exception of N-terminal disordered region that does not contain any phosphorylatable residue. On the contrary, the C-terminal disordered region contains the most numerous set of phosphorylatable residues (i.e. S780, S788, S795, S807, S811, T821 and T826), whose importance was only in some cases experimentally proved (i.e. S807, S811, T821 and T826). How much conserved is the overall domain organization and the occurrence of structural disorder and phosphorylation along the evolutionary tree of Rb? Among vertebrates, amino acid sequences of Rb proteins are highly conserved and traces of the evolutionary origin of regions A and B of pocket domain have been found in Archea (i.e. *Archeoglobus fulgidus* and *Methanococcus jannashii*) and poxviruses (Takemura 2005). To evaluate the evolutionary conservation of disordered regions, we compared human Rb to ortholog proteins from vertebrates *Mus musculus*, *Gallus gallus*, *Xenopus laevis*, *Danio rerio* that are highly conserved, and to hortologs from *Drosophila melanogaster*, *Arabidopsis thaliana*, *Pisum sativum* and *Caenorabditis elegans* that are much less conserved (Takemura, 2005). Since information on structures and experimental phosphorylation data are not homogenously available for all selected Rb sequence motifs and phosphorylatable residues were analyzed by the predictive tools Meme and GPS2.1 respectively. We observed that overall motifs pattern of human Rb is conserved among the orthologs and that a given functional domain of human Rb can contain several different sequence motifs identified by Meme algorithm, which turns out as a valuable tool to accomplish a detailed sequence analysis.

Thus we can conclude that the conservation of pocket domain organization is reflected by the similarity of motif patterns among organisms very distantly related in the evolutionary tree. The N- and C-terminal regions are confirmed in all orthologs as devoid of sequence

motifs, while phosphorylation pattern is conserved especially among vertebrates (see Fig. 3.1).

The analysis by PONDR algorithms VL-XT and VL3 (Fig. 3.1) allowed to point out that also the occurrence of disordered regions is conserved in all the analyzed proteins. More in detail, we recognized that each sequence could be overall divided in 4 disordered and 4 ordered blocks of sequence. Hence, the analysis of hortologs sequences was carried out for each block by multiple alignment analysis with ClustalW2 (Thompson *et al.* 1994), and by computing the overall mean evolutionary distance with Molecular Evolutionary Genetics Analysis (MEGA) software (Tamura *et al.* 2011). In general, as expected, in the whole group of 9 hortologs considered, disordered blocks were less conserved than ordered ones. This result is not surprising if we consider that, although structural disorder is a feature very conserved through the evolution of homologous proteins, it can be obtained by an almost unbiased array of amino-acids (Brown *et al.* 2010, Brown *et al.* 2002, Daughdrill *et al.* 2007) and it has been already observed that random substitutions are more common in disordered proteins than in ordered (Brown *et al.* 2010, Brown *et al.* 2002, Daughdrill *et al.* 2007).

Among disordered blocks, the first (N-terminus, amino acid from 1 – 60 in human Rb) and the fourth (C-terminal end, from position 780 to 928 in human Rb) are respectively the least and the most conserved. This was even more evident when we restricted our analysis to the group of 5 Rb proteins from vertebrates (*H. sapiens*, *M. musculus*, *G. gallus*, *X. laevis*, *D. rerio*).

We believe that it is not coincidental that the most diverse block does not contain any phosphorylatable site, while the most conserved contains the most numerous cluster of phosphorylatable sites (seven in the case of human Rb), with the only exception of *D. melanogaster* Rb.

We believe that it is not coincidental that these two blocks profoundly differ also for the occurrence of phosphorylatable sites. Indeed, there are no phosphorylatable sites at the N-terminal sequence, while the most conserved C-terminal region contains the most numerous cluster of phosphorylatable sites (seven in the case of human Rb), with the only exception of *D. melanogaster* Rb. This could sustain the hypothesis that the presence of phosphorylation sites, and more in general of post-translational modifications, can anchor nucleotide sequence of a disordered region through/against the genetic drift. In other terms, sites that specifically code for protein modifications can introduce functional constrains that make the evolution of disordered regions less "neutral" than expected to simply select for flexibility. This observation is in agreement with the interpretation given by Brown and Coworkers to the accumulation of specific amino acids in disordered proteins. For instance, the higher frequency of Asn residue in disordered sequences has been related to its role as a common site for glycosylation, while a higher frequence of Gly has been referred to its role in promoting the backbone flexibility (Brown *et al.* 2010).

The C-terminal phosphorylation sites in the group of 5 vertebrates were analysed by GPS2.1 for their probability to be phosphorylated by Cdk. We found that within each protein sequence, phosphorylatable residues were received very similar probability score, moreover the average score of analyzed sequences turned out very similar among them (5.26±0.77 for *H. sapiens*, 5.10±0.50 for *M. musculus*, 4.96±0.62 for *G. gallus*, 5.26±0.55 for *X. laevis*, 5.23±0.74 for *D. rerio*). The relatively high number of these equivalently phosphorylatable residues in a defined disordered region seems reminiscent of Sic1 phosphodegrons and promt us to suggest a mechanism of recognition similarly based on the allovalency.

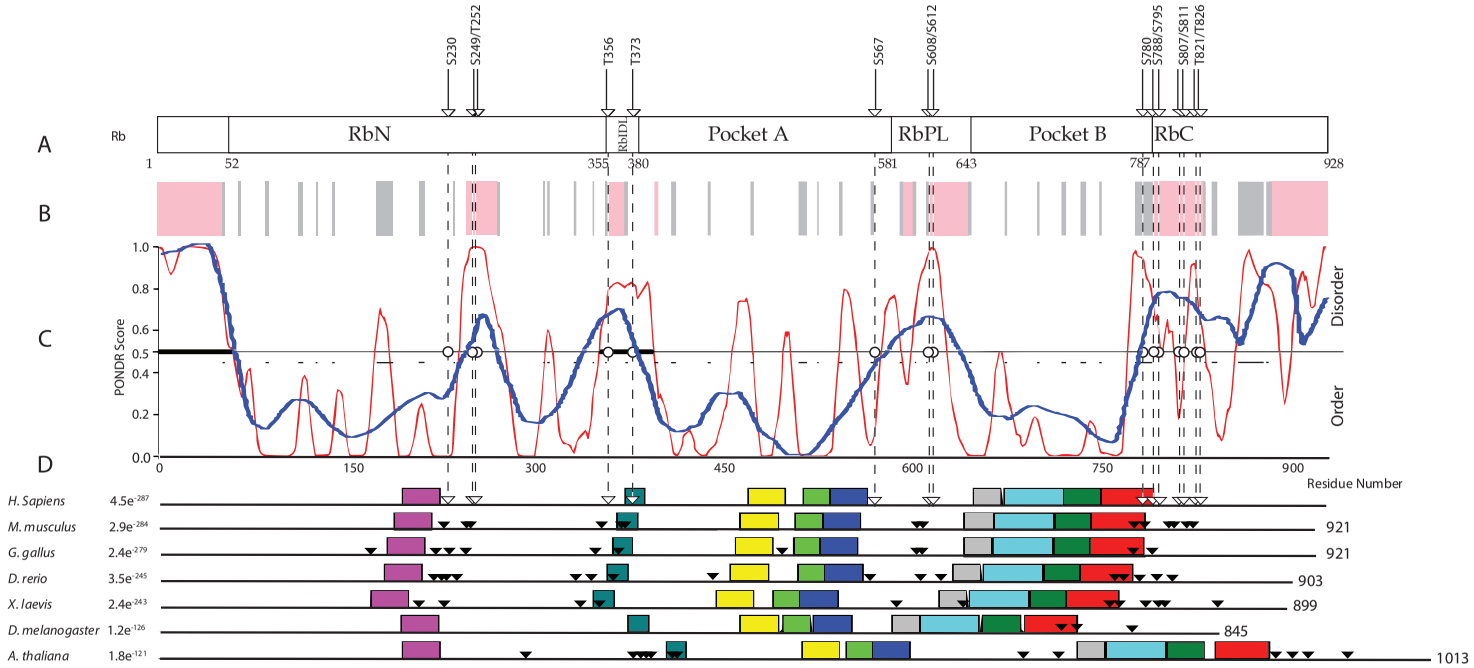


Fig. 3.1 Structural organization of human Rb along the evolution of multicellular Eukarya Schematic representation of Rb homologous from different multicellular Eukarya. The functional domain organization of human Rb (A) has been compared with conserved motifs (D) found by the Meme algorithm along the amino acid sequences of homologs. Triangles indicate the position of Cdk1 phosphorylatable residues. Disorder prediction by VL-XT and VL3 and unre-solved 3D structure with coiled coil region of human Rb are shown on C and B respectively. Grey shadow indicates experimentally determined coiled coil region, pink shadow indicates unsolved 3D structures.

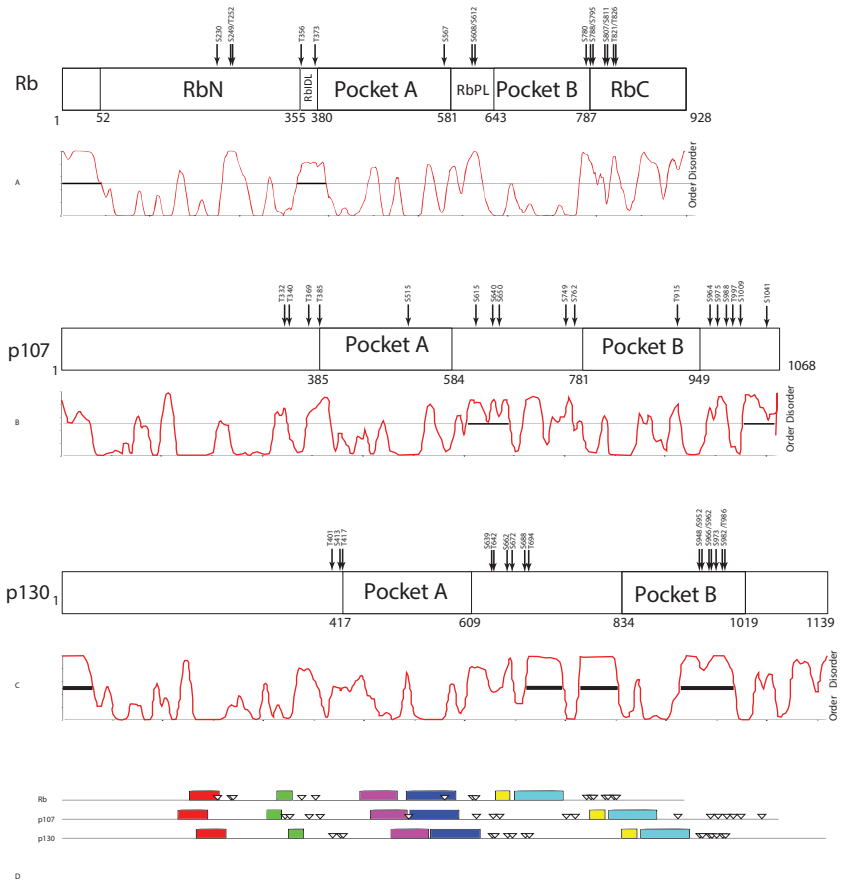


Fig. 3.2 Structural organization of human Rb related proteins. (A) Rb PONDR disorder prediction by VL-XT (continuous red line). (B, C) Disorder prediction by VL-XT (continuous red line) combined with functional domain organization for paralogs p107 and p130 with experimentaly determined Cdk phosporylation sites. (D) The pattern of conserved sequence motifs predicted by MEME in human Rb and its paralogs p107 and p130.

3.3.1.2 Domain organization, disorder distribution and phosphorylation pattern is conserved among human pocket proteins

Our analysis on the evolutionary persistence of "disordered blocks" in the orthologs of Rb are fully in agreement with the raising concept that structural disorder is evolutionary less "elusive" than commonly believed and witnesses the concept that "evolution of disordered proteins is driven by their structure and function just as the evolution of ordered proteins" (Brown *et al.* 2010). We complete our analysis on disordered blocks and phosphorylatable in the Rb-like family of proteins, including in our analysis paralogs of Rb. Indeed, human and in general mammalian genomes contain apart Rb, the Rb-like proteins, RBL1/p107 and RBL2/p130, that show both redundant and unique functions, as reviewed in (Wirt and Sage 2010). As previously showed by alignment analysis, Rb shares 32% and 31% homology with p107 and p130 respectively, while p107 and p130 share 53% between them (Mulligan and Jacks 1998).

As expected, the most structurally conserved region among the three proteins is the pocket domain, as reflected by sequence similarity and analogous interactions with viral proteins containing the LXCXE motif (Cobrinik 2005, Cobrinik *et al.* 1993, Hannon *et al.* 1993, Li *et al.* 1993, Mulligan and Jacks 1998). Functional/structural organization of p107 and p130 is reported in figure 3.2.

We analyzed the structural organization of p107 and p130 by open-source algorithms able to predict structural disorder and to analyze the conservation of sequence motifs. More in details, we used VL-XT to predict structural disorder, and the software Meme (Bailey and Elkan 1994) to discover motifs in the group of three proteins. Data on experimental phosphorylation sites are available for all the three proteins (Burke *et al.* 2010, Xiao *et al.* 1996). The overall results are resumed in figure 3.2. Our analysis indicates that the overall lengthening

of p107 (1068 amino acids) and p130 (1139 amino acids) on respect to Rb (908 amino acids) can be mainly accounted to the expansion of the disordered RbPL linker within the bipartite pocket domain; it is noteworthy also the occurrence in p130 Pocket B of a disordered sequence much longer than observed in the other paralogs.

We observed that overall motifs pattern is conserved among the paralogs proteins and that a given functional domain of human Rb can contain several different sequence motifs identified by Meme algorithm, which turns out as a valuable tool to accomplish a detailed sequence analysis.

Thus we can conclude that a similar pocket domain organization is reflected more in detail by the conservation of sequence motifs. The N- and C-terminal domains are confirmed in all three paralogs as devoid of sequence motifs, but only the C-terminal domain is rich of phosphorylation sites.

3.3.2 Analysis of disorder and phosphorylation sites of Whi5 proteins

3.3.2.1. Whi5 is a disordered protein

Results summarized in the previous paragraph indicate that Rb is a complex protein in which several autonomously folding domains alternate with intrinsically disordered regions. Multiple phosphorylations of specific Ser and Thr residues modulate Rb ability to interact with its partners (namely the E2F transcription factor) by promoting folding of disordered domains as well as by allosteric modulation of $E2F^{TD}$-pocket association and protein binding to the LxCxE site. Whi5 shares no sequence homology with Rb (Cooper, 2006 and Hasan *et al.* unpublished) and no structural information is yet available for all members of Whi5 protein family recorded by Pfam.

The only trait common to them is the so called "Whi5 domain", a stretch of ~25 amino acids(from amino acid 181 to 205 in Whi5 from *S. cerevisiae*, Whi5),predicted with high confidence to acquire an α-helical secondary structure. Therefore we decided to investigate Whi5 structure by bioinformatics tools devoted to structural disorder analysis (see Materials and Methods for details on the used tools).

The composition profile of Whi5is characterized by depletion of amino acids that promote order (i.e., Cys, Trp, Phe, Tyr, Val and Ile) and abundance of residues associated with disorder (i.e., Gln, Ser, Pro,Glu), (Figure 3.3 A). The very high content of Thr, Arg and Pro–the latter considered the main disorder-promoting amino acid - is remarkable. Indeed, Proline content and net charge of a polypeptide are considered the sequence determinants of compaction/disorder in IDPs(Marsh and Forman-Kay 2010).Consistently, the Uversky's plot, (Figure 3.3 B) an empirical graph where the mean average hydrophobicity (*R*) is plotted against the mean net charge (*H*) (Uversky *et al.* 2000) classifies Whi5 among disordered proteins.

Structural disorder was then analyzed by three different algorithms. FoldIndex allows to predict disorder propensity along the whole sequence. When applied to Whi5, it shows a roughly modular organization, with modules of approximately 50 amino acids. Structural disorder is recognized along the whole sequence with few exceptions, approximately around positions 190, 225 and 240 (Fig. 3.3C, green areas). The output of VSL1 (PONDR) and VSL2B (DisProt) (Fig. 3.3 D)- two predictors optimized for the recognition of short and long disordered regions- coherently predicts an extensive region of naturally disordered structure along the whole sequence, with the only exception of the region from 180to 200, pretty well coincident with the Whi5 domain recognized by Pfams as the signature of Whi5-like family. Finally, VL-XT (PONDR) (Fig. 3.3D) detects other than the peak of Whi5 signature

(amino acids 177-203, according to VL-XT output), two more downward peaks centered at positions 90 (amino acids 78-104) and 230 (amino acids 215-240), possibly corresponding to specific protein or DNA recognition sites within long disordered regions.

The structural disorder of Whi5 is indirectly confirmed by an independent analysis run with different secondary-structure prediction algorithms. Whi5 appears scarcely prone to form secondary structures, with α-helices accounting for 30% of amino acid sequence, whereas the remaining 70% has a random-coil conformation (Fig. 3.3 E). More in detail, the signature of Whi5 family appears to correspond to two close α-helices linked by a short random coil sequence.

An experimental confirmation of our prediction comes from the analyses on recombinant Whi5, fused to an N-terminal poly histidine tag and expressed in *Escherichia coli* cells (Appendix C Fig. 2). The recombinant Whi5 shows marked thermostability, oversensitivity to tryptic digestion and anomalous electrophoretic mobility (apparent ~43 kDa instead of the theoretical 32.8 kDa). Its retention time in gel filtration chromatography much lower than expected for a globular protein of ~ 33 kD allows to calculate a ratio between the experimental and theoretical hydrodynamic radius of 1.589, indicating that Whi5 behaves as a molten globule(Tompa 2009, Uversky 2002) (Appendix C Fig. 2).

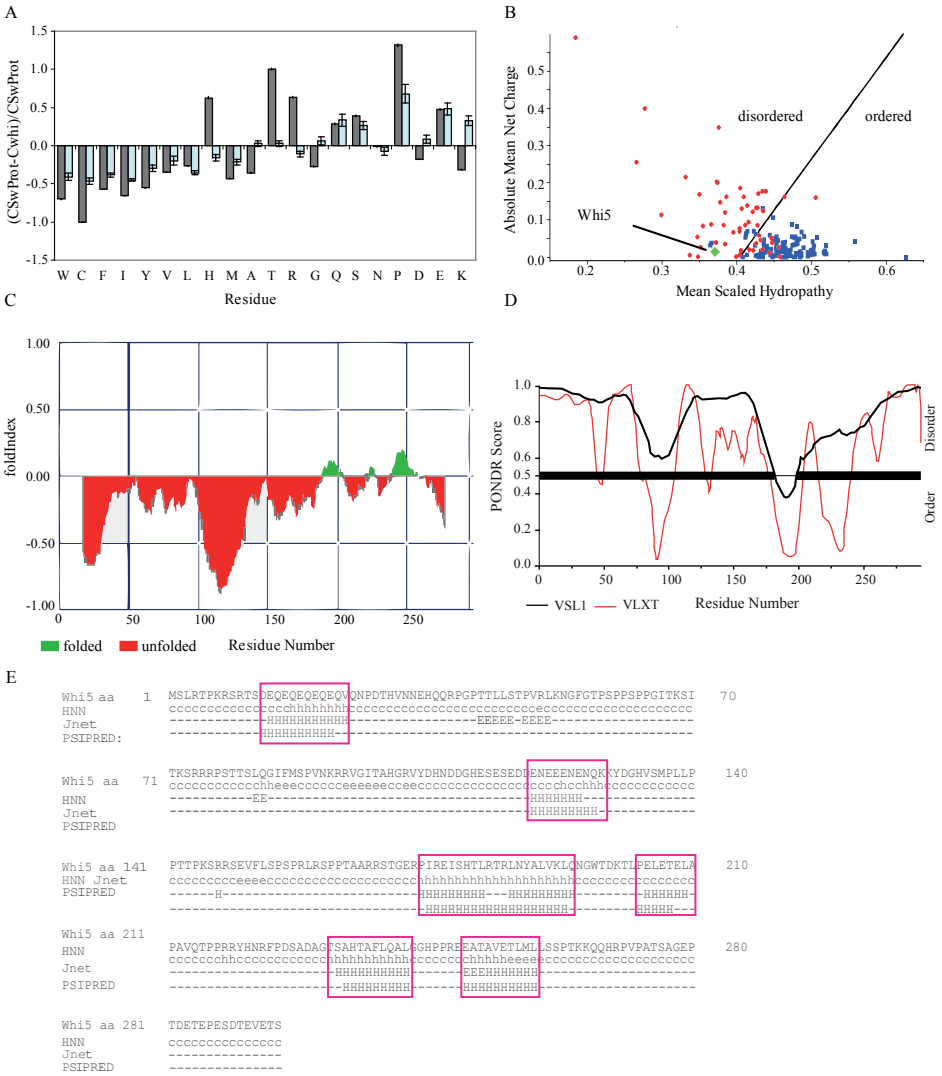


Fig. 3.3. Compositional and sequence analysis of *S. cerevisiae* Whi5 (A) Compositional profiling of Whi5. The plot shows the fractional difference in amino acid composition of Whi5 and of a set of intrinsically disordered proteins from the DisProt database (grey bars) relative to a reference set of ordered, globular proteins. Negative fractional difference indicates depletion, while positive difference indicates enrichment, in the corresponding amino acid. Amino acids are arranged on the x axis from the most rigid to the most flexible according to the Vihinen's flexibility scale (Vihinen, 1987). The error bars correspond to the confidence intervals evaluated by the 10,000 bootstrap iterations in the definition of the reference protein sets. (B) Charge-hydropathy plot of Whi5 (green diamond). The plot is an empirical graph representing data of net charge and mean hydrophobicity for a set of globular proteins (blue square) and a set of disordered proteins (red circle). The two groups are separated by a straight line <charge> = 2.743 <hydropathy> - 1.109 (Oldfield et al., 2005). (C) FoldIndex plot of the disorder of Whi5. Mean hydrophobicity and net charge calculated for residues included in a sliding window of 31 residues have been used to predict disorder of the middle residue (Prilusky et al., 2005). (D) PONDR disorder prediction by VL-XT (red line), VSL1 (black line). (E) Secondary structure prediction for Whi5 obtained by HNN, PSIPRED e Jpred (shown the only Jnet prediction). A simplified output of PSIPRED is used showing only secondary structure elements predicted with confidence ≥ 5, in the range 0-9. Boxes highlight sequences predicted to acquire α-helicalstructure from the three programs.

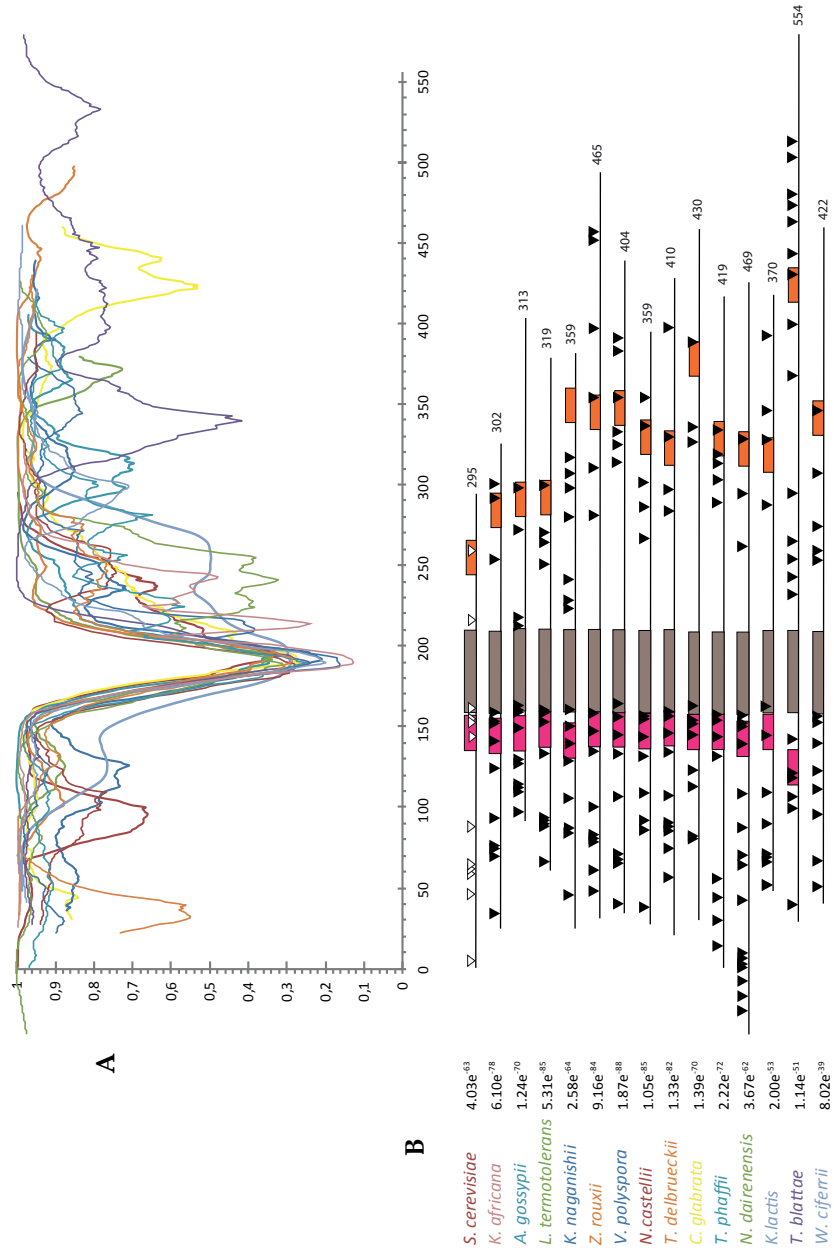


Fig. 3.4. Conservation of structural disorder among Whi5 homologues in Fungi. (A) The plots represent the prediction of structural disorder by VSL2B for Whi5 homologous (see text) from the same yeast species listed in the panel B, in the same color code; the plots were overposed by overlapping the deepest downward spike. This conserved sequence corresponds to the in the panel B. (B) Schematic representation of S. cerevisiae Whi5 homologous from different *Saccharomycetales*. Boxes represent conserved motifs found by the Meme algorithm, while their position along the amino acid sequences of homologues was refined as explained in Appendices. Vertical bars indicate the position of predicted Cdk1 phosphorylatable residues.

3.3.2.2. Whi5 sequence is lost in Eukarya genomes, except in Fungi

According to Pfam, the Whi5-like family contains 93 sequences sharing the family domain or signature, a 25-amino acid sequence that in *S. cerevisiae* Whi5 (Whi5) spans the positions from 175 to 200 and is predicted to acquire a helical secondary structure. All members of Whi5-like family are Eukarya, and 86 of them belong to Fungi Ascomycota. Which are the sequence features they share? Is the structural disorder conserved along the evolutionary tree of Whi5-like family? How similar are their sequences? An overview of the Pfam sequences of Whi5-like family, also including the Nmr1 proteins, pointed out to their wide heterogeneity. A BLAST search launched with budding yeast Whi5 against non-redundant protein data bank identifies putative homologs of Whi5 uniquely in the order of *Saccahromycetacetales*. Fifteen sequences were obtained filtering for a maximum e-value of 0.0001, and a minimum match length of 23% with respect to the length of the query sequence (see Appendix B Table 1).

In the group of 15 sequences (hereafter called "Whi5 homologous"), 14 belong to the family of *Saccahromycetaceae*, while *W. ciferrii* belongs to the family of *Wickerhamomycetaceae* and appears as the most isolated branch in the phylogenetic tree of Whi5 homologs (Appendix C Fig. 1).

Among Whi5 homologs, the relatively low identity to Whi5 (37-72%) is also due to differences in length and hence gaps, which indicates insertions and deletions. Thus, while Rb sequences are highly conserved in vertebrates as well as insects and plants (see section 3.1.), the sequence of Whi5 seems lost in the genomes of species more distantly related to *S. cerevisiae* than *Saccharomycetales*. This work confirms and extends the

conclusion attained by a previous overview of literature and sequence data (Takemura, 2005).

3.3.2.3. Conservation of disordered, phosphorylated motifs points to their functional relevance

All analyzed Whi5 homologs are predicted to be largely disordered by VSL2B with a profile similar to that of Whi5. In all cases, most sequences appear almost completely disordered, with the only exception of the Whi5-like family domain (Fig. 3.4 A). It has been observed that structural disorder is highly evolutionary conserved but, in a disordered protein, when an amino acid residue changes, it has more options than ordered residues (Brown *et al.* 2010, Brown *et al.* 2002, Daughdrill *et al.* 2007). The degree of evolutionary freedom is strictly and inversely related to the amount of physical constraints usually required for long-range interactions. Conformational ensembles of IDPs are dominated, on the contrary, by local interactions, depending just on rotational properties that are quite common to most amino acids (Brown *et al.* 2010). Thus, as expected for a disordered protein, Whi5 homologs show very low overall sequence identity or similarity. When we applied the software Meme (Bailey and Elkan 1994), we found that the signature of Whi5-like proteins, referred from now as "motif 2", is flanked by two other conserved motifs ("motif 1" and "motif 3") that fall in disordered regions (figure 3.4). The distribution of motifs along the 15 Whi5 homologues is shown in figure 3.4B. Referring to the amino acid numbering of Whi5/Whi5 homologs, motif 1 (pink box in figure 3.4 B), motif 2 (grey box in figure 3.4 B), and motif 3 (orange box in figure 3.4 B). The sequences of motifs obtained as output of Meme algorithm were aligned by ClustalW (Thompson et. al., 1994) and the motif boundaries were subjected to a manual refinement before any subsequent analysis (Appendix C Fig. 3). Referring to the amino acid numbering of refined Whi5, motif 1 (violet in figure 3.4 B) spans from amino acid 136 to 162,

motif 2 from amino acid 173 to 209 (blue in figure 3.4 B), and motif 3 (red box in figure 3.4 B) from amino acid 250 to 265 (Appendix C Fig. 4).

Only the Whi5 homologous from *W. ciferrii*, the only species not belonging to the family of *Saccahromycetaceae*, does not contain motif 1. The relatively strong conservation of disordered motifs 1 and 3 is surprising, although could be partially caused by an evolutionary constrain exerted by Cdk1-phosphorylatable consensus sites. Indeed, motif 1 contains 4 Cdk1 phosphosites (Thr143, Ser154, Ser156 and Ser161, Whi5 numbering), and motif 3 contains a single phosphosite (Ser262 in Whi5). Notably, motif 2 corresponding to the Whi5-like domain, does not contain any phosphorylatable residue in all analyzed sequences (revised version from Meme output see Appendix C Fig. 3,4).

Among conserved, disordered motifs we find a very high frequency of the non-polar Proline and of aromatic and hydrophobic residues (F, I, L, V) (Appendix C Fig 3). Aromatic and hydrophobic residues occur less frequently in disordered proteins and have been recognized to form the dominating contacts at the interface with ordered binding proteins (Gunasekaran *et al.* 2004). Hence, their conservation in Whi5 disordered motifs could witness their functional importance as contact points at the interface with interacting proteins.

Since charge segregation has been recognized as a major determinant of structural compactness of disordered proteins, we looked at the conservation of charged residues among Whi5 homologs by calculating the local pIs of conserved motifs 1-3 (Appendix B Table 2). We found that motifs 1 have an average pI of 10.23±1.38, being often remarkably more basic than the entire protein (pI ranging from 6.28 to 9.56). The average pI of the motifs 2 is 8.86 ±1.03, while motifs 3 have an acidic pI, with an average value of 4.84±1.18. The only exceptions to the rule that motif 1 is strongly basic and motif 3 markedly acidic is represented by

the acidic motif 1 of *K. lactis* (pI=5.81), and by the basic motif 3 of *K. africana* (pI=8.50). However, *K. africana* motif 3 remains relatively acidic with respect to the full-length protein pI (9.20).

3.3.3. Analysis of interactome of human pocket proteins and budding yeast Whi5

3.3.3.1. The Rb interactome

Protein interactors of human Rb, were obtained as described in Materials and Methods and are reported in Appendix B (Table 4) and Figure 3.5A in which Rb interactors are color-coded according to function. Most notable classes of Rb interactors are transcription factors, protein kinases and their regulatory subunits such as cyclins, histone modification enzymes and protein modification enzymes (Fig. 3.5A).

Rb interacts with many types of transcription factors. Table 4 in Appendix B contains list of transcription factors that interact with Rb. Among the transcription factors most notable transcription factors are E2F (1-4) and DP (1-2) (Giacinti and Giordano 2006). The hypophosphorylated Rb interacts heterodimeric E2F/DP complexes. The E2F family controls hundreds of transcription of genes including genes required for cell cycle progression.

Rb interacts with many types of kinases. Most notables are Cdks which recognize 16 potential phosphorylation sites (Knudsen and Wang, 1996). In early G1, CycD/Cdk4/6 may phosphorylate S249, T356, S807, S811 and T826 sites of Rb (Zarkowska and Mittnacht, 1997). CycE/Cdk2 phosphorylates S612 and T821 sites in G1/S transition. CycA/Cdk2 maintains hyperphosphorylated Rb throughout the S phase. Cdk1 (p34Cdc2) phosphorylates Rb on multiple serine-threonine sites *in vivo* but the outcome of this phosphorylation not reported (Lees *et al.* 1991). Cdk14 (PFTK1) phosphorylates C terminal of Rb (792–928 of Rb) *in vitro*

and the outcome is not known (Shu *et al.* 2007). Cdk5 also phosphorylate Rb multiple sites especially S807/S811 *in vitro* (Hamdane *et al.* 2005). Cyclin T- Cdk9 complex is not cell cycle regulated and involved in the regulation of transcription. This complex also phosphorylates C-terminus of Rb (793-928) *in vitro* (Simone *et al.* 2002). Other notable Rb interacting kinases are MAPK9/14 (p38 kinase) phosphorylating on Ser567 (delivers apoptotic signal) (Delston *et al.* 2011), Chek1/2 mediated phosphorylation of Rb at S612 (leading to the formation of Rb–E2F) (Inoue *et al.* 2007), AURKB phosphorylating on S780 (inhibits of endoreduplication) (Nair *et al.* 2009), ABL1 (c-Abl) binding in the C-terminal of Rb (resulting inhibition of kinase function of ABL1) (Welch and Wang 1993), FRK (Rak) binding Rb (Craven *et al.* 1995), p55 regulatory subunits of PIK3R3 (Phosphatidylinositol 3-kinase) (Xia *et al.* 2003), DGKZ (DGKξ) (Los *et al.* 2006), RAF1 (c-Raf) (Dasgupta *et al.* 2004) and Protein kinase C β 2 (PRKCB) (Suzuma *et al.* 2002). Rb not only interacts with kinases it also interacts with various kinase partner cyclins such as Cyclin A (CCNA1), Cyclin D (CCND1, CCND2 and CCND3), Cyclin E1 (CCNE1) Cyclin T (CCNT2) etc (Simone *et al.* 2002, Yang *et al.* 1999).

Rb interacts with protein modification enzymes. Protein phosphatase 1 (PPP1CA, PPP1CB and PPP1CC) also interact with Rb and may involve in regulation of reversible protein phosphorylation (Durfee *et al.* 1993, Flores-Delgado *et al.* 2007). PP1 harbors on the C terminal domain of Rb which is essential for PP1 activity (Hirschi *et al.* 2010). Sirturin 1 (SIRT1) can deacetylate Rb both *in vitro* and *in vivo* (Wong and Weber 2007). Set7/9 (SETD7) methylates Rb on K873 and K810 (Carr *et al.* 2011) Heterochromatin protein, HP1 (CBX1/5) binds to the methylated site at K873 of Rb and is involved in gene silencing (Fischle *et al.* 2003; Hediger and Gasser, 2006). SMYD2 also methylates Rb on K860 (Saddic *et al.* 2010). L3MBTL1 binds to the methylated site at K860 of Rb and is

involved in condensing chromatin and repressing gene expression (Bonasio *et al.* 2010). MDM2, a RING-finger type ubiquitin ligase, ubiquitinates Rb *in vivo* which leads its degradation via the ubiquitin-proteasome pathway (Uchida *et al.* 2006). Rb also undergoes SUMOylation on K720 which belongs to a cluster of lysine residues and modulate binding of LXCXE motif-containing proteins (Chan *et al.* 2001b, Ledl *et al.* 2005). (See Appendix B Table 4 for a list of other enzymes and factors that interact with Rb).

Rb interacts with histone modification enzymes HDAC1, HDAC2 and HDAC3 (Fajas *et al.* 2002, Sun *et al.* 2007). Histone Acetyl-transferase KAT2B (p300) acetylates Rb on K873/K874 during S phase (Chan *et al.* 2001a). It also interact with DNA methylation Enzyme DNMT1 and DNMT3A (Fuks *et al.* 2003, Robertson *et al.* 2000), Histone-lysine N-methyltransferase SUV39H1, SUV420H1 and SUV420H2 (Gonzalo *et al.* 2005, Nielsen *et al.* 2001) and others.

3.3.3.2. GO term enrichment of Rb interactome

GO term enrichment of Rb interactors, was obtained as described in Materials and Methods and are reported in Appendix B. Figure 3.5 B shows a hierarchical treemap of GO terms enriched in Rb interactors. The plot has been generated by Revigo (default parameters) (see Materials and Methods). In the treemap, single cluster representatives are shown as rectangles joined into superclusters of related terms, whose size reflects the p-value in the underlying GOA database. The largest supercluster (tagged as transcription from RNA polymerase II promoter) includes GO terms related to "regulation of metabolic process, gene expression, RNA metabolic process, regulation of biological process", as well as terms related to protein post-translational modifications, including "protein modification process".

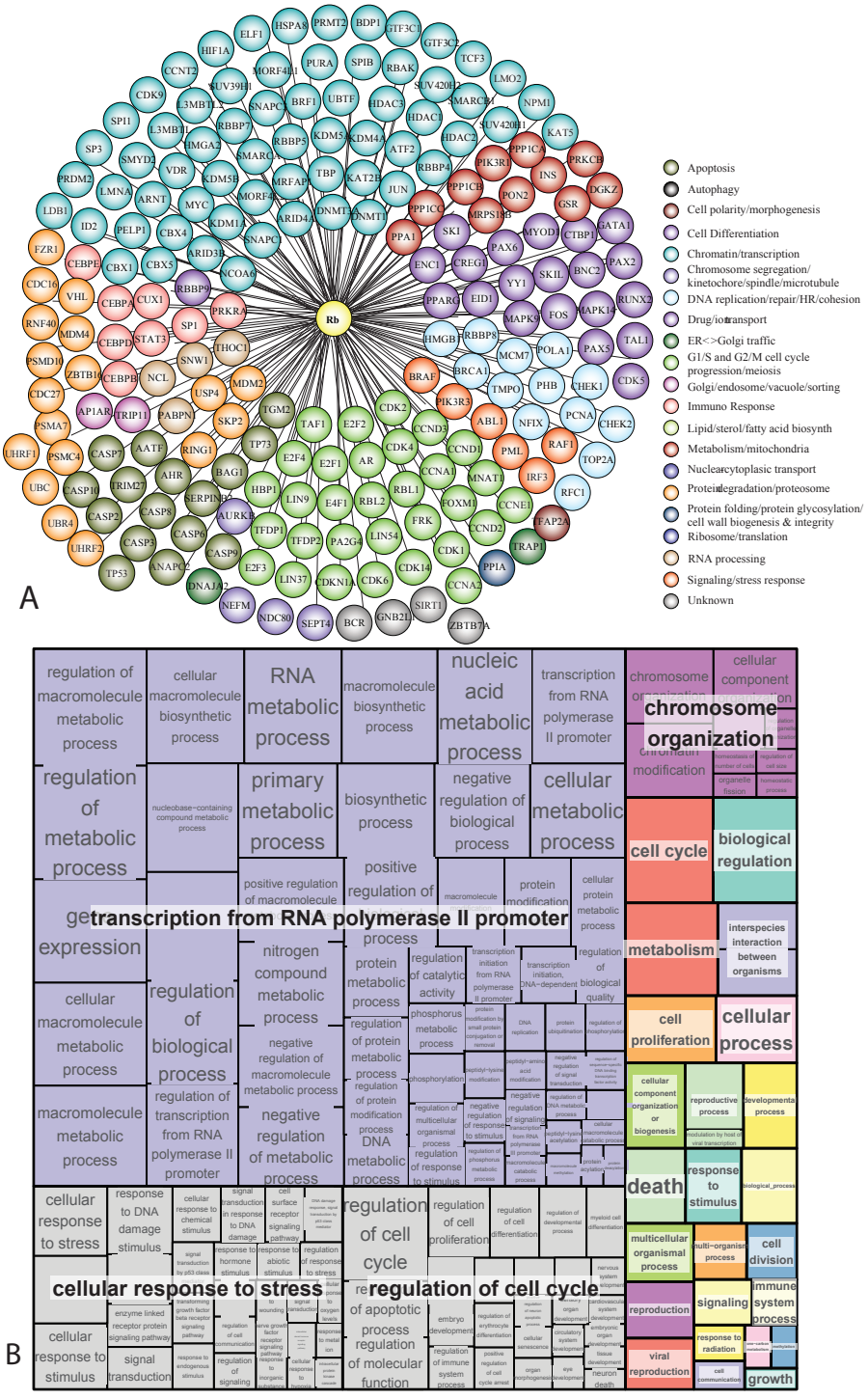


Fig. 3.5 Functional classification and GO enrichment map of physical interactors of Rb protein. A. The interaction network of Rb consists only of physical interactors, since all genetic interactors are also physical interactors. The functional classification of Rb interaction network was derived from database and literature search and color coded according to function. See text and tables for short description of the interactors. **B.** Treemap of GO terms enrichment of Rb interactors was generated by web service Revigo based on p value of GO terms enrichment of Biological Process. (see text for details)

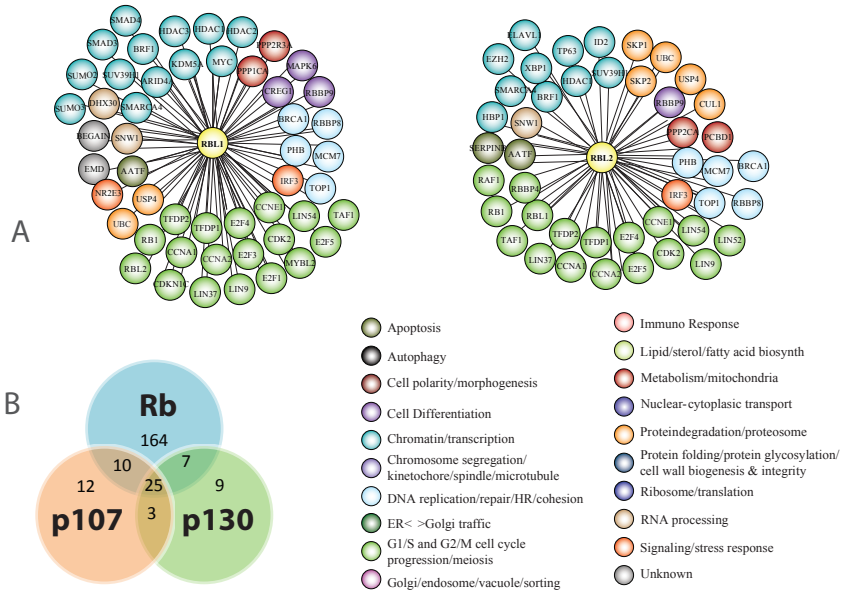
The cellular response to stress supercluster includes terms generically related to stress response as well as more specific response such as DNA damage, ions and estradiol and p53-mediated signal transduction events. The regulation of cell cycle supercluster includes regulation of apoptotic process, regulation of cell proliferation, regulation of cell differentiation etc. The chromosome organization supercluster includes mostly terms related to chromatin organization and remodeling (Fig. 3.5 B). (for a list of GO terms enrichment of Rb see Appendix B)

3.3.3.3. The p107 and p130 interactomes

Rb, p107 and p130 belong to a pocket protein family, which share common pocket domain despite their length and sequence differences (Cobrinik 2005). This conserved pocket domain serves as a binding site for numerous cellular proteins (Cobrinik 2005), although more physical interactors of Rb have been identified than its partners. Protein interactors of p107 and p130, were obtained as described in Materials and Methods and are reported in Table 5 and 6 in Appendix B. Like Rb, most notable p107 and p130 interactors are transcription factors, kinases, cyclins and histone modification enzymes.

p107 and p130 bind distinctly different set of transcription factors coupled to Rb, p107 and p130 preferably binding E2F4 and E2F5 (Litovchick *et al.* 2007) (Dyson *et al.* 1993, Hijmans *et al.* 1995). Both p107 and p130 proteins bind to DP (1-2) (TFDP1/2) (Litovchick *et al.* 2007, Wu *et al.* 1995) (Fig. 3.6 A).

Like Rb, p107 and p130 are phosphorylated by Cyclin/Cdk during cell cycle progression (Classon and Dyson 2001, Lacy and Whyte 1997, Xiao *et al.* 1996).

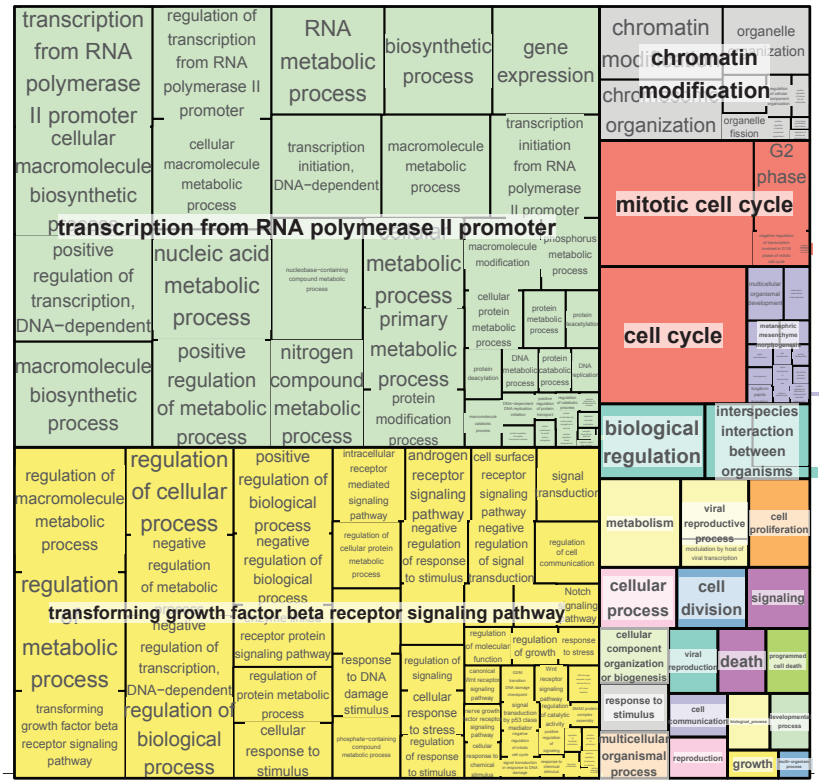


C

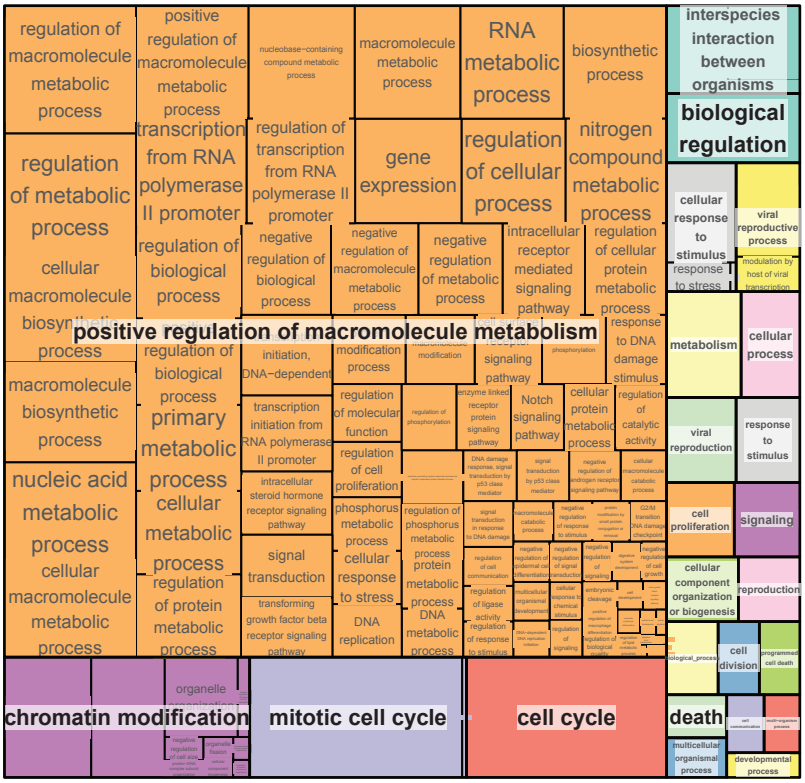
ID	Protein Name	Function
AATF	apoptosis antagonizing transcription factor	Apoptosis
BRCA1	breast cancer 1, early onset	DNA replication/repair/HR/cohesion
BRF1	BRF1 homolog, subunit of RNA polymerase III transcription initiation factor IIIB	Chromatin/transcription
CCNA1	Cyclin A1	G1/S and G2/M cell cycle progression/meiosis
CCNA2	Cyclin A2	G1/S and G2/M cell cycle progression/meiosis
CCNE1	Cyclin E1	G1/S and G2/M cell cycle progression/meiosis
CDK2	Cyclin-dependent kinase 2	G1/S and G2/M cell cycle progression/meiosis
E2F4	E2F transcription factor 4, p107/p130-binding	G1/S and G2/M cell cycle progression/meiosis
HDAC1	histone deacetylase 1	Chromatin/transcription
IRF3	interferon regulatory factor 3	signaling/stress response
LIN37	lin-37 homolog	G1/S and G2/M cell cycle progression/meiosis
LIN54	lin-54 homolog	G1/S and G2/M cell cycle progression/meiosis
LIN9	lin-9 homolog	G1/S and G2/M cell cycle progression/meiosis
MCM7	inichromosome maintenance complex component 7	DNA replication/repair/HR/cohesion
PHB	prohibitin	DNA replication/repair/HR/cohesion
RBBP8	Retinoblastoma binding protein 8	DNA replication/repair/HR/cohesion
RBBP9	Retinoblastoma binding protein 9	Cell Differentiation
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	Chromatin/transcription
SNW1	SNW domain containing 1	RNA processing
SUV39H1	Suppressor of variegation 3-9 homolog 1	Chromatin/transcription
TAF1	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 250kDa	G1/S and G2/M cell cycle progression/meiosis
TFDP1	Tanscription factor Dp-1 E2F dimerization partner 1)	G1/S and G2/M cell cycle progression/meiosis
TFDP2	Transcription factor Dp-2 (E2F dimerization partner 2)	G1/S and G2/M cell cycle progression/meiosis
UBC	Ubiquitin C	Protein degradation/proteosome
USP4	Ubiquitin specific peptidase 4 (proto-oncogene)	Protein degradation/proteosome

Fig. 3.6 Functional classification and GO enrichment map of physical interactors of Rb-like proteins. The pocket proteins (Rb, p107 and p130) share conserved pocket domain which serves as a binding site for numerous cellular proteins. The functional classification of p107 (RBL1) and p130 (RBL2) interaction network was derived from database and literature search and color coded according to function (A). See text and tables for short description of the interactors. According to BioGRID database these three proteins has some unique interactors and also shared interaction partners (B). The shared interactors are around 25 (C). Treemap of GO terms enrichment of p107 interactors (D) and p130 interactors (E) were generated by web service Revigo based on p value of GO terms enrichment of Biological Process. (see text for details)

D


transcription from RNA polymerase II promoter
regulation of transcription from RNA polymerase II promoter
RNA metabolic process
biosynthetic process
gene expression
cellular macromolecule biosynthetic process
cellular macromolecule metabolic process
transcription initiation, DNA−dependent
macromolecule metabolic process
transcription initiation from RNA polymerase II promoter
positive regulation of transcription, DNA−dependent
nucleic acid metabolic process
metabolic process
primary metabolic process
macromolecule biosynthetic process
positive regulation of metabolic process
nitrogen compound metabolic process
protein modification process
chromatin modification
chromatin organization
organelle fission
G2 phase
mitotic cell cycle
cell cycle
biological regulation
interspecies interaction between organisms
metabolism
viral reproductive process
cell proliferation
cellular process
cell division
signaling
regulation of macromolecule metabolic process
regulation of cellular process
positive regulation of biological process
intracellular receptor mediated signaling pathway
androgen receptor signaling pathway
cell surface receptor signaling pathway
signal transduction
regulation
negative regulation of metabolic process
negative regulation of biological process
negative regulation of response to stimulus
negative regulation of signal transduction
regulation of cell communication
metabolic process
transforming growth factor beta receptor signaling pathway
receptor protein signaling pathway
regulation of protein metabolic process
response to DNA damage stimulus
regulation of signaling
cellular response to stress
cellular response to stimulus
regulation of response to stimulus
Notch signaling pathway
regulation of molecular function
regulation of growth
response to stress
cellular component organization or biogenesis
viral reproduction
death
programmed cell death
response to stimulus
cell communication
developmental process
multicellular organismal process
reproduction
growth

E


positive regulation of macromolecule metabolism
regulation of macromolecule metabolic process
positive regulation of macromolecule metabolic process
nucleobase−containing compound metabolic process
macromolecule metabolic process
RNA metabolic process
biosynthetic process
interspecies interaction between organisms
biological regulation
transcription from RNA polymerase II promoter
regulation of transcription from RNA polymerase II promoter
gene expression
regulation of cellular process
nitrogen compound metabolic process
regulation of metabolic process
cellular macromolecule biosynthetic process
negative regulation of biological process
negative regulation of macromolecule metabolic process
negative regulation of metabolic process
intracellular receptor mediated signaling pathway
regulation of cellular protein metabolic process
cellular response to stimulus
viral reproductive process
response to stress
modulation by host of viral transcription
metabolism
cellular process
response to DNA damage stimulus
regulation of biological process
primary metabolic process
cellular metabolic process
nucleic acid metabolic process
cellular macromolecule metabolic process
regulation of protein metabolic process
viral reproduction
response to stimulus
cell proliferation
signaling
cellular component organization or biogenesis
reproduction
cell division
programmed cell death
death
multicellular organismal process
developmental process
chromatin modification
mitotic cell cycle
cell cycle

P107 and p130 interact with histone modification enzymes. Like Rb, p107 interacts with HDAC1, HDAC2 and HDAC3 whereas p130 interacts only with HDAC1 (Ferreira *et al.* 1998, Lai *et al.* 1999). Like Rb both proteins interact with Histone-lysine N-methyltransferase SUV39H1 (Nicolas *et al.* 2003).

These three pocket proteins (Rb, p107 and p130) share around 25 interactors among themselves (Fig. 3.6 B). Their shared interactor numbers are shown in the Venn diagram and corresponding list (Fig. 3.6 C) are highlighted. Most of the shared interactors are involved in either cell cycle progression or chromatin/transcription.

3.3.3.4. GO term enrichment of p107 and p130 interactome

GO term enrichment of p107 and p130 interactors, were obtained as described in Materials and Methods. Figure 3.6 shows a hierarchical treemap of GO terms enriched in interactors of both proteins. The plot has been generated by Revigo (default parameters) (see Materials and Methods). In case of p107, the largest supercluster (tagged as transcription from RNA polymerase II promoter) includes GO terms related to "gene expression, RNA metabolic process, macromolecule biosynthetic process, transcription initiation DNA−dependent". The second supercluster is transforming growth factor beta receptor signaling pathway includes terms generically related to signal transduction as well as more specific response such as androgen receptor signaling pathway, response to DNA damage stimulus, and regulation of cell communication events. The chromatin modification supercluster includes mostly terms related to chromatin organization and remodeling (Fig. 3.6. D).

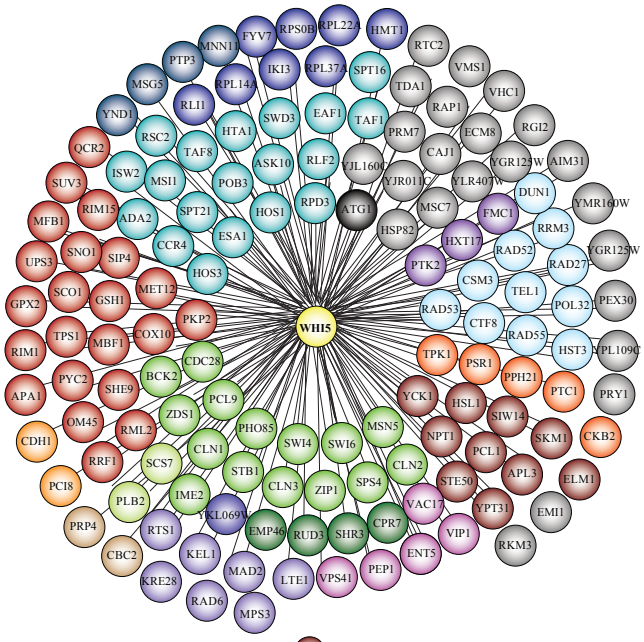
In case of p130, the largest supercluster (tagged as positive regulation of macromolecule metabolism) includes GO terms related to "regulation of metabolic process, gene expression, RNA metabolic process, regulation of biological process". The chromatin modification supercluster includes mostly terms related to chromatin organization and remodeling (Fig. 3.6E).

3.3.3.5. The Whi5 interactome

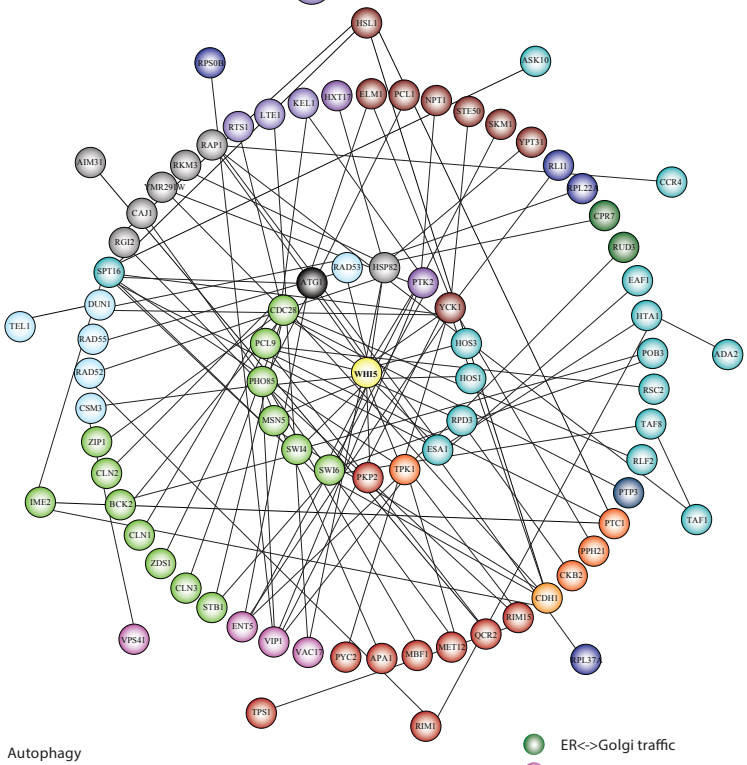
All genetic and physical protein interactors of budding yeast Whi5 were obtained as described in Materials and Methods (Appendix B Table 7). Figure 3.7A shows all interactors, using the same color-coding scheme used in. Of the 142 Whi5 interactors, only 17 are physical, the remaining having been characterized so far only as genetic interactors. With the exception of the histone deacetylase Hos 1, 3 and the proteins kinase Pkp2, all Whi5 physical interactors had one or more physical interactors among the Whi5 genetic interactors. In turn, some of these second level interactors took direct interaction with third level interactors. Thus, the Whi5 interactome is organized hierarchically (Fig. 3.7B).

As expected, physical interactors include the Swi4 and Swi6 components of the Whi5-inhibited SBF/MBF transcriptional activators involved in G1/S transition. Four more interactors are involved in chromatin remodeling, Esa1 being a histone acetylase and Hos1, Hos3, Rpd3 being histone deacetylases. Hsp82 is an ATP-dependent molecular chaperone and Msn5 is an exportin, possibly involved in regulated export of Whi5 that is required for SBF/MBF activation (Taberner *et al.* 2009). Strikingly, the nine remaining interactors are either regulatory (Pcl9) or catalytic (Atg1, Cdc28, Pho85, Pkp2, Ptk2, Rad53, Tpk1, Yck1) subunits of protein kinases.

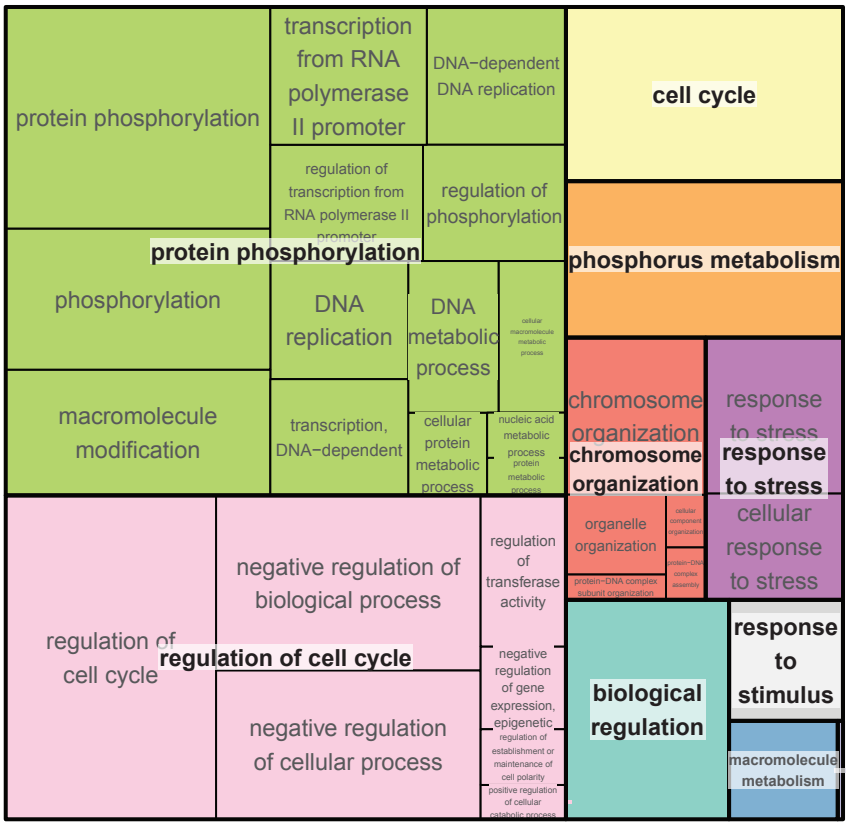
A



B



- Autophagy
- Metabolism/mitochondria
- G1/S and G2/M cell cycle progression/meiosis
- Signaling/stress response
- Cell polarity/morphogenesis
- Chromatin/transcription
- Drug/ion transport
- ER<->Golgi traffic
- Golgi/endosome/vacuole/sorting
- Chromosome segregation/ kinetochore/spindle/microtubule
- Protein degradation/proteosome
- DNA replication/repair/HR/cohesion
- Ribosome/translation
- Other/Unknown



C

Fig. 3.7 Functional classification and GO terms enrichment map of genetic and physical interactors of Whi5 protein. The interaction network of Whi5 includes both physical and genetic interactors (A). Functional classification of Whi5 interactors was derived from the classification model of Costanzo et al. The interaction network of Whi5 consists more genetic interactors than physical interactors. The first circle of Whi5 interaction network is all physical interactors. Some of the physical interactors physically interact with many genetic interactors (B). Some of the rest of genetic interactors which do not interact physically with physical interactors may physically interact with genetic interactors. Treemap of GO terms enrichment of Whi5 interactors (C) was generated by web service Revigo based on p value of GO terms enrichment of Biological Process. (see text for details)

3.3.3.6. The Whi5 interacting kinome

Phosphorylation sites are predicted by the GPS2.1 server and NetPhosYeast 1.0 Server (Ingrell *et al.* 2007) in the Whi5 protein. The putative phosphorylation sites and their cognate kinases are shown in Table 9 in Appendix B. Sites for 6 of the physically interacting protein kinases are found in the Whi5 sequence. These include kinases involved in the control of cell cycle (Cdk1 (Cdc28), Rad53 and Pho85), in regulating the cellular response to nutrient levels and environmental conditions (Pho85 and Tpk1), in septin assembly and endocytic trafficking (Yck1) and in negative regulation of activity of the mitochondrial pyruvate dehydrogenase complex (Pkp2). The prediction of these phosphorylation sites suggests, although it does not prove – that the physical interaction with the corresponding kinases may be of the enzyme-substrate type.

17 of the predicted phosphorylation sites have been experimentally verified by Wagner *et al.* (Wagner *et al.* 2009). The majority (twelve) of the phosphorylated sites are Cdk1 sites (one, S88 is not predicted as a Cdk site by GPS2.1 at high threshold, but is predicted as a phosphosite by the yeast-specific NetPhosYeast 1.0 Server), one for Rad53, a protein kinase, required for cell-cycle arrest in response to DNA damage (Sanchez *et al.* 1999). Rad53 plays a role in initiation of DNA replication and prevents collapse of replication fork when replication block (Lopes *et al.* 2001). Seven for Yck1, one for Pkp2 and two for Tpk1 predicted by GPS 2.1 are also experimentally detected phosphosites and the phosphorylation sites are predicted based human homolog of yeast kinases (see Table 9 in Appendix B).

3.3.3.7. GO term enrichment of Whi5 interactome

Figure 3.7C shows a hierarchical "treemap" of GO terms enriched in Whi5 interactors (genetic plus physical). The largest supercluster (tagged

as regulation of cell cycle) includes terms related to "regulation of cell cycle" and "regulation of macromolecular biosyntheses and gene expression". The presence in this supercluster of terms related to "regulation of metabolic and primary metabolic processes" is noteworthy. Together with the enrichment of terms related to "response to stress and stimuli and phosphorus metabolism", analysis of the treemap suggests a tight integration of Whi5 with internal and external signals, whose integration is required for proper regulation of the G1/S transition. The second largest supercluster - tagged as protein phosphorylation - includes terms related to protein phosphorylation per se, as well as terms related to DNA replication and related processes. (for a list of GO terms enrichment of Whi5 see Appendix B)

CHAPTER

4

Conclusions and Discussion

The G1/S transition and the initiation of DNA replication are major regulatory events in the eukaryotic cell cycle. The complex machinery involved in DNA replication has been reviewed in Chapter 2 and will not be further addressed in this Discussion, that will be limited to highlighting methodological approaches and major results obtained by comparative analysis of Rb and Whi5.

4.1 Conservation of Whi5 phosphorylation sites

As it has been shown in chapter 3, Whi5 is a highly disordered protein. Within the Whi5 sequence several phosphorylation sites have been identified. Interestingly many of them mapped to specific motifs that characterized the family of Whi5 homologs in *Saccharomycetales*. Several proteins kinases are direct physical interactors of Whi5. Consistently many phosphorylation sites have been identified in Whi5 (de Bruin *et al.* 2004, Wagner *et al.* 2009). Indeed, 12 out of 18 *in-vivo* phosphorylation sites found in asynchronously growing cells can be ascribed to Cdk1 kinase (Wagner *et al.* 2009). Cells bearing a deletion of *WHI5* are smaller than wild type (Costanzo *et al.* 2004). Expression of Whi5-12 Ala (all the Cdk putative serine consensus sites are mutated to alanine) rescues the small-cell phenotype of *whi5Δ* cells as well as wild-type Whi5 does. However overexpression of Whi5-12Ala in cells that harbor Swi6-SA4, a mutant lacking four Cdk sites (Sidorova *et al.* 1995), results in a severe growth defect (Wagner *et al.* 2009). Mutational analysis indicated that the four C-terminal Cdk sites 8, 9, 10, and 12 (Ser154, Ser156, Ser161 and Ser262) are specifically required for Whi5 inactivation. Swi6-SA4 cells co-expressing at physiologic levels a Whi5 mutant where sites 8-10 and 12 are changed to alanine (Whi5-4A) were 40% larger than wild type, indicating that phosphorylation of either Swi6 or Whi5 is necessary for regulation of cell size (Wagner *et al.* 2009). The cells are unviable when the two proteins are overexpressed from the *GAL1* promoter (Wagner *et*

al. 2009). Since retaining four or more N-terminal Cdk sites intact was not sufficient to prevent lethality, it has been concluded from Wagner *et al.*, (2009) that specific Cdk sites of Whi5 are required for inactivation, rather than simply a threshold number. The prediction of phosphorylation sites for all Whi5 homologues was carried out by the program PPSP (Prediction of PK-specific Phosphorylation site) (Xue *et al.* 2006), since the phosphorylation map predicted by PPSP for Whi5 was coincident with the experimental one (Wagner *et al.* 2009). A probability score was then assigned to each phosphorylation site by applying the algorithm GPS 2.1.

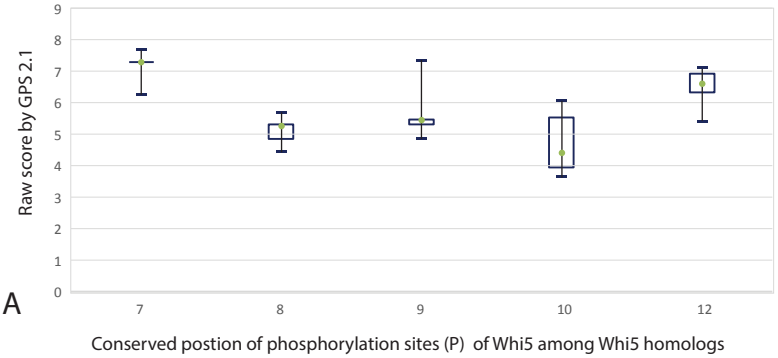
Although most Cdk1 consensus sites maps at the N-terminus of Whi5 (Figure 3.4 B), we would like to underscore that the four Whi5 phosphosites (8, 9, 10, 12) found to be relevant for the coordination of cell size with cycle progression (Wagner *et al.* 2009), are distributed along the conserved, disordered motifs 1 and 3. Moreover, according to our sequence analysis, also the Thr143 (site 7 of Whi5) appears highly conserved. By comparing motifs boundaries and phosphorylation sites of Whi5 homologous, a consensus map of conserved Cdk-phosphorylatable residues is proposed, including phosphosites 7-10 in the motif 1 and phosphosite 12 in the motif 3 (numbering of Whi5).As already outlined, motif 2 does not contain any phosphorylatable site, while the 11th phosphosite of Whi5 (T215) is included in an "inter-motif" sequence between motifs 2 and 3, whose expansion is associated to the presence of a variable number of phosphorylatable sites (from e.g. 1 in *S. cerevisiae*, to e.g. 8 in *T. blattae*) (Appendix B). In just two Whi5 homologs with expanded "inter-motif" sequence (i.e. from *A. gossypii* and *K. naganishii*), the motif 3 does not contain any Cdk-phosphorylatable site. The phosphosite 7 appears as one of the most conserved, but no correspondence was found in terms of sequence conservation with phosphosite 11, although phosphosites 7 and 11 have been suggested as

structural elements involved in the nuclear export of Whi5 in *S. cerevisiae* (Wagner *et al.* 2009).

In summary, it is not known whether specific Cdk sites or a threshold amount of phosphorylation determines Whi5 inactivation and dissociation from SBF but according to our sequence analysis, the most conserved phosphorylatable residues in the C-terminus of Whi5 (7-10 and 12) occur in structurally disordered regions, as also the four mutated Cdk sites of Swi6-SA4 (Appendix B). This observation makes conceivable an inactivation mechanism achieved by phosphorylating a threshold number of Cdk sites belonging to an overall disordered, trans-modular domain formed by the disordered motifs of both Whi5 and Swi6. Our model can be intended as an extreme case of allovalency (Borg *et al.* 2007, Klein *et al.* 2003, Mittag *et al.* 2008), where the phosphorylated sites participating in the SBF activation can equally come either from different regions of the same flexible molecule, or from different proteins/interactors.

The analysis of consensus of phosphorylation sites 7, 8, 9, 10 and 12 indicate they can be targeted by Cdk1 with a different, conserved probability (Fig. 4.1) due to the fact they do not perfectly match the canonical phosphorylation sequence. We infer that phosphorylation might occur hierarchically, sites with highest probability being phosphorylated first. Differences in phosphorylation kinetics might overall provide a kind of "phosphorylation rhythm" along the most phosphorylated region of the protein. In particular, we hypothesize that sites 7 and 12 (Whi5 numbering), when present, are the most probably and hence early phosphorylated, whereas sites 8, 9 and 10 are secondarily targeted (Fig. 4.1). Notably, phosphorylation sites among different yeast species are conserved not only in their reciprocal position, but also in their similarity to the consensus and hence their probability to be phosphorylated. In other terms, the "phosphorylation rhythm" within the motifs 1 and 3 appears strongly conserved, even when the

whole amino acid sequence is very poorly conserved. A similar analysis on the remaining phosphorylatable sites is hampered by the difficulty to recognize each site in a context of very poor sequence conservation. On the other hand, it is conceivable that the unusual conservation of primary structure in disordered, phosphorylated region of motif 1 and 3 could guarantee such a hierarchical phosphorylation. It can be observed that such a hierarchical phosphorylation mechanism is not necessarily antithetic with the hypothesis of trans modular allovalency, applicable to the complete set of P sites coming from different interaction partners. It has to be observed that Pfam includes a much larger number of Whi5-family proteins, classified according to the presence of the motif 2 alone, which is the protein family signature. We believe that the overall conservation of structural disorder, charge segregation and "phosphorylation rhythm" in motifs 1 and 3 are clues witnessing their functional importance, probably reflecting the structural constrains imposed by Whi5 interactors. On the other side, it is conceivable that members of Whi5 family that do not conserve motif 1 and motif 3, don't have to satisfy such structural constrains and might encounter structurally different binding partners. To understand the evolutionary process encountered by Whi5 and its molecular device it would be important/necessary to know how much conserved are the homologous physical interactors of Whi5 along the phylogenetic tree and when they start to diverge. This question concerning the "missing link" between the *Saccharomycetales* Whi5 homologues and the first evolutive occurrence of Rb will need the sequencing and also the structural characterization of many proteins among putative Whi5 interactors.



A

	7	8	9	10	12
S. cerevisiae	7.29	4.98	4.88	5.77	6.82
K africana	7.29	5.34	5.46	4.82	6.55
K. lactis	7.30	4.45		5.17	5.39
T. delbrueckii	7.29	5.18	5.32	4.22	6.21
C. glabrata	7.29	5.56	5.07	5.53	6.92
N. dairenensis	7.68	4.47	5.32	3.94	6.37
Z. rouxii	7.29	4.96		4.41	6.93
V. polyspora	7.29	5.34	5.46	3.65	7.12
L. thermot	6.97	5.34	5.46	3.65	6.58
N.castellii	7.68	5.34	5.46	4.41	7.12
T. phaffii	7.29	5.55	7.33	3.65	5.84
T. blattae	6.26	4.58		6.08	6.63
A. gossypii	7.29	5.68	5.44	6.08	
K. naganishii	7.29	4.77			

B

Fig. 4.1 Conservation of phosphorylation rhythm among Whi5 homologs in Fungi (A) The plot represents the probability of phosphorylation by CDK foremost conserved, C-terminal phosphorylatable sites of Whi5 homologs. The numbering P7-P12 refers to the phosphorylation map of *S. cerevisiae*. The probability was expressed in arbitrary units and calculated as average of scores assigned to a given position (P) by the algorithm GPS 2.1,when a medium threshold and a cutoff of 2.781 was set. (B) The scheme represents the occurrence and the probability score of conserved phosphorylatable sites P7-P12 in Whi5 homologs. Four P sites belonging to motif 1 are marked in orange gradations, P site 12, belonging to motif 3, is marked in blue.To note, P11 doesn't belong to any conserved motif, and in most of the analyzed sequences,its region is expanded to host up to 8 phosphorylatable sites (e.g.*T. blattae*).

4.2 Conserved motifs 1 and 3 may act as phosphorylation-dependent seeds in Whi5 folding/unfolding

As charge segregation has been recognized as a major determinant of structural compactness of disordered proteins like Whi5/Whi5 homologous, the heterogeneity of pIs calculated for different motifs of Whi5/Whi5 homologous reflects the heterogeneity of their primary structures. It can be hypothesized that conserved segregation of positive charges on motif 1 and negative ones on motif 3 has a functional role (for pI values see Table 2 in Appendix B).

We can argue that charge complementarity of disordered motifs might be functionally related to the structural compactness of the protein, a property tunable by phosphorylation. For instance, multisite phosphorylation of basic motif 1 could contribute to the functional unfolding (Uversky 2011) of Whi5 by attenuating the intramolecular interaction with the acidic motif 3. Consistently, peptides corresponding to motifs 1 and 3 are in fact able to interact in a BIAcore assay (Hasan and Vanoni, unpublished data) and phosphorylation of Cdk1 phosphosites within motif 1 destroys such interaction. Further experiments are under way to confirm and extend these results.

4.3 Towards an expanded model for Whi5 function

Although Whi5 and Rb/pocket proteins are evolutionary unrelated, a common core of functions can be identified when their interactors are examined. Although comparative interactomics needs to be taken with care, in order to avoid comparison of apples with pears (Kiemer and Cesareni 2007) the treemaps of GO terms (biological process) enriched within interactors common among Rb, p107 and p130 on one side, and GO terms (biological process) enriched within interactors common between Rb and Whi5 are strikingly similar (Figure 4.2.A and B, respectively).

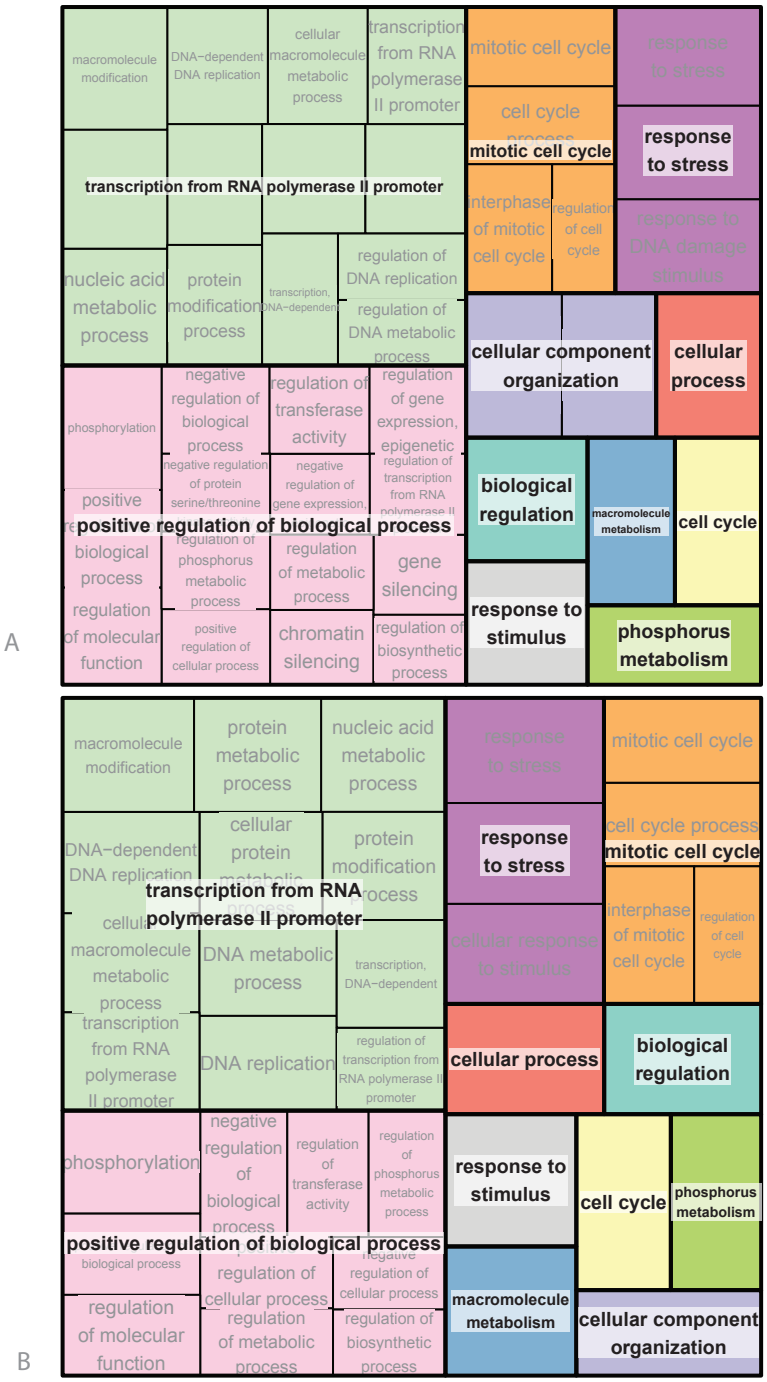


Fig 4.2 Treemaps of common GO terms enrichment of Whi5 and Rb interactors (A) and common GO terms enrichment of Whi5 and p107/p130 interactors (B) were generated by web service Revigo based on uniqueness of GO terms of Biological Process.

Interactomic data presented in Chapter 3 suggest that various pathways may be regulated by – or impinge upon - Whi5 function. On the contrary, in current models of yeast cell cycle (Barberis *et al.* 2007, Kaizu *et al.* 2010) a very limited subset of the Whi5 protein interactors are present or linked to Whi5. In order to improve our understanding of Whi5 function we present a revised, *work-in-progress* model of Whi5 function by step-wise incorporation of Whi5-interactors (Fig. 4.3). For sake of clarity we divided Whi5-related functions in four major blocks.

4.3.1 Synthesis and transport of Whi5

Following transcription of the *WHI5* gene, the Whi5 protein is translated in the cytoplasm. Transport in and out of the nucleus play a major role in controlling the function of Whi5. Whi5 nuclear import is mediated by the classical nuclear import pathway that comprises Kap95 (importin β1) and Kap60 (importin α) and recognizes the NLS which is present in N terminal portion of Whi5 sequences. The nuclear export of Whi5 requires a Nuclear Export Sequence (NES) that has been localized within amino acids 51-167. Nuclear export is mediated by the karyopherin Msn5 and regulated by phosphorylation of Ser154, Ser156 and Ser161 withinin the NES (Taberner *et al.* 2009). Recently a correlation between Whi5 translocation and activation of START has been shown using a live-cell video microscopy approach: at least 50% of Whi5 needs to exit the nucleus in order to commit cells to exit G1 and initiate a new cell cycle (Doncic *et al.* 2011).

4.3.2 Gene expression silencing

The Whi5 physical interactors Swi4 and Swi6 form the SBF complex on SCB element of promoters of SBF target genes in which the consensus sequence of SCB element is recognized by Swi6 (Koch *et al.* 1996). In late M/early G1, SBF recruits multiple components to promoters: Whi5 is recruited through interaction with Swi4, whereas Spt16 and Pob3 (i.e.,

the FACT complex (Wittmeyer *et al.* 1999) (Costanzo *et al.* 2003) involved in regulation and timing of transcription of SBF/MBF target genes) is recruited through interaction with Swi6 (Takahata *et al.* 2009). Stb1 and Whi5 both help to recruit Rpd3 (L)-a lysine deacetylases (KDAC) (Takahata *et al.* 2009). Other KDACs like HOS1 and HOS3 are also recruited to reorganize the chromatin thus inhibiting gene expression.

4.3.3 Protein modification-Protein folding

As mentioned above, protein kinases are the most abundant class of Whi5 physical interactors. Computational and experimental evidence indicates that cognate sites are present within the Whi5 sequence and are actually phosphorylated *in vivo*. For this reason we put Whi5 as substrate for these kinases in our model (fig. 4.3). When yeast cells pass through START in the cell cycle, Cnl3-Cdk1, first and then Cln1,2-Cdk1 kinase builds-up and phosphorylation of Whi5 – and possibly of some of it partners such as Swi4 , see 4.2 above- removes inhibition of transcription and leads to the G1/S transition. Interestingly, the interactions between Whi5 and KDACs, is interrupted by Cln3-Cdc28 and Pcl9-Pho85-dependent phosphorylation, leading to activation of transcription of a number of genes essential for G1/S transition including the *CLN1* and *CLN2*. Eventually these two cyclines bind with Cdc28 and further phosphorylate Whi5, promoting its dissociation from SBF and its nuclear export (Huang *et al.* 2009). Phosphorylated Stb1 may remain at the promoter and stimulate gene activation (Takahata *et al.* 2009) (Figure 4.3).

While the role of Cln1,2,3-Cdk1 kinase complexes in regulating Whi5 function and subcellular localization is known (section 3.3.3.6.), the role – if any – played by phosphorylation by the other kinases remains to be evaluated. Subcellular localization of these kinases was obtained from

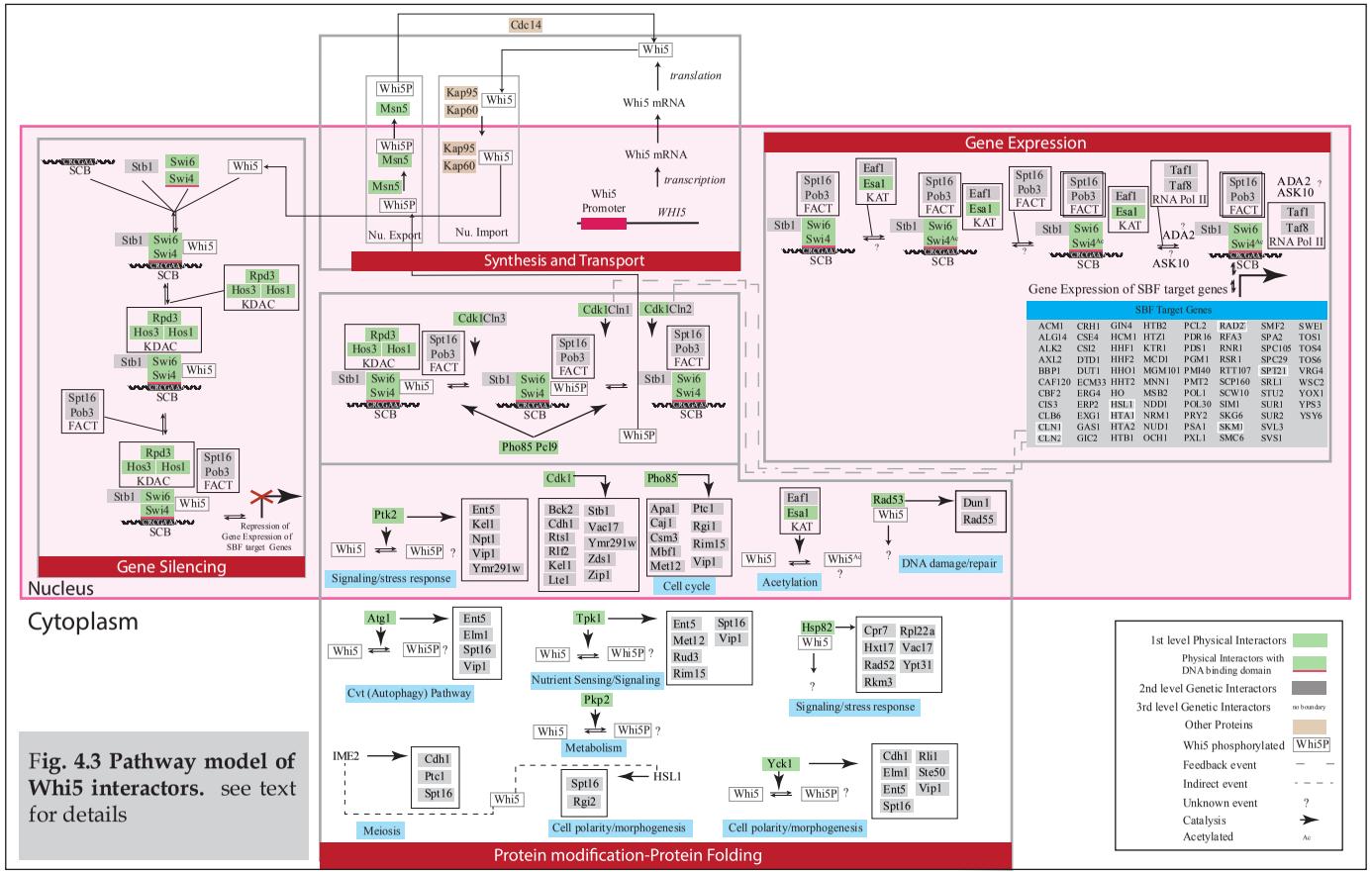


Fig. 4.3 Pathway model of Whi5 interactors. see text for details

YPL+ Database (Oskolkova, Leitner and Kohlwein, personal communication). Regulated phosphorylation by these kinases may link the functional state of Whi5 to different stimuli and/or cell fates. For instance Tpk1, one of the catalytic subunits of cAMP-dependent protein kinase, Ptk2 and Pho89/Pcl9 may link Whi5 to sensing of different nutrient such as carbon, nitrogen and inorganic phosphate. Some of the other kinases that phosphorylate Whi5 or its interactors may contribute to defining yeast cell fate. These kinases include Agt1 (autophagy), Yck1 and Hsl1 (cell polarity/morphogenesis), Ime2 (meiosis), Rad53 ((DNA damage-repair), while the KAT acetylating enzyme might convey information regarding the metabolic state. In the concept map reported in Figure 4.3. These pathways are not drawn in full, but the pathway is highlighted in light blue, simply to indicate which functional information these events (mostly phosphorylations) may convey to Whi5.

4.3.4 Gene Expression

As stated above, Whi5 inhibits SBF target gene expression. Upon Cdk1-mediated phosphorylation and nuclear exclusion of Whi5, promoters of SBF target gene recruit additional proteins that promote gene expression. These include Esa1 which is catalytic subunit of NuA4 histone acetyltransferase (KAT), involved in acetylation of histone proteins and other proteins and help to promote cell cycle progression. This protein also acetylates Whi5 and Swi4 proteins (Lin *et al.* 2009) though function of this acetylation remains unknown. Eaf1 - a Whi5 genetic interactor - acts as a platform for assembly of NuA4 subunits into the native complex. Some SBF target genes - such as *SKM1*, *RAD27*, *CLN2*, *CLN1*, *HTA1*, *HSL1*, *SPT21* (Fig. 4.2) - encode genetic interactors of Whi5. Notably, Cln1 and Cln2 act in a positive feedback loop to further phosphorylate Whi5 and increase gene expression. Recently a fine dissection of temporal expression of G1/S genes including SBF

target genes has been obtained (Eser *et al.*, 2011). Genes within the SBF/MBF regulon have a defined distribution of transcriptional activation times. Early transcription of CLN1 and CLN2 precedes the activation, providing early commitment to a new cell cycle. Conversely, late transcription of NRM1, helps to turn off MBF target genes (Eser *et al.* 2011).

4.4 Final remarks

According to molecular systems biology, biological processes are the result of complex, coordinated, dynamical, non-linear interactions that generate the corresponding function as an emergent property of the system, that therefore is not found in individual components, but only in their networking (Bhalla and Iyengar 1999). Conversely, it has been shown that even a single, multi-domain protein may present a system-level behavior that can be described by adapting the formalism used to describe inter-molecular networks (Sacco *et al.* 2012). Accordingly, a multi-scale approach is required to fully understand any given biological function (Alberghina *et al.* 2012, Kitano 2010).

In the second part of my PhD Thesis I concentrated on the protein Whi5, a regulator of G1/S-specific transcription in budding yeast and compared its structural properties with those of Rb – and members of the pocket family - that perform equivalent function(s) in mammalian cells.

Rb and Whi5 do not share any sequence identity, but both are disordered proteins. In Rb, that is a much larger protein than Whi5, structured domains alternate with disordered regions, whereas in Whi5 only a single structured domain is likely to exist.

- The family of Whi5-related proteins is only present in *Saccharomycetales*. The higher-than-expected conservation of

sequence in disordered regions, correlates with abundance of phosphorylation sites and allows to predict conservation of a similar "phosphorylation rhythm" with a potentially relevant functional role. Notably, conserved motifs 1 and 3 may act as phosphorylation-dependent seeds in Whi5 folding/unfolding.

- The (partial) disordered nature of both proteins allows them to act as protein hubs, able to interact with many partners. Interestingly, the smaller Whi5 acts as a hierarchical hub.
- Comparison of Rb and Whi5 interactors (both physical and genetical ones) allows to highlight a significant core of conserved common functionalities associated with the interactors indicating that network structure and function – rather than individual proteins, are conserved during evolution (Cross *et al.* 2011).
- By step-wise adding interactors to existing models for Whi5 function, an improved "concept map" of Whi5 function is being constructed. Such a map will allow to construct dynamic mathematical model(s) of increasing granularity and to design experiments to proof novel regulatory links within the Whi5 network.

References

Chapter 1

Alberghina L, Westerhoff HV. Systems Biology. Definitions and Perspectives. Top Curr Genet, 2005; 13.

Arias EE, Walter JC. Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. Genes Dev. 2007;21:497-518.

Bell SP, Dutta A. DNA replication in eukaryotic cells. Annu Rev Biochem. 2002;71:333-74.

Chicas A, Wang X, Zhang C, McCurrach M, Zhao Z, Mert O, et al. Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. Cancer Cell. 2010;17:376-87.

Cooper K. Rb, whi it's not just for metazoans anymore. Oncogene. 2006;25:5228-32.

Diffley JF. Regulation of early events in chromosome replication. Curr Biol. 2004;14:R778-86.

Groth A, Rocha W, Verreault A, Almouzni G. Chromatin challenges during DNA replication and repair. Cell. 2007;128:721-33.

Jorgensen P, Nishikawa JL, Breitkreutz BJ, Tyers M. Systematic identification of pathways that couple cell growth and division in yeast. Science. 2002;297:395-400.

Kiemer L, Cesareni G. Comparative interactomics: comparing apples and pears? Trends Biotechnol. 2007;25:448-54.

Kitano H. Systems biology: a brief overview. Science. 2002;295:1662-4.

Lin J, Qian J. Systems biology approach to integrative comparative genomics. Expert Rev Proteomics. 2007;4:107-19.

Liu J, Kipreos ET. Evolution of cyclin-dependent kinases (CDKs) and CDK-activating kinases (CAKs): differential conservation of CAKs in yeast and metazoa. Mol Biol Evol. 2000;17:1061-74.

Loyer P, Trembley JH, Katona R, Kidd VJ, Lahti JM. Role of CDK/cyclin complexes in transcription and RNA splicing. Cell Signal. 2005;17:1033-51.

Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer. 2009;9:153-66.

Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N, Oda M. Eukaryotic chromosome DNA replication: where, when, and how? Annu Rev Biochem. 2010;79:89-130.

Morgan DO. Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol. 1997;13:261-91.

Morgan DO. The Cell Cycle: Principles of Control London: New Science Press. 2007

Onn I, Heidinger-Pauli JM, Guacci V, Unal E, Koshland DE. Sister chromatid cohesion: a simple concept with a complex reality. Annu Rev Cell Dev Biol. 2008;24:105-29.

Pavletich NP. Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. J Mol Biol. 1999;287:821-8.

Peters JM. The anaphase promoting complex/cyclosome: a machine designed to destroy. Nat Rev Mol Cell Biol. 2006;7:644-56.

Phatnani HP, Greenleaf AL. Phosphorylation and functions of the RNA polymerase II CTD. Genes Dev. 2006;20:2922-36.

Wittenberg C, Reed SI. Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. Oncogene. 2005;24:2746-55.

Zhang J, Schneider C, Ottmers L, Rodriguez R, Day A, Markwardt J, et al. Genomic scale mutant hunt identifies cell size homeostasis genes in S. cerevisiae. Curr Biol. 2002;12:1992-2001.

Abdurashidova G, Radulescu S, Sandoval O, Zahariev S, Danailov MB, Demidovich A, et al. Functional interactions of DNA topoisomerases with a human replication origin. EMBO J 2007;26:998-1009.

Abe T, Yoshimura A, Hosono Y, Tada S, Seki M, Enomoto T. The N-terminal region of RECQL4 lacking the helicase domain is both essential and sufficient for the viability of vertebrate cells. Role of the N-terminal region of RECQL4 in cells. Biochim Biophys Acta 2011;1813:473–9.

Aggarwal P, Vaites LP, Kim JK, Mellert H, Gurung B, Nakagawa H, et al. Nuclear cyclin D1/CDK4 kinase regulates CUL4 expression and triggers neoplastic growth via activation of the PRMT5 methyltransferase. Cancer Cell 2010;18:329–40.

Alberghina L, Hofer T, Vanoni M. Molecular networks and system-level properties. J Biotechnol 2009;144:224–33.

Aparicio OM, Stout AM, Bell SP. Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication. Proc Natl Acad Sci U S A 1999;96: 9130–5.

Aparicio T, Ibarra A, Mendez J. Cdc45-MCM-GINS, a new power player for DNA replication. Cell Div 2006;1:18.

Aparicio T, Guillou E, Coloma J, Montoya G, Mendez J. The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. Nucleic Acids Res 2009;37:2087–95.

Araki H. Cyclin-dependent kinase-dependent initiation of chromosomal DNA replication. Curr Opin Cell Biol 2010a;22:766–71.

Araki H. Regulatory mechanism of the initiation step of DNA replication by CDK in budding yeast. Biochim Biophys Acta 2010b;1804:520–3.

Araki H, Leem SH, Phongdara A, Sugino A. Dpb11, which interacts with DNA polymerase II (epsilon) in *Saccharomyces cerevisiae*, has a dual role in S-phase progression and at a cell cycle checkpoint. Proc Natl Acad Sci U S A 1995;92:11791–5.

Arias EE, Walter JC. Replication-dependent destruction of Cdt1 limits DNA replication to a single round per cell cycle in *Xenopus* egg extracts. Genes Dev 2005;19:114–26.

Arias EE, Walter JC. PCNA functions as a molecular platform to trigger Cdt1 destruction and prevent re-replication. Nat Cell Biol 2006;8:84–90.

Arias EE, Walter JC. Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. Genes Dev 2007;21:497–518.

Balestrini A, Cosentino C, Errico A, Garner E, Costanzo V. GEMC1 is a TopBP1-interacting protein required for chromosomal DNA replication. Nat Cell Biol 2010;12:484–91.

Ballabeni A, Zamponi R, Caprara G, Melixetian M, Bossi S, Masiero L, et al. Human CDT1 associates with CDC7 and recruits CDC45 to chromatin during S phase. J Biol Chem 2009;284:3028–36.

Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP. A probability-based approach for highthroughput protein phosphorylation analysis and site localization. Nat Biotechnol 2006;24:1285–92.

Bell SP, Dutta A. DNA replication in eukaryotic cells. Annu Rev Biochem 2002;71: 333–74.

Bell SP, Stillman B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature 1992;357:128–34.

Bennetzen MV, Larsen DH, Bunkenborg J, Bartek J, Lukas J, Andersen JS. Site-specific phosphorylation dynamics of the nuclear proteome during the DNA damage response. Mol Cell Proteomics 2010;9:1314–23.

Berezney R, Dubey DD, Huberman JA. Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. Chromosoma 2000;108:471–84.

Bermudez VP, Farina A, Tappin I, Hurwitz J. Influence of the human cohesion establishment factor Ctf4/AND-1 on DNA replication. J Biol Chem 2010;285:9493–505.

Bermudez VP, Farina A, Raghavan V, Tappin I, Hurwitz J. Studies on human DNA polymerase epsilon and GINS complex and their role in DNA replication. J Biol Chem 2011;286:28963–77.

Blow JJ, Dutta A. Preventing re-replication of chromosomal DNA. Nat Rev Mol Cell Biol 2005;6:476–86.

Bochman ML, Schwacha A. The Mcm2–7 complex has in vitro helicase activity. Mol Cell 2008;31:287–93.

Bochman ML, Schwacha A. The Mcm complex: unwinding the mechanism of a replicative helicase. Microbiol Mol Biol Rev 2009;73:652–83.

Bork P, Hofmann K, Bucher P, Neuwald AF, Altschul SF, Koonin EV. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. FASEB J 1997;11:68–76.

Borlado LR, Mendez J. CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. Carcinogenesis 2008;29:237–43.

Bowers JL, Randell JC, Chen S, Bell SP. ATP hydrolysis by ORC catalyzes reiterative Mcm2–7 assembly at a defined origin of replication. Mol Cell 2004;16:967–78.

Bracken AP, Ciro M, Cocito A, Helin K. E2F target genes: unraveling the biology. Trends Biochem Sci 2004;29:409–17.

Brill LM, Xiong W, Lee KB, Ficarro SB, Crain A, Xu Y, et al. Phosphoproteomic analysis of human embryonic stem cells. Cell Stem Cell 2009;5:204–13.

Broderick R, Nasheuer HP. Regulation of Cdc45 in the cell cycle and after DNA damage. Biochem Soc Trans 2009;37:926–30.

Bruck I, Kaplan DL. GINS and Sld3 compete with one another for Mcm2–7 and Cdc45 binding. J Biol Chem 2011;286:14157–67.

Brummer A, Salazar C, Zinzalla V, Alberghina L, Hofer T. Mathematical modelling of DNA replication reveals a trade-off between coherence of origin activation and robustness against rereplication. PLoS Comput Biol 2010;6:e1000783.

Brush GS, Kelly TJ. Phosphorylation of the replication protein A large subunit in the *Saccharomyces cerevisiae* checkpoint response. Nucleic Acids Res 2000;28: 3725–32.

Burgers PM. Polymerase dynamics at the eukaryotic DNA replication fork. J Biol Chem 2009;284:4041–5.

Cadoret JC, Meisch F, Hassan-Zadeh V, Luyten I, Guillet C, Duret L, et al. Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc Natl Acad Sci U S A 2008;105:15837–42.

Caldecott KW. Cell signaling. The BRCT domain: signaling with friends? Science 2003;302:579–80.

Callan HG. Replication of DNA in eukaryotic chromosomes. Br Med Bull 1973;29:192–5.

Callebaut I, Mornon JP. From BRCA1 to RAP1: a widespread BRCT module closely associated with DNA repair. FEBS Lett 1997;400:25–30.

Calzada A, Sacristan M, Sanchez E, Bueno A. Cdc6 cooperates with Sic1 and Hct1 to inactivate mitotic cyclin-dependent kinases. Nature 2001;412:355–8.

Cantin GT, Yi W, Lu B, Park SK, Xu T, Lee JD, et al. Combining protein-based IMAC, peptidebased IMAC, and MudPIT for efficient phosphoproteomic analysis. J Proteome Res 2008;7:1346–51.

Charvin G, Oikonomou C, Siggia ED, Cross FR. Origin of irreversibility of cell cycle start in budding yeast. PLoS Biol 2010;8:e1000284.

Chattopadhyay S, Bielinsky AK. Human Mcm10 regulates the catalytic subunit of DNA polymerase-alpha and prevents DNA damage during replication. Mol Biol Cell 2007;18:4085–95.

Chen S, Bell SP. CDK prevents Mcm2–7 helicase loading by inhibiting Cdt1 interaction with Orc6. Genes Dev 2011;25:363–72.

Chen Z, Speck C, Wendel P, Tang C, Stillman B, Li H. The architecture of the DNA replication origin recognition complex in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 2008;105:10326–31.

Chowdhury A, Liu G, Kemp M, Chen X, Katrangi N, Myers S, et al. The DNA unwinding element binding protein DUE-B interacts with Cdc45 in preinitiation complex formation. Mol Cell Biol 2010;30:1495–507.

Chuang LC, Teixeira LK, Wohlschlegel JA, Henze M, Yates JR, Mendez J, et al. Phosphorylation of Mcm2 by Cdc7 promotes pre-replication complex assembly during cellcycle re-entry. Mol Cell 2009;35:206–16.

Chuang CH, Wallace MD, Abratte C, Southard T, Schimenti JC. Incremental genetic perturbations to MCM2–7 expression and subcellular distribution reveal exquisite sensitivity of mice to DNA replication stress. PLoS Genet 2010;6:e1001110.

Clay-Farrace L, Pelizon C, Santamaria D, Pines J, Laskey RA. Human replication protein Cdc6 prevents mitosis through a checkpoint mechanism that implicates Chk1. EMBO J 2003;22:704–12.

Cook JG. Replication licensing and the DNA damage checkpoint. Front Biosci 2009;14: 5013–30.

Cook JG, Chasse DA, Nevins JR. The regulated association of Cdt1 with minichromosome maintenance proteins and Cdc6 in mammalian cells. J Biol Chem 2004;279: 9625–33.

Costa A, Ilves I, Tamberg N, Petojevic T, Nogales E, Botchan MR, et al. The structural basis for MCM2–7 helicase activation by GINS and Cdc45. Nat Struct Mol Biol 2010;18:471–7.

Costanzo M, Nishikawa JL, Tang X, Millman JS, Schub O, Breitkreuz K, et al. CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. Cell 2004;117: 899–913.

Cotterill S, Kearsey SE. DNAReplication: a database of information and resources for the eukaryotic DNA replication community. Nucleic Acids Res 2009;37:D837–9.

de Bruin RA, McDonald WH, Kalashnikova TI, Yates III J, Wittenberg C. Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. Cell 2004;117:887–98.

De Falco M, Ferrari E, De Felice M, Rossi M, Hubscher U, Pisani FM. The human GINS complex binds to and specifically stimulates human DNA polymerase alpha-primase. EMBO Rep 2007;8:99-103.

De Marco V, Gillespie PJ, Li A, Karantzelis N, Christodoulou E, Klompmaker R, et al. Quaternary structure of the human Cdt1-Geminin complex regulates DNA replication licensing. Proc Natl Acad Sci U S A 2009;106:19807–12.

DePamphilis ML. Cell cycle dependent regulation of the origin recognition complex. Cell Cycle 2005;4:70–9.

Dephoure N, Zhou C, Villen J, Beausoleil SA, Bakalarski CE, Elledge SJ, et al. A quantitative atlas of mitotic phosphorylation. Proc Natl Acad Sci U S A 2008;105: 10762–7.

Desprat R, Thierry-Mieg D, Lailler N, Lajugie J, Schildkraut C, Thierry-Mieg J, et al. Predictable dynamic program of timing of DNA replication in human cells. Genome Res 2009;19:2288–99.

Devault A, Vallen EA, Yuan T, Green S, Bensimon A, Schwob E. Identification of Tah11/Sid2 as the ortholog of the replication licensing factor Cdt1 in *Saccharomyces cerevisiae*. Curr Biol 2002;12:689–94.

Devault A, Gueydon E, Schwob E. Interplay between S-cyclin-dependent kinase and Dbf4-dependent kinase in controlling DNA replication through phosphorylation of yeast Mcm4 N-terminal domain. Mol Biol Cell 2008;19:2267–77.

Dijkwel PA, Wang S, Hamlin JL. Initiation sites are distributed at frequent intervals in the Chinese hamster dihydrofolate reductase origin of replication but are used with very different efficiencies. Mol Cell Biol 2002;22:3053–65.

Dimitrova DS, Gilbert DM. The spatial position and replication timing of chromosomal domains are both established in early G1 phase. Mol Cell 1999;4:983–93.

Doksani Y, Bermejo R, Fiorani S, Haber JE, Foiani M. Replicon dynamics, dormant origin firing, and terminal fork integrity after double-strand break formation. Cell 2009;137:247–58.

Donaldson AD. Shaping time: chromatin structure and the DNA replication programme. Trends Genet 2005;21:444–9.

Drouin R, Lemieux N, Richer CL. Analysis of DNA replication during S-phase by means of dynamic chromosome banding at high resolution. Chromosoma 1990;99: 273–80.

Drury LS, Perkins G, Diffley JF. The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. Curr Biol 2000;10:231–40.

Duboule D. Vertebrate Hox genes and proliferation: an alternative pathway to homeosis? Curr Opin Genet Dev 1995;5:525–8.

Duncker BP, Chesnokov IN, McConkey BJ. The origin recognition complex protein family. Genome Biol 2009;10:214.

Eisenberg S, Korza G, Carson J, Liachko I, Tye BK. Novel DNA binding properties of the Mcm10 protein from *Saccharomyces cerevisiae*. J Biol Chem 2009;284: 25412–20.

Elsasser S, Lou F, Wang B, Campbell JL, Jong A. Interaction between yeast Cdc6 protein and B-type cyclin/Cdc28 kinases. Mol Biol Cell 1996;7:1723–35.

Evrin C, Clarke P, Zech J, Lurz R, Sun J, Uhle S, et al. A double-hexameric MCM2–7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. Proc Natl Acad Sci U S A 2009;106:20240–5.

Ferreira J, Carmo-Fonseca M. Genome replication in early mouse embryos follows a defined temporal and spatial order. J Cell Sci 1997;110(Pt 7):889–97.

Ferrezuelo F, Colomina N, Futcher B, Aldea M. The transcriptional network activated by Cln3 cyclin at the G1-to-S transition of the yeast cell cycle. Genome Biol 2010;11: R67.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, et al. The Pfam protein families database. Nucleic Acids Res 2010;38:D211–22.

Francis LI, Randell JC, Takara TJ, Uchima L, Bell SP. Incorporation into the prereplicative complex activates the Mcm2–7 helicase for Cdc7-Dbf4 phosphorylation. Genes Dev 2009;23:643–54.

Gambus A, Jones RC, Sanchez-Diaz A, Kanemaki M, van Deursen F, Edmondson RD, et al. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. Nat Cell Biol 2006;8:358–66.

Gambus A, van Deursen F, Polychronopoulos D, Foltman M, Jones RC, Edmondson RD, et al. A key role for Ctf4 in coupling the MCM2–7 helicase to DNA polymerase alpha within the eukaryotic replisome. EMBO J 2009;28:2992–3004.

Garcia V, Furuya K, Carr AM. Identification and functional analysis of TopBP1 and its homologs. DNA Repair (Amst) 2005;4:1227–39.

Gauci S, Helbig AO, Slijper M, Krijgsveld J, Heck AJ, Mohammed S. Lys-N and trypsin cover complementary parts of the phosphoproteome in a refined SCX-based approach. Anal Chem 2009;81:4493–501.

Gavin KA, Hidaka M, Stillman B. Conserved initiator proteins in eukaryotes. Science 1995;270:1667–71.

Ge XQ, Jackson DA, Blow JJ. Dormant origins licensed by excess Mcm2–7 are required for human cells to survive replicative stress. Genes Dev 2007;21:3331–41.

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, et al. Global analysis of protein expression in yeast. Nature 2003;425:737–41.

Ghosh S, Vassilev AP, Zhang J, Zhao Y, Depamphilis ML. Assembly of the human origin recognition complex occurs through independent nuclear localization of its components. J Biol Chem 2011;286:23831–41.

Giacca M, Zentilin L, Norio P, Diviacco S, Dimitrova D, Contreas G, et al. Fine mapping of a replication origin of human DNA. Proc Natl Acad Sci U S A 1994;91:7119–23.

Gilbert DM. Making sense of eukaryotic DNA replication origins. Science 2001;294:96-100.

Giordano-Coltart J, Ying CY, Gautier J, Hurwitz J. Studies of the properties of human origin recognition complex and its Walker A motif mutants. Proc Natl Acad Sci U S A 2005;102:69–74.

Glozak MA, Seto E. Acetylation/deacetylation modulates the stability of DNA replication licensing factor Cdt1. J Biol Chem 2009;284:11446–53.

Goldar A, Labit H, Marheineke K, Hyrien O. A dynamic stochastic model for DNA replication initiation in early embryos. PLoS One 2008;3:e2919.

Goren A, Cedar H. Replicating by the clock. Nat Rev Mol Cell Biol 2003;4:25–32.

Greer DA, Besley BD, Kennedy KB, Davey S. hRad9 rapidly binds DNA containing double-strand breaks and is required for damage-dependent topoisomerase II beta binding protein 1 focus formation. Cancer Res 2003;63:4829–35.

Gregan J, Lindner K, Brimage L, Franklin R, Namdar M, Hart EA, et al. Fission yeast Cdc23/Mcm10 functions after pre-replicative complex formation to promote Cdc45 chromatin binding. Mol Biol Cell 2003;14:3876–87.

Gupte RS, Weng Y, Liu L, Lee MY. The second subunit of the replication factor C complex (RFC40) and the regulatory subunit (RIalpha) of protein kinase A form a protein complex promoting cell survival. Cell Cycle (Georgetown, Tex) 2005;4:323–9.

Hansen JC, Nyborg JK, Luger K, Stargell LA. Histone chaperones, histone acetylation, and the fluidity of the chromogenome. J Cell Physiol 2010;224:289–99.

Hashimoto Y, Takisawa H. *Xenopus* Cut5 is essential for a CDK-dependent process in the initiation of DNA replication. EMBO J 2003;22:2526–35.

Hiratani I, Ryba T, Itoh M, Yokochi T, Schwaiger M, Chang CW, et al. Global reorganization of replication domains during embryonic stem cell differentiation. PLoS Biol 2008;6:e245.

Hiratani I, Ryba T, Itoh M, Rathjen J, Kulik M, Papp B, et al. Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. Genome Res 2010;20:155–69.

Hook SS, Lin JJ, Dutta A. Mechanisms to control rereplication and implications for cancer. Curr Opin Cell Biol 2007;19:663–71.

Hyrien O, Marheineke K, Goldar A. Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. Bioessays 2003;25:116–25.

Ibarra A, Schwob E, Mendez J. Excess MCM proteins protect human cells from replicative stress by licensing backup origins of replication. Proc Natl Acad Sci U S A 2008;105: 8956–61.

Ichikawa K, Noda T, Furuichi Y. Preparation of the gene targeted knockout mice for human premature aging diseases, Werner syndrome, and Rothmund–Thomson syndrome caused by the mutation of DNA helicases. Nippon Yakurigaku Zasshi 2002;119:219–26.

Iizuka M, Matsui T, Takisawa H, Smith MM. Regulation of replication licensing by acetyltransferase Hbo1. Mol Cell Biol 2006;26:1098–108.

Ilves I, Petojevic T, Pesavento JJ, Botchan MR. Activation of the MCM2–7 helicase by association with Cdc45 and GINS proteins. Mol Cell 2010;37:247–58.

Im JS, Ki SH, Farina A, Jung DS, Hurwitz J, Lee JK. Assembly of the Cdc45-Mcm2–7- GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins. Proc Natl Acad Sci U S A 2009;106:15628–32.

Ishimi Y, Komamura-Kohno Y. Phosphorylation of Mcm4 at specific sites by cyclindependent kinase leads to loss of Mcm4,6,7 helicase activity. J Biol Chem 2001;276:34428–33.

Izumi M, Yanagi K, Mizuno T, Yokoi M, Kawasaki Y, Moon KY, et al. The human homolog of *Saccharomyces cerevisiae* Mcm10 interacts with replication factors and dissociates from nuclease-resistant nuclear structures in G(2) phase. Nucleic Acids Res 2000;28:4769–77.

Izumi M, Yatagai F, Hanaoka F. Cell cycle-dependent proteolysis and phosphorylation of human Mcm10. J Biol Chem 2001;276:48526–31.

Jackson DA, Pombo A. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. J Cell Biol 1998;140:1285–95.

Jares P, Donaldson A, Blow JJ. The Cdc7/Dbf4 protein kinase: target of the S phase checkpoint? EMBO Rep 2000;1:319–22.

Jeon Y, Lee KY, Ko MJ, Lee YS, Kang S, Hwang DS. Human TopBP1 participates in cyclin E/CDK2 activation and preinitiation complex assembly during G1/S transition. J Biol Chem 2007;282:14882–90.

Jiang W, Wells NJ, Hunter T. Multistep regulation of DNA replication by Cdk phosphorylation of HsCdc6. Proc Natl Acad Sci U S A 1999;96:6193–8.

Johansson E, Macneill SA. The eukaryotic replicative DNA polymerases take shape.

Trends Biochem Sci 2010;35:339–47.

Kanemaki M, Labib K. Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. EMBO J 2006;25:1753–63.

Karnani N, Taylor CM, Malhotra A, Dutta A. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. Mol Biol Cell 2010;21:393–404.

Kawasaki Y, Hiraga S, Sugino A. Interactions between Mcm10p and other replication factors are required for proper initiation and elongation of chromosomal DNA replication in *Saccharomyces cerevisiae*. Genes Cells 2000;5:975–89.

Kinoshita Y, Johnson EM, Gordon RE, Negri-Bell H, Evans MT, Coolbaugh J, et al. Colocalization of MCM8 and MCM7 with proteins involved in distinct aspects of DNA replication. Microsc Res Tech 2008;71:288–97.

Klemm RD, Bell SP. ATP bound to the origin recognition complex is important for preRC formation. Proc Natl Acad Sci U S A 2001;98:8361–7.

Kumagai A, Lee J, Yoo HY, Dunphy WG. TopBP1 activates the ATR-ATRIP complex. Cell 2006;124:943–55.

Kumagai A, Shevchenko A, Shevchenko A, Dunphy WG. Treslin collaborates with TopBP1 in triggering the initiation of DNA replication. Cell 2010;140:349–59.

Labib K. How do Cdc7 and cyclin-dependent kinases trigger the initiation of chromosome replication in eukaryotic cells? Genes Dev 2010;24:1208–19.

Labib K, Diffley JF. Is the MCM2–7 complex the eukaryotic DNA replication fork helicase? Curr Opin Genet Dev 2001;11:64–70.

Labib K, Gambus A. A key role for the GINS complex at DNA replication forks. Trends Cell Biol 2007;17:271–8.

Larizza L, Roversi G, Volpi L. Rothmund–Thomson syndrome. Orphanet J Rare Dis 2010;5:2.

Lau E, Tsuji T, Guo L, Lu SH, Jiang W. The role of pre-replicative complex (pre-RC) components in oncogenesis. FASEB J 2007;21:3786–94.

Lebofsky R, Heilig R, Sonnleitner M, Weissenbach J, Bensimon A. DNA replication origin interference increases the spacing between initiation events in human cells. Mol Biol Cell 2006;17:5337–45.

Lee DG, Makhov AM, Klemm RD, Griffith JD, Bell SP. Regulation of origin recognition complex conformation and ATPase activity: differential effects of single-stranded and double-stranded DNA binding. EMBO J 2000;19:4774–82.

Lee JK, Seo YS, Hurwitz J. The Cdc23 (Mcm10) protein is required for the phosphorylation of minichromosome maintenance complex by the Dfp1-Hsk1 kinase. Proc Natl Acad Sci U S A 2003;100:2334–9.

Lee HO, Davidson JM, Duronio RJ. Endoreplication: polyploidy with purpose. Genes Dev 2009;23:2461–77.

Lee C, Liachko I, Bouten R, Kelman Z, Tye BK. Alternative mechanisms for coordinating polymerase alpha and MCM helicase. Mol Cell Biol 2010;30:423–35.

Levine K, Huang K, Cross FR. *Saccharomyces cerevisiae* G1 cyclins differ in their intrinsic functional specificities. Mol Cell Biol 1996;16:6794–803.

Li F, Chen J, Izumi M, Butler MC, Keezer SM, Gilbert DM. The replication timing program of the Chinese hamster beta-globin locus is established coincident with its repositioning near peripheral heterochromatin in early G1 phase. J Cell Biol 2001;154: 283–92.

Li X, Zhao Q, Liao R, Sun P, Wu X. The SCF(Skp2) ubiquitin ligase complex interacts with the human replication licensing factor Cdt1 and regulates Cdt1 degradation. J Biol Chem 2003;278:30854–8.

Lin DI, Aggarwal P, Diehl JA. Phosphorylation of MCM3 on Ser-112 regulates its incorporation into the MCM2–7 complex. Proc Natl Acad Sci U S A 2008;105: 8079–84.

Liu Y. Rothmund–Thomson syndrome helicase, RECQ4: on the crossroad between DNA replication and repair. DNA Repair (Amst) 2010;9:325–30.

Liu E, Li X, Yan F, Zhao Q, Wu X. Cyclin-dependent kinases phosphorylate human Cdt1 and induce its degradation. J Biol Chem 2004;279:17283–8.

Longhese MP, Plevani P, Lucchini G. Replication factor A is required in vivo for DNA replication, repair, and recombination. Mol Cell Biol 1994;14:7884–90.

Lord PG, Wheals AE. Asymmetrical division of *Saccharomyces cerevisiae*. J Bacteriol 1980;142:808–18.

Lovejoy CA, Lock K, Yenamandra A, Cortez D. DDB1 maintains genome integrity through regulation of Cdt1. Mol Cell Biol 2006;26:7977–90.

Lu Z, Hunter T. Ubiquitylation and proteasomal degradation of the p21 (Cip1), p27 (Kip1) and p57 (Kip2) CDK inhibitors. Cell Cycle 2010;9:2342–52.

Lutzmann M, Mechali M. MCM9 binds Cdt1 and is required for the assembly of prereplication complexes. Mol Cell 2008;31:190–200.

Ma H, Samarabandu J, Devdhar RS, Acharya R, Cheng PC, Meng C, et al. Spatial and temporal dynamics of DNA replication sites in mammalian cells. J Cell Biol 1998;143: 1415–25.

MacAlpine DM, Bell SP. A genomic view of eukaryotic DNA replication. Chromosome Res 2005;13:309–26.

MacNeill SA. Structure and function of the GINS complex, a key component of the eukaryotic replisome. Biochem J 2010;425:489–500.

Mailand N, Diffley JF. CDKs promote DNA replication origin licensing in human cells by protecting Cdc6 from APC/C-dependent proteolysis. Cell 2005;122:915–26.

Maiorano D, Lemaitre JM, Mechali M. Stepwise regulated chromatin assembly of MCM2–7 proteins. J Biol Chem 2000;275:8426–31.

Maiorano D, Cuvier O, Danis E, Mechali M. MCM8 is an MCM2–7-related protein that functions as a DNA helicase during replication elongation and not initiation. Cell 2005;120:315–28.

Maire G, Yoshimoto M, Chilton-MacNeill S, Thorner PS, Zielenska M, Squire JA. Recurrent RECQL4 imbalance and increased gene expression levels are associated with structural chromosomal instability in sporadic osteosarcoma. Neoplasia 2009;11: 260–8. (3 pp. following 8).

Makiniemi M, Hillukkala T, Tuusa J, Reini K, Vaara M, Huang D, et al. BRCT domain-containing protein TopBP1 functions in DNA replication and damage response. J Biol Chem 2001;276:30399–406.

Malumbres M, Barbacid M. Mammalian cyclin-dependent kinases. Trends Biochem Sci 2005;30:630–41.

Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev 2009;9:153–66.

Marchetti L, Comelli L, D'Innocenzo B, Puzzi L, Luin S, Arosio D, et al. Homeotic proteins participate in the function of human-DNA replication origins. Nucleic Acids Res 2010;38:8105–19.

Masai H, Arai K. Regulation of DNA replication during the cell cycle: roles of Cdc7 kinase and coupling of replication, recombination, and repair in response to replication fork arrest. IUBMB Life 2000;49:353–64.

Masai H, Arai K. Cdc7 kinase complex: a key regulator in the initiation of DNA replication. J Cell Physiol 2002;190:287–96.

Masai H, Taniyama C, Ogino K, Matsui E, Kakusho N, Matsumoto S, et al. Phosphorylation of MCM4 by Cdc7 kinase facilitates its interaction with Cdc45 on the chromatin. J Biol Chem 2006;281:39249–61.

Masai H, Matsumoto S, You Z, Yoshizawa-Sugata N, Oda M. Eukaryotic chromosome DNA replication: where, when, and how? Annu Rev Biochem 2010;79: 89-130.

Masumoto H, Muramatsu S, Kamimura Y, Araki H. S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. Nature 2002;415:651–5.

Matsuno K, Kumano M, Kubota Y, Hashimoto Y, Takisawa H. The N-terminal noncatalytic region of *Xenopus* RecQ4 is required for chromatin binding of DNA polymerase alpha in the initiation of DNA replication. Mol Cell Biol 2006;26: 4843–52.

Matsushime H, Ewen ME, Strom DK, Kato JY, Hanks SK, Roussel MF, et al. Identification and properties of an atypical catalytic subunit (p34PSK-J3/cdk4) for mammalian D type G1 cyclins. Cell 1992;71:323–34.

Maya-Mendoza A, Jackson DA. Replication timing: findings from fibers. Cell Cycle 2009;8:3073–4.

Maya-Mendoza A, Olivares-Chauvet P, Shaw A, Jackson DA. S phase progression in human cells is dictated by the genetic continuity of DNA foci. PLoS Genet 2010;6: e1000900.

Mayer ML, Gygi SP, Aebersold R, Hieter P. Identification of RFC(Ctf18p, Ctf8p, Dcc1p): an alternative RFC complex required for sister chromatid cohesion in *S. cerevisiae*. Mol Cell 2001;7:959–70.

Mayya V, Lundgren DH, Hwang SI, Rezaul K, Wu L, Eng JK, et al. Quantitative phosphoproteomic analysis of T cell receptor signaling reveals system-wide modulation of protein–protein interactions. Sci Signal 2009;2: ra 46.

McGarry TJ, Kirschner MW. Geminin, an inhibitor of DNA replication, is degraded during mitosis. Cell 1998;93:1043–53.

Mechali M. Eukaryotic DNA replication origins: many choices for appropriate answers. Nat Rev Mol Cell Biol 2010;11:728–38.

Mendez J, Stillman B. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. Mol Cell Biol 2000;20:8602–12.

Mendez J, Zou-Yang XH, Kim SY, Hidaka M, Tansey WP, Stillman B. Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. Mol Cell 2002;9:481–91.

Meng TC, Somani S, Dhar P. Modeling and simulation of biological systems with stochasticity. In Silico Biol 2004;4:293–309.

Merchant AM, Kawasaki Y, Chen Y, Lei M, Tye BK. A lesion in the DNA replication initiation factor Mcm10 induces pausing of elongation forks through chromosomal replication origins in *Saccharomyces cerevisiae*. Mol Cell Biol 1997;17:3261–71.

Meyerson M, Harlow E. Identification of G1 kinase activity for cdk6, a novel cyclin D partner. Mol Cell Biol 1994;14:2077–86.

Miller ME, Cross FR. Distinct subcellular localization patterns contribute to functional specificity of the Cln2 and Cln3 cyclins of *Saccharomyces cerevisiae*. Mol Cell Biol 2000;20:542–55.

Mimura S, Seki T, Tanaka S, Diffley JF. Phosphorylation-dependent binding of mitotic cyclins to Cdc6 contributes to DNA replication control. Nature 2004;431:1118–23.

Miotto B, Struhl K. HBO1 histone acetylase is a coactivator of the replication licensing factor Cdt1. Genes Dev 2008;22:2633–8.

Miotto B, Struhl K. HBO1 histone acetylase activity is essential for DNA replication licensing and inhibited by Geminin. Mol Cell 2010;37:57–66.

Mizushima T, Takahashi N, Stillman B. Cdc6p modulates the structure and DNA binding activity of the origin recognition complex in vitro. Genes Dev 2000;14: 1631–41.

Montagnoli A, Valsasina B, Brotherton D, Troiani S, Rainoldi S, Tenca P, et al. Identification of Mcm2 phosphorylation sites by S-phase-regulating kinases. J Biol Chem 2006;281:10281–90.

Montanari M, Macaluso M, Cittadini A, Giordano A. Role of geminin: from normal control of DNA replication to cancer formation and progression? Cell Death Differ 2006;13:1052–6.

Mossi R, Jonsson ZO, Allen BL, Hardin SH, Hubscher U. Replication factor C interacts with the C-terminal side of proliferating cell nuclear antigen. J Biol Chem 1997;272:1769–76.

Motegi A, Liaw HJ, Lee KY, Roest HP, Maas A, Wu X, et al. Polyubiquitination of proliferating cell nuclear antigen by HLTF and SHPRH prevents genomic instability from stalled replication forks. Proc Natl Acad Sci U S A 2008;105:12411–6.

Moyer SE, Lewis PW, Botchan MR. Isolation of the Cdc45/Mcm2–7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. Proc Natl Acad Sci U S A 2006;103:10236–41.

Murakami H, Yanow SK, Griffiths D, Nakanishi M, Nurse P. Maintenance of replication forks and the S-phase checkpoint by Cdc18p and Orp1p. Nat Cell Biol 2002;4: 384–8.

Muramatsu S, Hirai K, Tak YS, Kamimura Y, Araki H. CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon), and GINS in budding yeast. Genes Dev 2010;24:602–12.

Nash P, Tang X, Orlicky S, Chen Q, Gertler FB, Mendenhall MD, et al. Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. Nature 2001;414:514–21.

Natsume T, Tanaka TU. Spatial regulation and organization of DNA replication within the nucleus. Chromosome Res 2010;18:7-17.

Navadgi-Patil VM, Burgers PM. Yeast DNA replication protein Dpb11 activates the Mec1/ATR checkpoint kinase. J Biol Chem 2008;283:35853–9.

Newport J, Kirschner M. A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell 1982;30:675–86.

Nieduszynski CA, Hiraga S, Ak P, Benham CJ, Donaldson AD. OriDB: a DNA replication origin database. Nucleic Acids Res 2007;35:D40–6.

Noguchi E, Shanahan P, Noguchi C, Russell P. CDK phosphorylation of Drc1 regulates DNA replication in fission yeast. Curr Biol 2002;12:599–605.

Oehlmann M, Score AJ, Blow JJ. The role of Cdc6 in ensuring complete genome licensing and S phase checkpoint activation. J Cell Biol 2004;165:181–90.

Ohta S, Tatsumi Y, Fujita M, Tsurimoto T, Obuse C. The ORC1 cycle in human cells: II. Dynamic changes in the human ORC complex during the cell cycle. J Biol Chem 2003;278:41535–40.

Olsen JV, Vermeulen M, Santamaria A, Kumar C, Miller ML, Jensen LJ, et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal 2010;3:ra3.

Pacek M, Walter JC. A requirement for MCM7 and Cdc45 in chromosome unwinding during eukaryotic DNA replication. EMBO J 2004;23:3667–76.

Paolinelli R, Mendoza-Maldonado R, Cereseto A, Giacca M. Acetylation by GCN5 regulates CDC6 phosphorylation in the S phase of the cell cycle. Nat Struct Mol Biol 2009;16:412–20.

Patel S, Latterich M. The AAA team: related ATPases with diverse functions. Trends Cell Biol 1998;8:65–71.

Petersen BO, Lukas J, Sorensen CS, Bartek J, Helin K. Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. EMBO J 1999;18: 396–410.

Petersen BO, Wagener C, Marinoni F, Kramer ER, Melixetian M, Lazzerini Denchi E, et al. Cell cycle- and cell growth-regulated proteolysis of mammalian CDC6 is dependent on APC-CDH1. Genes Dev 2000;14:2330–43.

Petropoulou C, Kotantaki P, Karamitros D, Taraviras S. Cdt1 and Geminin in cancer: markers or triggers of malignant transformation? Front Biosci 2008;13:4485–94.

Piergiovanni G, Costanzo V. GEMC1 is a novel TopBP1-interacting protein involved in chromosomal DNA replication. Cell Cycle 2010;9:3662–6.

Pollok S, Grosse F. Cdc45 degradation during differentiation and apoptosis. Biochem Biophys Res Commun 2007;362:910–5.

Pospiech H, Grosse F, Pisani FM. The initiation step of eukaryotic DNA replication. Subcell Biochem 2010;50:79-104.

Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, et al. Replication dynamics of the yeast genome. Science 2001;294:115–21.

Randell JC, Bowers JL, Rodriguez HK, Bell SP. Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2–7 helicase. Mol Cell 2006;21:29–39.

Remus D, Diffley JF. Eukaryotic DNA replication control: lock and load, then fire. Curr Opin Cell Biol 2009;21:771–7.

Resar LM. The high mobility group A1 gene: transforming inflammatory signals into cancer? Cancer Res 2010;70:436–9.

Resat H, Costa MN, Shankaran H. Spatial aspects in biological system simulations. Methods Enzymol 2011;487:485–511.

Rhind N. DNA replication timing: random thoughts about origin firing. Nat Cell Biol 2006;8:1313–6.

Rhind N, Yang SC, Bechhoefer J. Reconciling stochastic origin firing with defined replication timing. Chromosome Res 2010;18:35–43.

Ricke RM, Bielinsky AK. Mcm10 regulates the stability and chromatin association of DNA polymerase-alpha. Mol Cell 2004;16:173–85.

Rivin CJ, Fangman WL. Replication fork rate and origin activation during the S phase of *Saccharomyces cerevisiae*. J Cell Biol 1980;85:108–15.

Sadoni N, Cardoso MC, Stelzer EH, Leonhardt H, Zink D. Stable chromosomal units determine the spatial and temporal organization of DNA replication. J Cell Sci 2004;117:5353–65.

Saha P, Chen J, Thome KC, Lawlis SJ, Hou ZH, Hendricks M, et al. Human CDC6/Cdc18 associates with Orc1 and cyclin-cdk and is selectively eliminated from the nucleus at the onset of S phase. Mol Cell Biol 1998;18:2758–67.

Salazar C, Hofer T. Multisite protein phosphorylation—from molecular mechanisms to kinetic models. FEBS J 2009;276:3177–98.

Salazar C, Brummer A, Alberghina L, Hofer T. Timing control in regulatory networks by multisite protein modifications. Trends Cell Biol 2010;20:634–41.

Salsi V, Ferrari S, Ferraresi R, Cossarizza A, Grande A, Zappavigna V. HOXD13 binds DNA replication origins to promote origin licensing and is inhibited by geminin. Mol Cell Biol 2009;29:5775–88.

Sanchez I, Dynlacht BD. New insights into cyclins, CDKs, and cell cycle control. Semin Cell Dev Biol 2005;16:311–21.

Sanchez-Pulido L, Diffley JF, Ponting CP. Homology explains the functional similarities of Treslin/Ticrr and Sld3. Curr Biol 2010;20:R509–10.

Sangrithi MN, Bernal JA, Madine M, Philpott A, Lee J, Dunphy WG, et al. Initiation of DNA replication requires the RECQL4 protein mutated in Rothmund–Thomson syndrome. Cell 2005;121:887–98.

Sawyer SL, Cheng IH, Chai W, Tye BK. Mcm10 and Cdc45 cooperate in origin activation in *Saccharomyces cerevisiae*. J Mol Biol 2004;340:195–202.

Saxena S, Dutta A. Geminin-Cdt1 balance is critical for genetic stability. Mutat Res 2005;569:111–21.

Schmidt U, Wollmann Y, Franke C, Grosse F, Saluz HP, Hanel F. Characterization of the interaction between the human DNA topoisomerase IIbeta-binding protein 1 (TopBP1) and the cell division cycle 45 (Cdc45) protein. Biochem J 2008;409:169–77.

Schwob E, Nasmyth K. CLB5 and CLB6, a new pair of B cyclins involved in DNA replication in *Saccharomyces cerevisiae*. Genes Dev 1993;7:1160–75.

Sclafani RA. Cdc7p-Dbf4p becomes famous in the cell cycle. J Cell Sci 2000;113(Pt 12): 2111–7.

Sclafani RA, Holzen TM. Cell cycle regulation of DNA replication. Annu Rev Genet 2007;41:237–80.

Semple JW, Da-Silva LF, Jervis EJ, Ah-Kee J, Al-Attar H, Kummer L, et al. An essential role for Orc6 in DNA replication through maintenance of pre-replicative complexes. EMBO J 2006;25:5150–8.

Senga T, Sivaprasad U, Zhu W, Park JH, Arias EE, Walter JC, et al. PCNA is a cofactor for Cdt1 degradation by CUL4/DDB1-mediated N-terminal ubiquitination. J Biol Chem 2006;281:6246–52.

Sequeira-Mendes J, Diaz-Uriarte R, Apedaile A, Huntley D, Brockdorff N, Gomez M. Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet 2009;5:e1000446.

Sherr CJ, Roberts JM. CDK inhibitors: positive and negative regulators of G1-phase progression. Genes Dev 1999;13:1501–12.

Sherr CJ, Roberts JM. Living with or without cyclins and cyclin-dependent kinases. Genes Dev 2004;18:2699–711.

Sheu YJ, Stillman B. Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. Mol Cell 2006;24:101–13.

Sheu YJ, Stillman B. The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. Nature 2010;463:113–7.

Siddiqui K, Stillman B. ATP-dependent assembly of the human origin recognition complex. J Biol Chem 2007;282:32370–83.

Sokka M, Parkkinen S, Pospiech H, Syvaoja JE. Function of TopBP1 in genome stability. Subcell Biochem 2010;50:119–41.

Sporbert A, Gahl A, Ankerhold R, Leonhardt H, Cardoso MC. DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters. Mol Cell 2002;10:1355–65.

Stuart D, Wittenberg C. CLN3, not positive feedback, determines the timing of CLN2 transcription in cycling cells. Genes Dev 1995;9:2780–94.

Sugimoto N, Tatsumi Y, Tsurumi T, Matsukage A, Kiyono T, Nishitani H, et al. Cdt1 phosphorylation by cyclin A-dependent kinases negatively regulates its function without affecting geminin binding. J Biol Chem 2004;279:19691–7.

Sugimoto N, Kitabayashi I, Osano S, Tatsumi Y, Yugawa T, Narisawa-Saito M, et al. Identification of novel human Cdt1-binding proteins by a proteomics approach: proteolytic regulation by APC/CCdh1. Mol Biol Cell 2008;19:1007–21.

Taddei A, Hediger F, Neumann FR, Gasser SM. The function of nuclear architecture: a genetic approach. Annu Rev Genet 2004;38:305–45.

Tak YS, Tanaka Y, Endo S, Kamimura Y, Araki H. A CDK-catalysed regulatory phosphorylation for formation of the DNA replication complex Sld2-Dpb11. EMBO J 2006;25:1987–96.

Takebayashi SI, Manders EM, Kimura H, Taguchi H, Okumura K. Mapping sites where replication initiates in mammalian cells using DNA fibers. Exp Cell Res 2001;271: 263–8.

Takeda DY, Parvin JD, Dutta A. Degradation of Cdt1 during S phase is Skp2-independent and is required for efficient progression of mammalian cells through S phase. J Biol Chem 2005;280:23416–23.

Tanaka S, Diffley JF. Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2–7 during G1 phase. Nat Cell Biol 2002;4:198–207.

Tanaka S, Umemori T, Hirai K, Muramatsu S, Kamimura Y, Araki H. CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. Nature 2007;445:328–32.

Tanaka T, Umemori T, Endo S, Muramatsu S, Kanemaki M, Kamimura Y, et al. Sld7, an Sld3-associated protein required for efficient chromosomal DNA replication in budding yeast. EMBO J 2011;30:2019–30.

Tatsumi Y, Ohta S, Kimura H, Tsurimoto T, Obuse C. The ORC1 cycle in human cells: I. cell cycle-regulated oscillation of human ORC1. J Biol Chem 2003;278:41528–34.

Thangavel S, Mendoza-Maldonado R, Tissino E, Sidorova JM, Yin J, Wang W, et al. Human RECQ1 and RECQ4 helicases play distinct roles in DNA replication initiation. Mol Cell Biol 2010;30:1382–96.

Thomae AW, Pich D, Brocher J, Spindler MP, Berens C, Hock R, et al. Interaction between HMGA1a and the origin recognition complex creates site-specific replication origins. Proc Natl Acad Sci U S A 2008;105:1692–7.

Thomae AW, Baltin J, Pich D, Deutsch MJ, Ravasz M, Zeller K, et al. Different roles of the human Orc6 protein in the replication initiation process. Cell Mol Life Sci 2011. doi:10.1007/s00018-011-0675-9.

Todorovic V, Falaschi A, Giacca M. Replication origins of mammalian chromosomes: the happy few. Front Biosci 1999;4:D859–68.

Truong LN, Wu X. Prevention of DNA re-replication in eukaryotic cells. J Mol Cell Biol 2011;3:13–22.

Tsakraklides V, Bell SP. Dynamics of pre-replicative complex assembly. J Biol Chem 2010;285:9437–43.

Tsuji T, Ficarro SB, Jiang W. Essential role of phosphorylation of MCM2 by Cdc7/Dbf4 in the initiation of DNA replication in mammalian cells. Mol Biol Cell 2006;17: 4459–72.

Tsutsui Y, Morishita T, Natsume T, Yamashita K, Iwasaki H, Yamao F, et al. Genetic and physical interactions between *Schizosaccharomyces pombe* Mcl1 and Rad2, Dna2 and DNA polymerase alpha: evidence for a multifunctional role of Mcl1 in DNA replication and repair. Curr Genet 2005;48:34–43.

Tyers M. The cyclin-dependent kinase inhibitor p40SIC1 imposes the requirement for Cln G1 cyclin function at Start. Proceedings of the National Academy of Sciences of the United States of America 1996;93:7772–6.

Tyers M, Tokiwa G, Futcher B. Comparison of the *Saccharomyces cerevisiae* G1 cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins. EMBO J 1993;12:1955–68.

Ullah Z, Lee CY, Lilly MA, DePamphilis ML. Developmentally programmed endoreduplication in animals. Cell Cycle 2009;8:1501–9.

Van Hatten RA, Tutter AV, Holway AH, Khederian AM, Walter JC, Michael WM. The *Xenopus* Xmus101 protein is required for the recruitment of Cdc45 to origins of DNA replication. J Cell Biol 2002;159:541–7.

Vanoni M, Vai M, Popolo L, Alberghina L. Structural heterogeneity in populations of the budding yeast *Saccharomyces cerevisiae*. J Bacteriol 1983;156:1282–91.

Volkening M, Hoffmann I. Involvement of human MCM8 in prereplication complex assembly by recruiting hcdc6 to chromatin. Mol Cell Biol 2005;25:1560–8.

Wang B, Feng L, Hu Y, Huang SH, Reynolds CP, Wu L, et al. The essential role of *Saccharomyces cerevisiae* CDC6 nucleotide-binding site in cell growth, DNA synthesis, and Orc1 association. J Biol Chem 1999;274:8291–8.

Wang B, Malik R, Nigg EA, Korner R. Evaluation of the low-specificity protease elastase for large-scale phosphoproteome analysis. Anal Chem 2008;80:9526–33.

Warren EM, Huang H, Fanning E, Chazin WJ, Eichman BF. Physical interactions between Mcm10, DNA, and DNA polymerase alpha. J Biol Chem 2009;284:24662–72.

Weisshart K, Pestryakov P, Smith RW, Hartmann H, Kremmer E, Lavrik O, et al. Coordinated regulation of replication protein A activities by its subunits p14 and p32. J Biol Chem 2004;279:35368–76.

Wilmes GM, Archambault V, Austin RJ, Jacobson MD, Bell SP, Cross FR. Interaction of the S-phase cyclin Clb5 with an "RXL" docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. Genes Dev 2004;18: 981–91.

Wohlschlegel JA, Dwyer BT, Dhar SK, Cvetic C, Walter JC, Dutta A. Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. Science 2000;290:2309–12.

Wohlschlegel JA, Dhar SK, Prokhorova TA, Dutta A, Walter JC. *Xenopus* Mcm10 binds to origins of DNA replication after Mcm2–7 and stimulates origin binding of Cdc45. Mol Cell 2002;9:233–40.

Woodfine K, Fiegler H, Beare DM, Collins JE, McCann OT, Young BD, et al. Replication timing of the human genome. Hum Mol Genet 2004;13:191–202.

Woodward AM, Gohler T, Luciani MG, Oehlmann M, Ge X, Gartner A, et al. Excess Mcm2–7 license dormant origins of replication that can be used under conditions of replicative stress. J Cell Biol 2006;173:673–83.

Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, et al. Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. Science 2001;294:2357–60.

Xu X, Liu Y. Dual DNA unwinding activities of the Rothmund–Thomson syndrome protein, RECQ4. EMBO J 2009;28:568–77.

Xu X, Rochette PJ, Feyissa EA, Su TV, Liu Y. MCM10 mediates RECQ4 association with MCM2–7 helicase complex during DNA replication. EMBO J 2009;28:3005–14.

Yabuki N, Terashima H, Kitada K. Mapping of early firing origins on a replication profile of budding yeast. Genes Cells 2002;7:781–9.

Yamane K, Kawabata M, Tsuruo T. A DNA-topoisomerase-II-binding protein with eight repeating regions similar to DNA-repair enzymes and to a cell-cycle regulator. Eur J Biochem 1997;250:794–9.

Yamane K, Wu X, Chen J. A DNA damage-regulated BRCT-containing protein, TopBP1, is required for cell survival. Mol Cell Biol 2002;22:555–66.

Yamane K, Chen J, Kinsella TJ. Both DNA topoisomerase II-binding protein 1 and BRCA1 regulate the G2-M cell cycle checkpoint. Cancer Res 2003;63:3049–53.

Yang SC, Rhind N, Bechhoefer J. Modeling genome-wide replication kinetics reveals a mechanism for regulation of replication timing. Mol Syst Biol 2010;6:404.

Yao N, Coryell L, Zhang D, Georgescu RE, Finkelstein J, Coman MM, et al. Replication factor C clamp loader subunit arrangement within the circular pentamer and its attachment points to proliferating cell nuclear antigen. J Biol Chem 2003;278: 50744–53.

Yoshida K. Identification of a novel cell-cycle-induced MCM family protein MCM9. Biochem Biophys Res Commun 2005;331:669–74.

Yu X, Chini CC, He M, Mer G, Chen J. The BRCT domain is a phospho-protein binding domain. Science 2003;302:639–42.

Zegerman P, Diffley JF. Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature 2007;445:281–5.

Zhang X, Morera S, Bates PA, Whitehead PC, Coffer AI, Hainbucher K, et al. Structure of an XRCC1 BRCT domain: a new protein–protein interaction module. EMBO J 1998;17:6404–11.

Zhang G, Gibbs E, Kelman Z, O'Donnell M, Hurwitz J. Studies on the interactions between human replication factor C and human proliferating cell nuclear antigen. Proc Natl Acad Sci U S A 1999;96:1869–74.

Zhu W, Ukomadu C, Jha S, Senga T, Dhar SK, Wohlschlegel JA, et al. Mcm10 and And1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. Genes Dev 2007;21:2288–99.

Zink D. The temporal program of DNA replication: new insights into old questions. Chromosoma 2006;115:273–87.

Zou L, Stillman B. Assembly of a complex containing Cdc45p, replication protein A, and Mcm2p at replication origins controlled by S-phase cyclin-dependent kinases and Cdc7p-Dbf4p kinase. Mol Cell Biol 2000;20:3086–96.

Zou Y, Liu Y, Wu X, Shell SM. Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. J Cell Physiol 2006;208: 267–73.

Chapter 3

Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994;2:28-36.

Bonasio R, Lecona E, Reinberg D. MBT domain proteins in development and disease. Semin Cell Dev Biol. 2010;21:221-30.

Brown CJ, Johnson AK, Daughdrill GW. Comparing models of evolution for ordered and disordered proteins. Mol Biol Evol. 2010;27:609-21.

Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, et al. Evolutionary rate heterogeneity in proteins with long disordered regions. J Mol Evol. 2002;55:104-10.

Burke JR, Deshong AJ, Pelton JG, Rubin SM. Phosphorylation-induced conformational changes in the retinoblastoma protein inhibit E2F transactivation domain binding. J Biol Chem. 2010;285:16286-93.

Calzone L, Gelay A, Zinovyev A, Radvanyi F, Barillot E. A comprehensive modular map of molecular interactions in RB/E2F pathway. Mol Syst Biol. 2008;4:173.

Carr SM, Munro S, Kessler B, Oppermann U, La Thangue NB. Interplay between lysine methylation and Cdk phosphorylation in growth control by the retinoblastoma protein. EMBO J. 2011;30:317-27.

Chan HM, Krstic-Demonacos M, Smith L, Demonacos C, La Thangue NB. Acetylation control of the retinoblastoma tumour-suppressor protein. Nat Cell Biol. 2001a;3:667-74.

Chan HM, Smith L, La Thangue NB. Role of LXCXE motif-dependent interactions in the activity of the retinoblastoma protein. Oncogene. 2001b;20:6152-63.

Chicas A, Wang X, Zhang C, McCurrach M, Zhao Z, Mert O, et al. Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. Cancer Cell. 2010;17:376-87.

Classon M, Dyson N. p107 and p130: versatile proteins with interesting pockets. Exp Cell Res. 2001;264:135-47.

Cobrinik D. Pocket proteins and cell cycle control. Oncogene. 2005;24:2796-809.

Cobrinik D, Whyte P, Peeper DS, Jacks T, Weinberg RA. Cell cycle-specific association of E2F with the p130 E1A-binding protein. Genes Dev. 1993;7:2392-404.

Cooper K. Rb, whi it's not just for metazoans anymore. Oncogene. 2006;25:5228-32.

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The genetic landscape of a cell. Science. 2010;327:425-31.

Craven RJ, Cance WG, Liu ET. The nuclear tyrosine kinase Rak associates with the retinoblastoma protein pRb. Cancer Res. 1995;55:3969-72.

Cuff JA, Barton GJ. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. Proteins. 2000;40:502-11.

Cusick ME, Yu H, Smolyar A, Venkatesan K, Carvunis AR, Simonis N, et al. Literature-curated protein interaction datasets. Nat Methods. 2009;6:39-46.

Dasgupta P, Sun J, Wang S, Fusaro G, Betts V, Padmanabhan J, et al. Disruption of the Rb--Raf-1 interaction inhibits tumor growth and angiogenesis. Mol Cell Biol. 2004;24:9527-41.

Daughdrill GW, Narayanaswami P, Gilmore SH, Belczyk A, Brown CJ. Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. J Mol Evol. 2007;65:277-88.

Delston RB, Matatall KA, Sun Y, Onken MD, Harbour JW. p38 phosphorylates Rb on Ser567 by a novel, cell cycle-independent mechanism that triggers Rb-Hdm2 interaction and apoptosis. Oncogene. 2011;30:588-99.

Durfee T, Becherer K, Chen PL, Yeh SH, Yang Y, Kilburn AE, et al. The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. Genes Dev. 1993;7:555-69.

Dyson N, Dembski M, Fattaey A, Ngwu C, Ewen M, Helin K. Analysis of p107-associated proteins: p107 associates with a form of E2F that differs from pRB-associated E2F-1. J Virol. 1993;67:7641-7.

Fajas L, Egler V, Reiter R, Hansen J, Kristiansen K, Debril MB, et al. The retinoblastoma-histone deacetylase 3 complex inhibits PPARgamma and adipocyte differentiation. Dev Cell. 2002;3:903-10.

Ferreira R, Magnaghi-Jaulin L, Robin P, Harel-Bellan A, Trouche D. The three members of the pocket proteins family share the ability to repress E2F activity through recruitment of a histone deacetylase. Proc Natl Acad Sci U S A. 1998;95:10493-8.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, et al. The Pfam protein families database. Nucleic Acids Res. 2010;38:D211-22.

Flores-Delgado G, Liu CW, Sposto R, Berndt N. A limited screen for protein interactions reveals new roles for protein phosphatase 1 in cell cycle control and apoptosis. J Proteome Res. 2007;6:1165-75.

Fuks F, Hurd PJ, Deplus R, Kouzarides T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res. 2003;31:2305-12.

Giacinti C, Giordano A. RB and cell cycle progression. Oncogene. 2006;25:5220-7.

Gonzalo S, Garcia-Cao M, Fraga MF, Schotta G, Peters AH, Cotter SE, et al. Role of the RB1 family in stabilizing histone methylation at constitutive heterochromatin. Nat Cell Biol. 2005;7:420-8.

Gunasekaran K, Tsai CJ, Nussinov R. Analysis of ordered and disordered protein complexes reveals structural features discriminating between stable and unstable monomers. J Mol Biol. 2004;341:1327-41.

Hamdane M, Bretteville A, Sambo AV, Schindowski K, Begard S, Delacourte A, et al. p25/Cdk5-mediated retinoblastoma phosphorylation is an early event in neuronal cell death. J Cell Sci. 2005;118:1291-8.

Hannon GJ, Demetrick D, Beach D. Isolation of the Rb-related p130 through its interaction with CDK2 and cyclins. Genes Dev. 1993;7:2378-91.

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 2004;32:D258-61.

Hijmans EM, Voorhoeve PM, Beijersbergen RL, van 't Veer LJ, Bernards R. E2F-5, a new E2F family member that interacts with p130 in vivo. Mol Cell Biol. 1995;15:3082-9.

Hirschi A, Cecchini M, Steinhardt RC, Schamber MR, Dick FA, Rubin SM. An overlapping kinase and phosphatase docking site regulates activity of the retinoblastoma protein. Nat Struct Mol Biol. 2010;17:1051-7.

Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, et al. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 2004;32:1037-49.

Ingrell CR, Miller ML, Jensen ON, Blom N. NetPhosYeast: prediction of protein phosphorylation sites in yeast. Bioinformatics. 2007;23:895-7.

Inoue Y, Kitagawa M, Taya Y. Phosphorylation of pRB at Ser612 by Chk1/2 leads to a complex between pRB and E2F-1 after DNA damage. EMBO J. 2007;26:2083-93.

Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999;292:195-202.

Jorgensen P, Nishikawa JL, Breitkreutz BJ, Tyers M. Systematic identification of pathways that couple cell growth and division in yeast. Science. 2002;297:395-400.

Kaizu K, Ghosh S, Matsuoka Y, Moriya H, Shimizu-Yoshida Y, Kitano H. A comprehensive molecular interaction map of the budding yeast cell cycle. Mol Syst Biol. 2010;6:415.

Lacy S, Whyte P. Identification of a p130 domain mediating interactions with cyclin A/cdk 2 and cyclin E/cdk 2 complexes. Oncogene. 1997;14:2395-406.

Lai A, Lee JM, Yang WM, DeCaprio JA, Kaelin WG, Jr., Seto E, et al. RBP1 recruits both histone deacetylase-dependent and -independent repression activities to retinoblastoma family proteins. Mol Cell Biol. 1999;19:6632-41.

Ledl A, Schmidt D, Muller S. Viral oncoproteins E1A and E7 and cellular LxCxE proteins repress SUMO modification of the retinoblastoma tumor suppressor. Oncogene. 2005;24:3810-8.

Lee S, Cha JY, Kim H, Yu U. GoBean: a Java GUI application for visual exploration of GO term enrichments. BMB Rep. 2012;45:120-5.

Lees JA, Buchkovich KJ, Marshak DR, Anderson CW, Harlow E. The retinoblastoma protein is phosphorylated on multiple sites by human cdc2. EMBO J. 1991;10:4279-90.

Li Y, Graham C, Lacy S, Duncan AM, Whyte P. The adenovirus E1A-associated 130-kD protein is encoded by a member of the retinoblastoma gene family and physically interacts with cyclins A and E. Genes Dev. 1993;7:2366-77.

Litovchick L, Sadasivam S, Florens L, Zhu X, Swanson SK, Velmurugan S, et al. Evolutionarily conserved multisubunit RBL2/p130 and E2F4 protein complex represses human cell cycle-dependent genes in quiescence. Mol Cell. 2007;26:539-51.

Lopes M, Cotta-Ramusino C, Pellicioli A, Liberi G, Plevani P, Muzi-Falconi M, et al. The DNA replication checkpoint response stabilizes stalled replication forks. Nature. 2001;412:557-61.

Los AP, Vinke FP, de Widt J, Topham MK, van Blitterswijk WJ, Divecha N. The retinoblastoma family proteins bind to and activate diacylglycerol kinase zeta. J Biol Chem. 2006;281:858-66.

Marsh JA, Forman-Kay JD. Sequence determinants of compaction in intrinsically disordered proteins. Biophys J. 2010;98:2383-90.

Mulligan G, Jacks T. The retinoblastoma gene family: cousins with overlapping interests. Trends Genet. 1998;14:223-9.

Nair JS, Ho AL, Tse AN, Coward J, Cheema H, Ambrosini G, et al. Aurora B kinase regulates the postmitotic endoreduplication checkpoint via phosphorylation of the retinoblastoma protein at serine 780. Mol Biol Cell. 2009;20:2218-28.

Nicolas E, Roumillac C, Trouche D. Balance between acetylation and methylation of histone H3 lysine 9 on the E2F-responsive dihydrofolate reductase promoter. Mol Cell Biol. 2003;23:1614-22.

Nielsen SJ, Schneider R, Bauer UM, Bannister AJ, Morrison A, O'Carroll D, et al. Rb targets histone H3 methylation and HP1 to promoters. Nature. 2001;412:561-5.

Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins. 2005;61 Suppl 7:176-82.

Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK. Comparing and combining predictors of mostly disordered proteins. Biochemistry. 2005;44:1989-2000.

Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics. 2006;7:208.

Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, et al. FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics. 2005;21:3435-8.

Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet. 2000;25:338-42.

Rubin SM, Gall AL, Zheng N, Pavletich NP. Structure of the Rb C-terminal domain bound to E2F1-DP1: a mechanism for phosphorylation-induced E2F release. Cell. 2005;123:1093-106.

Saddic LA, West LE, Aslanian A, Yates JR, 3rd, Rubin SM, Gozani O, et al. Methylation of the retinoblastoma tumor suppressor by SMYD2. J Biol Chem. 2010;285:37733-40.

Sanchez Y, Bachant J, Wang H, Hu F, Liu D, Tetzlaff M, et al. Control of the DNA damage checkpoint by chk1 and rad53 protein kinases through distinct mechanisms. Science. 1999;286:1166-71.

Shu F, Lv S, Qin Y, Ma X, Wang X, Peng X, et al. Functional characterization of human PFTK1 as a cyclin-dependent kinase. Proc Natl Acad Sci U S A. 2007;104:9248-53.

Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, et al. DisProt: the Database of Disordered Proteins. Nucleic Acids Res. 2007;35:D786-93.

Simone C, Bagella L, Bellan C, Giordano A. Physical interaction between pRb and cdk9/cyclinT2 complex. Oncogene. 2002;21:4158-65.

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 2011;27:431-2.

Sonnhammer EL, Eddy SR, Durbin R. Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins. 1997;28:405-20.

Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006;34:D535-9.

Sun JM, Chen HY, Davie JR. Differential distribution of unmodified and phosphorylated histone deacetylase 2 in chromatin. J Biol Chem. 2007;282:33227-36.

Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6:e21800.

Suzuma K, Takahara N, Suzuma I, Isshiki K, Ueki K, Leitges M, et al. Characterization of protein kinase C beta isoform's action on retinoblastoma protein phosphorylation, vascular endothelial growth factor-induced endothelial cell proliferation, and retinal neovascularization. Proc Natl Acad Sci U S A. 2002;99:721-6.

Taberner FJ, Quilis I, Igual JC. Spatial regulation of the start repressor Whi5. Cell Cycle. 2009;8:3010-8.

Takemura M. Evolutionary history of the retinoblastoma gene from archaea to eukarya. Biosystems. 2005;82:266-72.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28:2731-9.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673-80.

Tompa P. Structural disorder in amyloid fibrils: its implication in dynamic interactions of proteins. FEBS J. 2009;276:5406-15.

Turner B, Razick S, Turinsky AL, Vlasblom J, Crowdy EK, Cho E, et al. iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. Database (Oxford). 2010;2010:baq023.

Uchida C, Miwa S, Isobe T, Kitagawa K, Hattori T, Oda T, et al. Effects of MdmX on Mdm2-mediated downregulation of pRB. FEBS Lett. 2006;580:1753-8.

Uversky VN. What does it mean to be natively unfolded? Eur J Biochem. 2002;269:2-12.

Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins. 2000;41:415-27.

Vacic V, Uversky VN, Dunker AK, Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. BMC Bioinformatics. 2007;8:211.

Vihinen M. Relationship of protein flexibility to thermostability. Protein Eng. 1987;1:477-80.

Vihinen M, Torkkila E, Riikonen P. Accuracy of protein flexibility predictions. Proteins. 1994;19:141-9.

Wagner MV, Smolka MB, de Bruin RA, Zhou H, Wittenberg C, Dowdy SF. Whi5 regulation by site specific CDK-phosphorylation in *Saccharomyces cerevisiae*. PLoS One. 2009;4:e4300.

Welch PJ, Wang JY. A C-terminal protein-binding domain in the retinoblastoma protein regulates nuclear c-Abl tyrosine kinase in the cell cycle. Cell. 1993;75:779-90.

Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999;112:531-52.

Wirt SE, Sage J. p107 in the public eye: an Rb understudy and more. Cell Div. 2010;5:9.

Wong S, Weber JD. Deacetylation of the retinoblastoma tumour suppressor protein by SIRT1. Biochem J. 2007;407:451-60.

Wu CL, Zukerberg LR, Ngwu C, Harlow E, Lees JA. In vivo association of E2F and DP family proteins. Mol Cell Biol. 1995;15:2536-46.

Xia X, Cheng A, Akinmade D, Hamburger AW. The N-terminal 24 amino acids of the p55 gamma regulatory subunit of phosphoinositide 3-kinase binds Rb and induces cell cycle arrest. Mol Cell Biol. 2003;23:1717-25.

Xiao ZX, Ginsberg D, Ewen M, Livingston DM. Regulation of the retinoblastoma protein-related protein p107 by G1 cyclin-associated kinases. Proc Natl Acad Sci U S A. 1996;93:4633-7.

Xue Y, Li A, Wang L, Feng H, Yao X. PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. BMC Bioinformatics. 2006;7:163.

Xue Y, Ren J, Gao X, Jin C, Wen L, Yao X. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. Mol Cell Proteomics. 2008;7:1598-608.

Yang R, Muller C, Huynh V, Fung YK, Yee AS, Koeffler HP. Functions of cyclin A1 in the cell cycle and its interactions with transcription factor E2F-1 and the Rb family of proteins. Mol Cell Biol. 1999;19:2400-7.

Zhang J, Schneider C, Ottmers L, Rodriguez R, Day A, Markwardt J, et al. Genomic scale mutant hunt identifies cell size homeostasis genes in S. cerevisiae. Curr Biol. 2002;12:1992-2001.

Chapter 4

Alberghina L, Mavelli G, Drovandi G, Palumbo P, Pessina S, Tripodi F, et al. Cell growth and cell cycle in *Saccharomyces cerevisiae*: basic regulatory design and protein-protein interaction network. Biotechnol Adv. 2012;30:52-72.

Barberis M, Klipp E, Vanoni M, Alberghina L. Cell size at S phase initiation: an emergent property of the G1/S network. PLoS Comput Biol. 2007;3:e64.

Bhalla US, Iyengar R. Emergent properties of networks of biological signaling pathways. Science. 1999;283:381-7.

Borg M, Mittag T, Pawson T, Tyers M, Forman-Kay JD, Chan HS. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. Proc Natl Acad Sci U S A. 2007;104:9650-5.

Costanzo M, Nishikawa JL, Tang X, Millman JS, Schub O, Breitkreuz K, et al. CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. Cell. 2004;117:899-913.

Costanzo M, Schub O, Andrews B. G1 transcription factors are differentially regulated in *Saccharomyces cerevisiae* by the Swi6-binding protein Stb1. Mol Cell Biol. 2003;23:5064-77.

Cross FR, Buchler NE, Skotheim JM. Evolution of networks and sequences in eukaryotic cell cycle control. Philos Trans R Soc Lond B Biol Sci. 2011;366:3532-44.

de Bruin RA, McDonald WH, Kalashnikova TI, Yates J, 3rd, Wittenberg C. Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. Cell. 2004;117:887-98.

Doncic A, Falleur-Fettig M, Skotheim JM. Distinct interactions select and maintain a specific cell fate. Mol Cell. 2011;43:528-39.

Eser U, Falleur-Fettig M, Johnson A, Skotheim JM. Commitment to a cellular transition precedes genome-wide transcriptional change. Mol Cell. 2011;43:515-27.

Huang D, Kaluarachchi S, van Dyk D, Friesen H, Sopko R, Ye W, et al. Dual regulation by pairs of cyclin-dependent protein kinases and histone deacetylases controls G1 transcription in budding yeast. PLoS Biol. 2009;7:e1000188.

Kaizu K, Ghosh S, Matsuoka Y, Moriya H, Shimizu-Yoshida Y, Kitano H. A comprehensive molecular interaction map of the budding yeast cell cycle. Mol Syst Biol. 2010;6:415.

Kiemer L, Cesareni G. Comparative interactomics: comparing apples and pears? Trends Biotechnol. 2007;25:448-54.

Kitano H. Grand challenges in systems physiology. Front Physiol. 2010;1:3.

Klein P, Pawson T, Tyers M. Mathematical modeling suggests cooperative interactions between a disordered polyvalent ligand and a single receptor site. Curr Biol. 2003;13:1669-78.

Koch C, Schleiffer A, Ammerer G, Nasmyth K. Switching transcription on and off during the yeast cell cycle: Cln/Cdc28 kinases activate bound transcription factor SBF (Swi4/Swi6) at start, whereas Clb/Cdc28 kinases displace it from the promoter in G2. Genes Dev. 1996;10:129-41.

Lin YY, Lu JY, Zhang J, Walter W, Dang W, Wan J, et al. Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. Cell. 2009;136:1073-84.

Mittag T, Orlicky S, Choy WY, Tang X, Lin H, Sicheri F, et al. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. Proc Natl Acad Sci U S A. 2008;105:17772-7.

Sacco E, Farina M, Greco C, Lamperti S, Busti S, Degioia L, et al. Regulation of hSos1 activity is a system-level property generated by its multi-domain structure. Biotechnol Adv. 2012;30:154-68.

Sidorova JM, Mikesell GE, Breeden LL. Cell cycle-regulated phosphorylation of Swi6 controls its nuclear localization. Mol Biol Cell. 1995;6:1641-58.

Taberner FJ, Quilis I, Igual JC. Spatial regulation of the start repressor Whi5. Cell Cycle. 2009;8:3010-8.

Takahata S, Yu Y, Stillman DJ. The E2F functional analogue SBF recruits the Rpd3(L) HDAC, via Whi5 and Stb1, and the FACT chromatin reorganizer, to yeast G1 cyclin promoters. EMBO J. 2009;28:3378-89.

Uversky VN. Intrinsically disordered proteins may escape unwanted interactions via functional misfolding. Biochim Biophys Acta. 2011;1814:693-712.

Wagner MV, Smolka MB, de Bruin RA, Zhou H, Wittenberg C, Dowdy SF. Whi5 regulation by site specific CDK-phosphorylation in *Saccharomyces cerevisiae*. PLoS One. 2009;4:e4300.

Wittmeyer J, Joss L, Formosa T. Spt16 and Pob3 of *Saccharomyces cerevisiae* form an essential, abundant heterodimer that is nuclear, chromatin-associated, and copurifies with DNA polymerase alpha. Biochemistry. 1999;38:8961-71.

Xue Y, Li A, Wang L, Feng H, Yao X. PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. BMC Bioinformatics. 2006;7:163.

Appendices

APPENDIX A

List of Abbreviations

Abbreviations	*Explanations*
ARS	Autonomously Replicating Sequence
ATP	Adenosine Tri Phosphate
Chk1	Checkpoint kinase 1
Cdc	cell division cycle
CDK	cyclin-dependent kinase
Cdt	Cdc10-dependent transcription
CMG	(CDC45-MCM-GINS Complex)
Ctf4	chromosome transmission fidelity protein 4
DDK	Dbf4-dependent kinase
DNA Pol	DNA polymerase
Dpb	DNA polymerase B subunit
GEMC1	Geminin coiled-coil containing protein 1
GINS	Go, Ichi, Nii, and San
MCM	mini-chromosome maintenance
ORC	origin recognition complex
Pre-IC	Pre-Initiation Complex
Pre-LC	Pre-Loading Complex
Pre-RC	Pre-Replication Complex
Psf	Partner of Sld five
RPA	replication protein A
RPC	Replication Progression Complex
Sld	synthetically lethal with Dpb11-1
ssDNA	single-stranded DNA
TopBP1	DNA topoisomerase II binding protein

APPENDIX B

Table 1: List of top fifteen sequences derived from BLAST search

Accession	Source	Description	Seq length	Max ident	% Query cover age	e-value
CCF57878.1	*Kazachstaniaafricana*	Hypot. protein KAFR_0D02310	302	47%	52	9e-32
NP_984329.1	*Ashbyagossypii ATCC 10895*	ADR233Wp	313	46%	47	3e-27
XP_002553543.1	*Lachanceathermotolerans CBS 6340*	KLTH0E01254p	319	47%	45	4e-27
CCK72823.1	*Kazachstanianagaishii CBS 8797*	KNAG_0L02050	359	42%	48	1e-24
XP_002498613.1	*Zygosaccharomycesrouxii*	ZYRO0G14564p	465	38%	66%	2e-24
XP_001645731.1	*Vanderwaltozymapolyspora DSM 70294*	hypothetical protein Kpol_1043p64	404	72%	24%	3e-24
XP_003677271.1	*Naumovozymacastellii CBS 4309*	hypothetical protein NCAS_0G00310	369	47%	35%	7e-24
XP_003679570.1	*Torulasporadelbrueckii*	hypothetical protein TDEL_0B02300	410	38%	43%	2e-23
XP_448471.1	*Candida glabrata CBS 138*	hypothetical protein	430	67%	24%	2e-22
XP_003685559.1	*Tetrapisisporaphaffii CBS 4417*	hypothetical protein TPHA_0E00290	419	37%	43%	2e-21
XP_003670599.1	*Naumovozymadairenensis CBS 421*	hypothetical protein NDAI_0F00370	469	43%	33%	2e-15
XP_454541.1	*Kluyveromyceslactis NRRL Y-1140*	hypothetical protein	370	32%	44%	1e-10
CCH58702.1	*Tetrapisisporablattae CBS 6284*	hypothetical protein TBLA_0A09140	554	42%	25%	4e-07
CCH44769.1	*Wickerhamomycesciferrii*	Muscle M-line assembly protein unc-89	422	44%	23%	1e-05

Table 2: pI value of Whi5 proteins and their motifs

Source	pI of entire protein	pI of Motif 1 (disordered)	pI of Motif 2 (α-helix in Whi5 family signature)	pI of Motif 3 (disordered)
S. cerevisiae	6.28	12.00	8.50	5.50
V. polyspora	8.37	11.10	8.38	4.59
L. thermotolerans	9.48	11.10	8.38	4.49
C. glabrata	7.01	11.72	8.34	6.23
Z. rouxii	9.38	11.10	9.70	4.25
A. gossypii	6.41	11.71	8.38	4.48
K. lactis	6.33	5.46	9.99	4.75
T.phaffii	5.95	10.46	9.70	4.41
K.africana	9.20	11.17	8.38	8.50
N. castellii	5.82	10.46	8.38	4.33
T. delbrueckii	6.14	11.00	8.38	4.41
K. naganishii	8.71	9.99	8.34	3.43
N. dairenensis	9.28	9.73	8.43	4.33
T. blattae	9.56	11.73	6.24	4.41
W. ciferrii	5.38	-	9.44	4.25

Table 3: Available structures of human Rb protein

	PDB ID	Resolution (Å)	Chain	Positions
1.	1AD6	2.30	A	378-562
2.	1GH6	3.20	B	379-772
3.	1GUX	1.85	A	372-589
			B	636-787
4.	1H25	2.50	E	868-878
5.	1N4M	2.20	A/B	380-785
6.	1O9K	2.60	A/C/E/G	372-589
			B/D/F/H	636-787
7.	1PJM	2.50	A	860-876
8.	2AZE	2.55	C	829-874
9.	2QDJ	2.00	A	52-355
10.	2R7G	1.67	A/C	380-787
11.	3N5U	3.20	C	870-882
12.	3POM	2.50	A/B	380-787
13.	4ELJ	2.70	A	53-787
14.	4ELL	1.98	A/B	380-787

Table 4: Interactors of human Rb protein

Gene Name	Protein Name	Description	Functional classification in Figure
AATF	apoptosis antagonizing transcription factor	It contains a leucine zipper, which is a characteristic motif of transcription factors, and was shown to exhibit strong transactivation activity when fused to Gal4 DNA binding domain. Overexpression of this gene interfered with MAP3K12 induced apoptosis.	Apoptosis
ABL1	c-abl oncogene 1, non-receptor tyrosine kinase	The ABL1 protooncogene encodes a cytoplasmic and nuclear protein tyrosine kinase that has been implicated in processes of cell differentiation, cell division, cell adhesion, and stress response. Activity of c-Abl protein is negatively regulated by its SH3 domain, and deletion of the SH3 domain turns ABL1 into an oncogene. The t(9;22) translocation results in the head-to-tail fusion of the BCR (MIM:151410) and ABL1 genes present in many cases of chronic myelogeneous leukemia. The DNA-binding activity of the ubiquitously expressed ABL1 tyrosine kinase is regulated by CDC2-mediated phosphorylation, suggesting a cell cycle function for ABL1.	Signaling/stress response
AHR	aryl hydrocarbon receptor	It is a ligand-activated transcription factor involved in the regulation of biological responses to planar aromatic hydrocarbons. This receptor has been shown to regulate xenobiotic-metabolizing enzymes such as cytochrome P450. Its ligands included a variety of aromatic hydrocarbons.	Apoptosis
ANAPC2	anaphase promoting complex subunit 2	A large protein complex, termed the anaphase-promoting complex (APC), or the cyclosome, promotes metaphase-anaphase transition by ubiquitinating its specific substrates such as mitotic cyclins and anaphase inhibitor, which are subsequently degraded by the 26S proteasome.	Apoptosis
AR	androgen receptor	The androgen receptor gene is more than 90 kb long and codes for a protein that has 3 major functional domains: the N-terminal domain, DNA-binding domain, and androgen-binding domain. The protein functions as a steroid-hormone activated transcription factor. Upon binding the hormone ligand, the receptor dissociates from accessory proteins, translocates into the nucleus, dimerizes, and then stimulates transcription of androgen responsive genes.	G1/S and G2/M cell cycle progression/ meiosis
ARID3B	AT rich interactive domain 3B (BRIGHT-like)	It is a member of the ARID (AT-rich interaction domain) family of DNA-binding proteins. Members of the ARID family have roles in embryonic patterning, cell lineage gene regulation, cell cycle control, transcriptional regulation and possibly in chromatin structure modification.	Chromatin/transcription
ARID4A	AT rich interactive domain 4A (RBP1-like)	It binds directly, with several other proteins, to retinoblastoma protein (pRB). This protein, in turn, serves as a bridging molecule to recruit HDACs and, in addition, provides a second HDAC-independent repression function. The protein possesses transcriptional repression activity.	Chromatin/transcription
ARNT	aryl hydrocarbon receptor nuclear translocator	The aryl hydrocarbon (Ah) receptor is involved in the induction of several enzymes that participate in xenobiotic metabolism. The ligand-free, cytosolic form of the Ah receptor is complexed to heat shock protein 90. Binding of ligand, which includes dioxin and polycyclic aromatic hydrocarbons, results in translocation of the ligand-binding subunit only to the nucleus. Induction of enzymes involved in xenobiotic metabolism occurs through binding of the ligand-bound Ah receptor to xenobiotic responsive elements in the promoters of genes for these enzymes.	Chromatin/transcription
ATF2	activating transcription factor 2	It is a member of the leucine zipper family of DNA binding proteins. This protein binds to the cAMP-responsive element (CRE), an octameric palindrome. It forms a homodimer or a heterodimer with c-Jun and stimulates CRE-dependent transcription. This protein is also a histone acetyltransferase (HAT) that specifically acetylates histones H2B and H4 in vitro.	Chromatin/transcription
AURKB	aurora kinase B	It is a member of the aurora kinase subfamily of serine/threonine kinases. These kinases participate in the regulation of segregation of chromosomes during mitosis and meiosis through association with microtubules.	chromosome segregation/kinetochore/spindle/microtubule
BAG1	BCL2-associated athanogene	The oncogene BCL2 is a membrane protein that blocks a step in a pathway leading to apoptosis or programmed cell death. The protein binds to BCL2 and is referred to as BCL2-associated athanogene. It enhances the anti-apoptotic effects of BCL2 and represents a link between growth factor receptors and anti-apoptotic mechanisms.	Apoptosis
BCR	breakpoint cluster region	It is involved in Philadelphia chromosome. Although the BCR-ABL fusion protein has been extensively studied, the function of the normal BCR gene product is not clear. The protein has serine/threonine kinase activity and is a GTPase-activating protein for p21rac.	Unknown
BRCA1	breast cancer 1, early onset	It is a nuclear phosphoprotein that plays a role in maintaining genomic stability, and it also acts as a tumor suppressor. The protein combines with other tumor suppressors, DNA damage sensors, and signal transducers to form a large multi-subunit protein complex known as the BRCA1-associated genome surveillance complex (BASC). This protein associates with RNA polymerase II, and through the C-terminal domain, also interacts with histone deacetylase complexes. This protein thus plays a role in transcription, DNA repair of double-stranded breaks, and recombination.	DNA replication/repair/HR/cohesion
BRF1	BRF1 homolog, subunit of RNA polymerase III transcription initiation factor IIIB (S. cerevisiae)	It is one of the three subunits of the RNA polymerase III transcription factor complex. This complex plays a central role in transcription initiation by RNA polymerase III on genes encoding tRNA, 5S rRNA, and other small structural RNAs. The gene product belongs to the TF2B family.	Chromatin/transcription
CBX1	chromobox homolog 1	It is a highly conserved nonhistone protein, which is a member of the heterochromatin protein family. The protein is enriched in the	Chromatin/transcription

		heterochromatin and associated with centromeres. The protein has a single N-terminal chromodomain which can bind to histone proteins via methylated lysine residues, and a C-terminal chromo shadow-domain (CSD) which is responsible for the homodimerization and interaction with a number of chromatin-associated nonhistone proteins.	
CBX4	chromobox homolog 4	Same as CBX1	Chromatin/transcription
CBX5	chromobox homolog 5	Same as CBX1	Chromatin/transcription
CCNA1	cyclin A1	The cyclin encoded by this gene was shown to be expressed in testis and brain, as well as in several leukemic cell lines, and is thought to primarily function in the control of the germline meiotic cell cycle. This cyclin binds both CDK2 and CDC2 kinases, which give two distinct kinase activities, one appearing in S phase, the other in G2, and thus regulate separate functions in cell cycle. This cyclin was found to bind to important cell cycle regulators, such as Rb family proteins, transcription factor E2F-1, and the p21 family proteins.	G1/S and G2/M cell cycle progression/ meiosis
CCNA2	cyclin A2	In contrast to cyclin A1, which is present only in germ cells, this cyclin is expressed in all tissues tested. This cyclin binds and activates CDC2 or CDK2 kinases, and thus promotes both cell cycle G1/S and G2/M transitions.	G1/S and G2/M cell cycle progression/ meiosis
CCND1	cyclin D1	This cyclin forms a complex with and functions as a regulatory subunit of CDK4 or CDK6, whose activity is required for cell cycle G1/S transition. This protein has been shown to interact with tumor suppressor protein Rb and the expression of this gene is regulated positively by Rb.	G1/S and G2/M cell cycle progression/ meiosis
CCND2	cyclin D2	Same as CCND1. Knockout studies of the homologous gene in mouse suggest the essential roles of this gene in ovarian granulosa and germ cell proliferation. High level expression of this gene was observed in ovarian and testicular tumors.	G1/S and G2/M cell cycle progression/ meiosis
CCND3	cyclin D3	Same as CCND1. The CDK4 activity associated with this cyclin was reported to be necessary for cell cycle progression through G2 phase into mitosis after UV radiation. Several transcript variants encoding different isoforms have been found for this gene.	G1/S and G2/M cell cycle progression/ meiosis
CCNE1	cyclin E1	This cyclin forms a complex with and functions as a regulatory subunit of CDK2, whose activity is required for cell cycle G1/S transition. This protein accumulates at the G1-S phase boundary and is degraded as cells progress through S phase. Overexpression of this gene has been observed in many tumors, which results in chromosome instability, and thus may contribute to tumorigenesis. This protein was found to associate with, and be involved in, the phosphorylation of NPAT protein (nuclear protein mapped to the ATM locus), which participates in cell-cycle regulated histone gene expression and plays a critical role in promoting cell-cycle progression in the absence of pRb.	G1/S and G2/M cell cycle progression/ meiosis
CCNT2	cyclin T2	This cyclin and its kinase partner CDK9 were found to be subunits of the transcription elongation factor p-TEFb. The p-TEFb complex containing this cyclin was reported to interact with, and act as a negative regulator of human immunodeficiency virus type 1 (HIV-1) Tat protein.	Chromatin/transcription
CDC16	cell division cycle 16 homolog (S. cerevisiae)	It is a component protein of the APC complex, which is composed of eight proteins and functions as a protein ubiquitin ligase. This protein and two other APC complex proteins, CDC23 and CDC27, contain a tetratricopeptide repeat (TPR), a protein domain that may be involved in protein-protein interaction.	protein degradation/ proteosome
CDC27	cell division cycle 27 homolog (S. cerevisiae)	This protein is a component of anaphase-promoting complex (APC), which is composed of eight protein subunits and highly conserved in eucaryotic cells. APC catalyzes the formation of cyclin B-ubiquitin conjugate that is responsible for the ubiquitin-mediated proteolysis of B-type cyclins. This protein and 3 other members of the APC complex contain the TPR (tetratricopeptide repeat), a protein domain important for protein-protein interaction. This protein was shown to interact with mitotic checkpoint proteins including Mad2, p55CDC and BUBR1, and thus may be involved in controlling the timing of mitosis.	protein degradation/ proteosome
CDK1	cyclin-dependent kinase 1	This protein is a catalytic subunit of the highly conserved protein kinase complex known as M-phase promoting factor (MPF), which is essential for G1/S and G2/M phase transitions of eukaryotic cell cycle. Mitotic cyclins stably associate with this protein and function as regulatory subunits. The kinase activity of this protein is controlled by cyclin accumulation and destruction through the cell cycle. The phosphorylation and dephosphorylation of this protein also play important regulatory roles in cell cycle control.	G1/S and G2/M cell cycle progression/ meiosis
CDK2	cyclin-dependent kinase 2	This protein kinase is highly similar to the gene products of S. cerevisiae cdc28, and S. pombe cdc2. It is a catalytic subunit of the cyclin-dependent protein kinase complex, whose activity is restricted to the G1-S phase, and essential for cell cycle G1/S phase transition. This protein associates with and regulated by the regulatory subunits of the complex including cyclin A or E, CDK inhibitor p21Cip1 (CDKN1A) and p27Kip1 (CDKN1B). Its activity is also regulated by its protein phosphorylation.	G1/S and G2/M cell cycle progression/ meiosis
CDK4	cyclin-dependent kinase 4	This protein is highly similar to the gene products of *S. cerevisiae* cdc28 and S. pombe cdc2. It is a catalytic subunit of the protein kinase complex that is important for cell cycle G1 phase progression. The activity of this kinase is restricted to the G1-S phase, which is controlled by the regulatory subunits D-type cyclins and CDK inhibitor p16(INK4a). This kinase was shown to be responsible for the phosphorylation pRb. Mutations in this gene as well as in its related proteins including D-type cyclins, p16(INK4a) and Rb were all found to be associated with tumorigenesis of a variety of cancers.	G1/S and G2/M cell cycle progression/ meiosis

CDK5	cyclin-dependent kinase 5	Proline-directed serine/threonine-protein kinase essential for neuronal cell cycle arrest and differentiation and may be involved in apoptotic cell death in neuronal diseases by triggering abortive cell cycle re-entry. Interacts with D1 and D3-type G1 cyclins.	Cell Differentiation
CDK6	cyclin-dependent kinase 6	This kinase is a catalytic subunit of the protein kinase complex that is important for cell cycle G1 phase progression and G1/S transition. The activity of this kinase first appears in mid-G1 phase, which is controlled by the regulatory subunits including D-type cyclins and members of INK4 family of CDK inhibitors. This kinase, as well as CDK4, has been shown to phosphorylate, and thus regulate the activity of, tumor suppressor protein Rb. Expression of this gene is up-regulated in some types of cancer.	G1/S and G2/M cell cycle progression/ meiosis
CDK9	cyclin-dependent kinase 9	This kinase was found to be a component of the multiprotein complex TAK/P-TEFb, which is an elongation factor for RNA polymerase II-directed transcription and functions by phosphorylating the C-terminal domain of the largest subunit of RNA polymerase II. This protein forms a complex with and is regulated by its regulatory subunit cyclin T or cyclin K. HIV-1 Tat protein was found to interact with this protein and cyclin T, which suggested a possible involvement of this protein in AIDS.	Chromatin/transcription
CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha	The protein encoded by this intronless gene is a bZIP transcription factor which can bind as a homodimer to certain promoters and enhancers. It can also form heterodimers with the related proteins CEBP-beta and CEBP-gamma. The protein has been shown to bind to the promoter and modulate the expression of the gene encoding leptin, a protein that plays an important role in body weight homeostasis. Also, the protein can interact with CDK2 and CDK4, thereby inhibiting these kinases and causing growth arrest in cultured cells.	Immune Response
CEBPB	CCAAT/enhancer binding protein (C/EBP), beta	The encoded protein is important in the regulation of genes involved in immune and inflammatory responses and has been shown to bind to the IL-1 response element in the IL-6 gene, as well as to regulatory regions of several acute-phase and cytokine genes. In addition, the encoded protein can bind the promoter and upstream element and stimulate the expression of the collagen type I gene.	Immune Response
CREG1	cellular repressor of E1A-stimulated genes 1	The adenovirus E1A protein both activates and represses gene expression to promote cellular proliferation and inhibit differentiation. The protein encoded by this gene antagonizes transcriptional activation and cellular transformation by E1A. This protein shares limited sequence similarity with E1A and binds both the general transcription factor TBP and the tumor suppressor pRb in vitro.	Cell Differentiation
CUX1	cut-like homeobox 1	The protein encoded by this gene is a member of the homeodomain family of DNA binding proteins. It may regulate gene expression, morphogenesis, and differentiation and it may also play a role in the cell cycle progession.	Immune Response
DNAJA2	DnaJ (Hsp40) homolog, subfamily A, member 2	The protein encoded by this gene belongs to the evolutionarily conserved DNAJ/HSP40 family of proteins, which regulate molecular chaperone activity by stimulating ATPase activity. DNAJ proteins may have up to 3 distinct domains: a conserved 70-amino acid J domain, usually at the N terminus; a glycine/phenylalanine (G/F)-rich region; and a cysteine-rich domain containing 4 motifs resembling a zinc finger domain. The product of this gene works as a cochaperone of Hsp70s in protein folding and mitochondrial protein import in vitro.	protein folding/protein glycosylation /cell wall biogenesis&integrity
DNMT1	DNA (cytosine-5-)-methyltransferase 1	DNA (cytosine-5-)-methyltransferase 1 has a role in the establishment and regulation of tissue-specific patterns of methylated cytosine residues. Aberrant methylation patterns are associated with certain human tumors and developmental abnormalities.	Chromatin/transcription
DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha	It is a DNA methyltransferase that is thought to function in de novo methylation, rather than maintenance methylation. The protein localizes to the cytoplasm and nucleus and its expression is developmentally regulated.	Chromatin/transcription
E2F1	E2F transcription factor 1	The E2F family plays a crucial role in the control of cell cycle and action of tumor suppressor proteins and is also a target of the transforming proteins of small DNA tumor viruses. The E2F proteins contain several evolutionally conserved domains found in most members of the family. These domains include a DNA binding domain, a dimerization domain which determines interaction with the differentiation regulated transcription factor proteins (DP), a transactivation domain enriched in acidic amino acids, and a tumor suppressor protein association domain which is embedded within the transactivation domain. This protein binds preferentially to retinoblastoma protein pRB in a cell-cycle dependent manner. It can mediate both cell proliferation and p53-dependent/independent apoptosis.	G1/S and G2/M cell cycle progression/ meiosis
E2F2	E2F transcription factor 2	Same as E2F1. This protein binds specifically to retinoblastoma protein pRB in a cell-cycle dependent manner, and it exhibits overall 46% amino acid identity to E2F1.	G1/S and G2/M cell cycle progression/ meiosis
E2F4	E2F transcription factor 4, p107/p130-binding	Same as E2F1. This protein binds to all three of the tumor suppressor proteins pRB, p107 and p130, but with higher affinity to the last two. It plays an important role in the suppression of proliferation-associated genes, and its gene mutation and increased expression may be associated with human cancer.	G1/S and G2/M cell cycle progression/ meiosis
E4F1	E4F transcription factor 1	The zinc finger protein encoded by this gene is one of several cellular transcription factors whose DNA-binding activities are regulated through the action of adenovirus E1A. A 50-kDa amino-terminal product is generated from the full-length protein through proteolytic cleavage. The protein is differentially regulated by E1A-induced phosphorylation. The full-length gene product represses transcription from the E4 promoter in the absence of E1A, while the 50-kDa form acts as a transcriptional activator in its presence.	G1/S and G2/M cell cycle progression/ meiosis
EID1	EP300 interacting	Interacts with pRb and EP300 and acts as a repressor of MYOD1 transactivation. Inhibits EP300 and CBP histone acetyltransferase activity. May be involved in coupling cell cycle exit to the transcriptional activation of genes	Cell Differentiation

	inhibitor of differentiation 1	required for cellular differentiation. May act as a candidate coinhibitory factor for NR0B2 that can be directly linked to transcription inhibitory mechanisms.	
ENC1	ectodermal-neural cortex 1 (with BTB-like domain)	It is a member of the kelch-related family of actin-binding proteins. The encoded protein plays a role in the oxidative stress response as a regulator of the transcription factor Nrf2, and expression of this gene may play a role in malignant transformation..	Cell Differentiation
FOXM1	forkhead box M1	The protein encoded by this gene is a transcriptional activator involved in cell proliferation. The encoded protein is phosphorylated in M phase and regulates the expression of several cell cycle genes, such as cyclin B1 and cyclin D1. Several transcript variants encoding different isoforms have been found for this gene.	G1/S and G2/M cell cycle progression/ meiosis
FRK	fyn-related kinase	The protein encoded by this gene belongs to the TYR family of protein kinases. This tyrosine kinase is a nuclear protein and may function during G1 and S phase of the cell cycle and suppress growth.	G1/S and G2/M cell cycle progression/ meiosis
FZR1	fizzy/cell division cycle 20 related 1 (Drosophila)	Key regulator of ligase activity of the anaphase promoting complex/cyclosome (APC/C), which confers substrate specificity upon the complex. Associates with the APC/C in late mitosis, in replacement of CDC20, and activates the APC/C during anaphase and telophase. At the G1/S transition FZR1 is phosphorylated, leading to its dissociation from the APC/C. Following DNA damage, it is required for the G2 DNA damage checkpoint: its dephosphorylation and reassociation with the APC/C leads to the ubiquitination of PLK1, preventing entry into mitosis.	protein degradation/ proteosome
GNB2L1	guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1	Involved in the recruitment, assembly and/or regulation of a variety of signaling molecules. Interacts with a wide variety of proteins and plays a role in many cellular processes. Component of the 40S ribosomal subunit involved in translational repression. Binds to and stabilizes activated protein kinase C (PKC), increasing PKC-mediated phosphorylation. May recruit activated PKC to the ribosome, leading to phosphorylation of EIF6.	Unknown
GTF3C1	general transcription factor IIIC, polypeptide 1, alpha 220kDa	Required for RNA polymerase III-mediated transcription. Component of TFIIIC that initiates transcription complex assembly on tRNA and is required for transcription of 5S rRNA and other stable nuclear and cytoplasmic RNAs. Binds to the box B promoter element.	Chromatin/transcription
HBP1	HMG-box transcription factor 1	Transcriptional repressor that binds to the promoter region of target genes. Plays a role in the regulation of the cell cycle and of the Wnt pathway. Binds preferentially to the sequence 5'-TTCATTCATTCA-3'. Binding to the H1F0 promoter is enhanced by interaction with RB1. Disrupts the interaction between DNA and TCF4.	G1/S and G2/M cell cycle progression/ meiosis
HDAC1	histone deacetylase 1	The protein belongs to the histone deacetylase/acuc/apha family and is a component of the histone deacetylase complex. It also interacts with retinoblastoma tumor-suppressor protein and this complex is a key element in the control of cell proliferation and differentiation. Together with metastasis-associated protein-2, it deacetylates p53 and modulates its effect on cell growth and apoptosis.	Chromatin/transcription
HDAC2	histone deacetylase 2	The protein belongs to the histone deacetylase family. Histone deacetylases act via the formation of large multiprotein complexes, and are responsible for the deacetylation of lysine residues at the N-terminal regions of core histones (H2A, H2B, H3 and H4). This protein forms transcriptional repressor complexes by associating with many different proteins, including YY1, a mammalian zinc-finger transcription factor. Thus, it plays an important role in transcriptional regulation, cell cycle progression and developmental events.	Chromatin/transcription
HDAC3	histone deacetylase 3	The protein belongs to the histone deacetylase/acuc/apha family. It has histone deacetylase activity and represses transcription when tethered to a promoter. It may participate in the regulation of transcription through its binding with the zinc-finger transcription factor YY1. This protein can also down-regulate p53 function and thus modulate cell growth and apoptosis.	Chromatin/transcription
HMGA2	high mobility group AT-hook 2	It is a protein that belongs to the non-histone chromosomal high mobility group (HMG) protein family. HMG proteins function as architectural factors and are essential components of the enhancesome. This protein contains structural DNA-binding domains and may act as a transcriptional regulating factor. Identification of the deletion, amplification, and rearrangement of this gene that are associated with myxoid liposarcoma suggests a role in adipogenesis and mesenchymal differentiation. A gene knock out study of the mouse counterpart demonstrated that this gene is involved in diet-induced obesity.	Chromatin/transcription
HMGB1	high mobility group box 1	DNA binding proteins that associates with chromatin and has the ability to bend DNA. Binds preferentially single-stranded DNA. Involved in V(D)J recombination by acting as a cofactor of the RAG complex. Acts by stimulating cleavage and RAG protein binding at the 23 bp spacer of conserved recombination signal sequences (RSS).	DNA replication/repair/HR/cohesion
IRF3	interferon regulatory factor 3	It is a member of the interferon regulatory transcription factor (IRF) family. The encoded protein is found in an inactive cytoplasmic form that upon serine/threonine phosphorylation forms a complex with CREBBP. This complex translocates to the nucleus and activates the transcription of interferons alpha and beta, as well as other interferon-induced genes.	signaling/stress response
JUN	jun proto-oncogene	The protein which is highly similar to the viral protein, and which interacts directly with specific target DNA sequences to regulate gene expression. This gene is intronless and is mapped to 1p32-p31, a chromosomal region involved in both translocations and deletions in human malignancies.	Chromatin/transcription
KAT2B	K(lysine) acetyltransferase 2B	CBP and p300 are large nuclear proteins that bind to many sequence-specific factors involved in cell growth and/or differentiation, including c-jun and the adenoviral oncoprotein E1A. The protein encoded by this gene associates with p300/CBP. It has in vitro and in vivo binding activity with CBP and p300, and	Chromatin/transcription

		competes with E1A for binding sites in p300/CBP. It has histone acetyl transferase activity with core histones and nucleosome core particles, indicating that this protein plays a direct role in transcriptional regulation.	
KAT5	K(lysine) acetyltransferase 5	The protein belongs to the MYST family of histone acetyl transferases (HATs) and was originally isolated as an HIV-1 TAT-interactive protein. HATs play important roles in regulating chromatin remodeling, transcription and other nuclear processes by acetylating histone and nonhistone proteins. This protein is a histone acetylase that has a role in DNA repair and apoptosis and is thought to play an important role in signal transduction.	Chromatin/transcription
KDM1A	lysine (K)-specific demethylase 1A	It is a nuclear protein containing a SWIRM domain, a FAD-binding motif, and an amine oxidase domain. This protein is a component of several histone deacetylase complexes, though it silences genes by functioning as a histone demethylase.	Chromatin/transcription
KDM4A	lysine (K)-specific demethylase 4A	This gene is a member of the Jumonji domain 2 (JMJD2) family and encodes a protein containing a JmjN domain, a JmjC domain, a JD2H domain, two TUDOR domains, and two PHD-type zinc fingers. This nuclear protein functions as a trimethylation-specific demethylase, converting specific trimethylated histone residues to the dimethylated form, and as a transcriptional repressor.	Chromatin/transcription
KDM5A	lysine (K)-specific demethylase 5A	The protein is a ubiquitously expressed nuclear protein. This protein also interacts with rhombotin-2 which functions distinctly in erythropoiesis and in T-cell leukemogenesis. Rhombotin-2 is thought to either directly affect the activity of the encoded protein or may indirectly modulate the functions of the retinoblastoma protein by binding to this protein.	Chromatin/transcription
KDM5B	lysine (K)-specific demethylase 5B	Histone demethylase that demethylates 'Lys-4' of histone H3, thereby playing a central role in histone code. Does not demethylate histone H3 'Lys-9' or H3 'Lys-27'. Demethylates trimethylated, dimethylated and monomethylated H3 'Lys-4'. Acts as a transcriptional corepressor for FOXG1B and PAX9. Favors the proliferation of breast cancer cells by repressing tumor suppressor genes such as BRCA1 and HOXA5. In contrast, may act as a tumor suppressor for melanoma.	Chromatin/transcription
L3MBTL1	l(3)mbt-like 1 (Drosophila)	This gene represents a polycomb group gene. The encoded protein functions to regulate gene activity, likely via chromatin modification. The encoded protein may also be necessary for mitosis.	Chromatin/transcription
L3MBTL2	l(3)mbt-like 2 (Drosophila)	Same as L3MBTL1. Binds to monomethylated and dimethylated 'Lys-20' on histone H4. Binds histone H3 peptides that are monomethylated or dimethylated on 'Lys-4', 'Lys-9' or 'Lys-27'.	Chromatin/transcription
LDB1	LIM domain binding 1	Binds to the LIM domain of a wide variety of LIM domain-containing transcription factors. May regulate the transcriptional activity of LIM-containing proteins by determining specific partner interactions. May play a role in the development of motor neurons. Acts synergistically with LHX1/LIM1 in axis formation and activation of gene expression. Acts with LMO2 in the regulation of red blood cell development, maintaining erythroid precursors in an immature state	Chromatin/transcription
LIN54	lin-54 homolog (C. elegans)	LIN54 is a component of the LIN, or DREAM, complex, an essential regulator of cell cycle genes	G1/S and G2/M cell cycle progression/ meiosis
LIN9	lin-9 homolog (C. elegans)	LIN9 is a component of the LIN, or DREAM, complex, an essential regulator of cell cycle genes	G1/S and G2/M cell cycle progression/ meiosis
LMNA	lamin A/C	The nuclear lamina consists of a two-dimensional matrix of proteins located next to the inner nuclear membrane. The lamin family of proteins make up the matrix and are highly conserved in evolution. During mitosis, the lamina matrix is reversibly disassembled as the lamin proteins are phosphorylated. Lamin proteins are thought to be involved in nuclear stability, chromatin structure and gene expression.	Chromatin/transcription
LMO2	LIM domain only 2 (rhombotin-like 1)	LMO2 encodes a cysteine-rich, two LIM-domain protein that is required for yolk sac erythropoiesis. The LMO2 protein has a central and crucial role in hematopoietic development and is highly conserved. The LMO2 transcription start site is located approximately 25 kb downstream from the 11p13 T-cell translocation cluster (11p13 ttc), where a number T-cell acute lymphoblastic leukemia-specific translocations occur.	Chromatin/transcription
MAPK14	mitogen-activated protein kinase 14	The protein is a member of the MAP kinase family. The activation requires its phosphorylation by MAP kinase kinases (MKKs), or its autophosphorylation triggered by the interaction of MAP3K7IP1/TAB1 protein with this kinase. The substrates of this kinase include transcription regulator ATF2, MEF2C, and MAX, cell cycle regulator CDC25B, and tumor suppressor p53, which suggest the roles of this kinase in stress related transcription and cell cycle regulation, as well as in genotoxic stress response.	Cell Differentiation
MCM7	minichromosome maintenance complex component 7	The hexameric protein complex formed by the MCM proteins is a key component of the pre-replication complex (pre_RC) and may be involved in the formation of replication forks and in the recruitment of other DNA replication related proteins. The MCM complex consisting of this protein and MCM2, 4 and 6 proteins possesses DNA helicase activity, and may act as a DNA unwinding enzyme. Cyclin D1-dependent kinase, CDK4, is found to associate with this protein, and may regulate the binding of this protein with the tumorsuppressor protein RB1/RB.	DNA replication/repair/HR/cohesion
MDM2	Mdm2, p53 E3 ubiquitin protein ligase homolog (mouse)	This gene is a target gene of the transcription factor tumor protein p53. The encoded protein is a nuclear phosphoprotein that binds and inhibits transactivation by tumor protein p53, as part of an autoregulatory negative feedback loop. Overexpression of this gene can result in excessive inactivation of tumor protein p53, diminishing its tumor suppressor function. This protein has E3 ubiquitin ligase activity, which targets tumor protein p53 for proteasomal degradation. This protein also affects the cell cycle, apoptosis, and tumorigenesis through interactions with other proteins, including retinoblastoma 1 and ribosomal protein L5.	protein degradation/ proteosome

MNAT 1	menage a trois homolog 1, cyclin H assembly factor	The protein encoded by this gene, along with cyclin H and CDK7, forms the CDK-activating kinase (CAK) enzymatic complex. This complex activates several cyclin-associated kinases and can also associate with TFIIH to activate transcription by RNA polymerase II.	G1/S and G2/M cell cycle progression/ meiosis
MORF 4L1	mortality factor 4 like 1	Component of the NuA4 histone acetyltransferase (HAT) complex which is involved in transcriptional activation of select genes principally by acetylation of nucleosomal histones H4 and H2A. This modification may both alter nucleosome - DNA interactions and promote interaction of the modified histones with other proteins which positively regulate transcription. This complex may be required for the activation of transcriptional programs associated with oncogene and proto-oncogene mediated growth induction, tumor suppressor mediated growth arrest and replicative senescence, apoptosis, and DNA repair.	Chromatin/tr anscription
MORF 4L2	mortality factor 4 like 2	Same as MORF4L1.	Chromatin/tr anscription
MRFA P1	Morf4 family associated protein 1	Found in a complex composed of MORF4L1, MRFAP1 and RB1. Interacts via its N-terminus with MORF4L1. Interacts with CSTB and MORF4L2.	Chromatin/tr anscription
MYC	v-myc myelocytomato sis viral oncogene homolog (avian)	The protein encoded by this gene is a multifunctional, nuclear phosphoprotein that plays a role in cell cycle progression, apoptosis and cellular transformation. It functions as a transcription factor that regulates transcription of specific target genes. Mutations, overexpression, rearrangement and translocation of this gene have been associated with a variety of hematopoietic tumors, leukemias and lymphomas, including Burkitt lymphoma.	Chromatin/tr anscription
MYOD 1	myogenic differentiation 1	It is a nuclear protein that belongs to the basic helix-loop-helix family of transcription factors and the myogenic factors subfamily. It regulates muscle cell differentiation by inducing cell cycle arrest, a prerequisite for myogenic initiation. The protein is also involved in muscle regeneration. It activates its own transcription which may stabilize commitment to myogenesis.	Cell Differentiatio n
NCOA 6	nuclear receptor coactivator 6	The protein encoded by this gene is a transcriptional coactivator that can interact with nuclear hormone receptors to enhance their transcriptional activator functions. This protein has been shown to be involved in the hormone-dependent coactivation of several receptors, including prostanoid, retinoid, vitamin D3, thyroid hormone, and steroid receptors.	Chromatin/tr anscription
NFIX	nuclear factor I/X (CCAAT-binding transcription factor)	Recognizes and binds the palindromic sequence 5'-TTGGCNNNNNGCCAA-3' present in viral and cellular promoters and in the origin of replication of adenovirus type 2. These proteins are individually capable of activating transcription and replication.	DNA replication/re pair/HR/coh esion
NPM1	nucleophosmin (nucleolar phosphoprotein B23, numatrin)	It is a phosphoprotein which moves between the nucleus and the cytoplasm. The gene product is thought to be involved in several processes including regulation of the ARF/p53 pathway. A number of genes are fusion partners have been characterized, in particular the anaplastic lymphoma kinase gene on chromosome 2. Mutations in this gene are associated with acute myeloid leukemia.	Chromatin/tr anscription
PA2G4	proliferation-associated 2G4, 38kDa	It is an RNA-binding protein that is involved in growth regulation. This protein is present in pre-ribosomal ribonucleoprotein complexes and may be involved in ribosome assembly and the regulation of intermediate and late steps of rRNA processing. This protein can interact with the cytoplasmic domain of the ErbB3 receptor and may contribute to transducing growth regulatory signals. This protein is also a transcriptional co-repressor of androgen receptor-regulated genes and other cell cycle regulatory genes through its interactions with histone deacetylases. This protein has been implicated in growth inhibition and the induction of differentiation of human cancer cells.	G1/S and G2/M cell cycle progression/ meiosis
PAX6	paired box 6	The protein contains a homeo box domain. Both domains are known to bind DNA, and function as regulators of gene transcription. This gene is expressed in the developing nervous system, and in developing eyes. Mutations in this gene are known to cause ocular disorders such as aniridia and Peter's anomaly.	Cell Differentiatio n
PELP1	proline, glutamate and leucine rich protein 1	It is a transcription factor which coactivates transcription of estrogen receptor responsive genes and corepresses genes activated by other hormone receptors or sequence-specific transcription factors. Expression of this gene is regulated by both members of the estrogen receptor family.	Chromatin/tr anscription
PHB	prohibitin	Prohibitin is an evolutionarily conserved gene that is ubiquitously expressed. It is thought to be a negative regulator of cell proliferation and may be a tumor suppressor. Mutations in PHB have been linked to sporadic breast cancer. Prohibitin is expressed as two transcripts with varying lengths of 3' untranslated region. The longer transcript is present at higher levels in proliferating tissues and cells, suggesting that this longer 3' untranslated region may function as a trans-acting regulatory RNA.	DNA replication/re pair/HR/coh esion
PIK3R 3	phosphoinositid e-3-kinase, regulatory subunit 3 (gamma)	Binds to activated (phosphorylated) protein-tyrosine kinases through its SH2 domain and regulates their kinase activity. During insulin stimulation, it also binds to IRS-1.	Metabolism/ mitochondria
PML	promyelocytic leukemia	The protein encoded by this gene is a member of the tripartite motif (TRIM) family. The TRIM motif includes three zinc-binding domains, a RING, a B-box type 1 and a B-box type 2, and a coiled-coil region. This phosphoprotein localizes to nuclear bodies where it functions as a transcription factor and tumor suppressor. Its expression is cell-cycle related and it regulates the p53 response to oncogenic signals. The gene is often involved in the translocation with the retinoic acid receptor alpha gene associated with acute promyelocytic leukemia (APL).	signaling/stre ss response
POLA1	polymerase (DNA directed),	It is the catalytic subunit of DNA polymerase, which together with a regulatory and two primase subunits, forms the DNA polymerase alpha	DNA replication/re

	alpha 1, catalytic subunit	complex. The catalytic subunit plays an essential role in the initiation of DNA replication.	pair/HR/cohesion
PPA1	pyrophosphatase (inorganic) 1	The protein encoded by this gene is a member of the inorganic pyrophosphatase (PPase) family. PPases catalyze the hydrolysis of pyrophosphate to inorganic phosphate, which is important for the phosphate metabolism of cells.	Metabolism/mitochondria
PPARG	peroxisome proliferator-activated receptor gamma	It is a member of the peroxisome proliferator-activated receptor (PPAR) subfamily of nuclear receptors. PPARs form heterodimers with retinoid X receptors (RXRs) and these heterodimers regulate transcription of various genes. Three subtypes of PPARs are known: PPAR-alpha, PPAR-delta, and PPAR-gamma. The protein encoded by this gene is PPAR-gamma and is a regulator of adipocyte differentiation. Additionally, PPAR-gamma has been implicated in the pathology of numerous diseases including obesity, diabetes, atherosclerosis and cancer.	Cell Differentiation
PPP1CA	protein phosphatase 1, catalytic subunit, alpha isozyme	The protein encoded by this gene is one of the three catalytic subunits of protein phosphatase 1 (PP1). PP1 is a serine/threonine specific protein phosphatase known to be involved in the regulation of a variety of cellular processes, such as cell division, glycogen metabolism, muscle contractility, protein synthesis, and HIV-1 viral transcription. Increased PP1 activity has been observed in the end stage of heart failure. Studies in both human and mice suggest that PP1 is an important regulator of cardiac function. Mouse studies also suggest that PP1 functions as a suppressor of learning and memory.	Metabolism/mitochondria
PRDM2	PR domain containing 2, with ZNF domain	This tumor suppressor gene is a member of a nuclear histone/protein methyltransferase superfamily. It encodes a zinc finger protein that can bind to retinoblastoma protein, estrogen receptor, and the TPA-responsive element (MTE) of the heme-oxygenase-1 gene. Although the functions of this protein have not been fully characterized, it may (1) play a role in transcriptional regulation during neuronal differentiation and pathogenesis of retinoblastoma, (2) act as a transcriptional activator of the heme-oxygenase-1 gene, and (3) be a specific effector of estrogen action.	Chromatin/transcription
PRKRA	protein kinase, interferon-inducible double stranded RNA dependent activator	It is a protein kinase activated by double-stranded RNA which mediates the effects of interferon in response to viral infection. Mutations in this gene have been associated with dystonia.	Immune Response
PSMA7	proteasome (prosome, macropain) subunit, alpha type, 7	The proteasome is a multicatalytic proteinase complex with a highly ordered ring-shaped 20S core structure. The core structure is composed of 4 rings of 28 non-identical subunits; 2 rings are composed of 7 alpha subunits and 2 rings are composed of 7 beta subunits. Proteasomes are distributed throughout eukaryotic cells at a high concentration and cleave peptides in an ATP/ubiquitin-dependent process in a non-lysosomal pathway. An essential function of a modified proteasome, the immunoproteasome, is the processing of class I MHC peptides. The core alpha subunit is also involved in regulating the hypoxia-inducible factor-1alpha, a transcription factor important for cellular responses to oxygen tension.	Protein degradation/proteosome
PSMD10	proteasome (prosome, macropain) 26S subunit, non-ATPase, 10	It is a subunit of the PA700/19S complex, which is the regulatory component of the 26S proteasome. The 26S proteosome complex is required for ubiquitin-dependent protein degradation. This protein is a non-ATPase subunit that may be involved in protein-protein interactions. Aberrant expression of this gene may play a role in tumorigenesis.	Protein degradation/proteosome
RAF1	v-raf-1 murine leukemia viral oncogene homolog 1	This gene is the cellular homolog of viral raf gene (v-raf). The encoded protein is a MAP kinase kinase kinase (MAP3K). Once activated, the cellular RAF1 protein can phosphorylate to activate the dual specificity protein kinases MEK1 and MEK2, which in turn phosphorylate to activate the serine/threonine specific protein kinases, ERK1 and ERK2. Activated ERKs are pleiotropic effectors of cell physiology and play an important role in the control of gene expression involved in the cell division cycle, apoptosis, cell differentiation and cell migration. Mutations in this gene are associated with Noonan syndrome 5 and LEOPARD syndrome 2.	signaling/stress response
RBBP4	retinoblastoma binding protein 4	It is a ubiquitously expressed nuclear protein which belongs to a highly conserved subfamily of WD-repeat proteins. It is present in protein complexes involved in histone acetylation and chromatin assembly. It is part of the Mi-2 complex which has been implicated in chromatin remodeling and transcriptional repression associated with histone deacetylation. This encoded protein is also part of co-repressor complexes, which is an integral component of transcriptional silencing. It is found among several cellular proteins that bind directly to retinoblastoma protein to regulate cell proliferation. This protein also seems to be involved in transcriptional repression of E2F-responsive genes.	Chromatin/transcription
RBBP5	retinoblastoma binding protein 5	It is a ubiquitously expressed nuclear protein which belongs to a highly conserved subfamily of WD-repeat proteins. The encoded protein binds directly to retinoblastoma protein, which regulates cell proliferation. It interacts preferentially with the underphosphorylated retinoblastoma protein via the E1A-binding pocket B.	Chromatin/transcription
RBBP7	retinoblastoma binding protein 7	This protein is a ubiquitously expressed nuclear protein and belongs to a highly conserved subfamily of WD-repeat proteins. It is found among several proteins that binds directly to retinoblastoma protein, which regulates cell proliferation. The encoded protein is found in many histone deacetylase complexes, including mSin3 co-repressor complex. It is also present in protein complexes involved in chromatin assembly. This protein can interact with BRCA1 tumor-suppressor gene and may have a role in the regulation of cell proliferation and differentiation.	Chromatin/transcription
RBBP8	retinoblastoma binding protein 8	The protein encoded by this gene is a ubiquitously expressed nuclear protein. It is found among several proteins that bind directly to pRb. This protein complexes with transcriptional co-repressor CTBP. It is also associated with BRCA1 and is thought to modulate the functions of BRCA1 in transcriptional	DNA replication/re

		regulation, DNA repair, and/or cell cycle checkpoint control. It is suggested that this gene may itself be a tumor suppressor acting in the same pathway as BRCA1. Three transcript variants encoding two different isoforms have been found for this gene. More transcript variants exist, but their full-length natures have not been determined.	pair/HR/cohesion
RBBP9	retinoblastoma binding protein 9	The protein encoded by this gene is a retinoblastoma binding protein that may play a role in the regulation of cell proliferation and differentiation.	Cell Differentiation
RBL1	retinoblastoma-like 1 (p107)	The protein encoded by this gene is similar in sequence and possibly function to the product of the retinoblastoma 1 (RB1) gene. The RB1 gene product is a tumor suppressor protein that appears to be involved in cell cycle regulation, as it is phosphorylated in the S to M phase transition and is dephosphorylated in the G1 phase of the cell cycle. Both the RB1 protein and the product of this gene can form a complex with adenovirus E1A protein and SV40 large T-antigen, with the SV40 large T-antigen binding only to the unphosphorylated form of each protein. In addition, both proteins can inhibit the transcription of cell cycle genes containing E2F binding sites in their promoters. Due to the sequence and biochemical similarities with the RB1 protein, it is thought that the protein encoded by this gene may also be a tumor suppressor.	G1/S and G2/M cell cycle progression/ meiosis
RBL2	retinoblastoma-like 2 (p130)	Directly involved in heterochromatin formation by maintaining overall chromatin structure and, in particular, that of constitutive heterochromatin by stabilizing histone methylation. Recruits and targets histone methyltransferases SUV420H1 and SUV420H2, leading to epigenetic transcriptional repression. Controls histone H4 'Lys-20' trimethylation. Probably acts as a transcription repressor by recruiting chromatin-modifying enzymes to promoters. Potent inhibitor of E2F-mediated trans-activation, associates preferentially with E2F5	G1/S and G2/M cell cycle progression/ meiosis
RFC1	replication factor C (activator 1) 1, 145kDa	It is the large subunit of replication factor C, a five subunit DNA polymerase accessory protein, which is a DNA-dependent ATPase required for eukaryotic DNA replication and repair. The large subunit acts as an activator of DNA polymerases, binds to the 3' end of primers, and promotes coordinated synthesis of both strands. It may also have a role in telomere stability.	DNA replication/repair/HR/cohesion
RING1	ring finger protein 1	This gene belongs to the RING finger family, members of which encode proteins characterized by a RING domain, a zinc-binding motif related to the zinc finger domain. The gene product can bind DNA and can act as a transcriptional repressor. It is associated with the multimeric polycomb group protein complex. The gene product interacts with the polycomb group proteins BMI1, EDR1, and CBX4, and colocalizes with these proteins in large nuclear domains. It interacts with the CBX4 protein via its glycine-rich C-terminal domain. The gene maps to the HLA class II region, where it is contiguous with the RING finger genes FABGL and HKE4.	Protein degradation/ proteosome
RNF40	ring finger protein 40, E3 ubiquitin protein ligase	The protein encoded by this gene contains a RING finger, a motif known to be involved in protein-protein and protein-DNA interactions. This protein was reported to interact with the tumor suppressor protein RB1. Studies of the rat counterpart suggested that this protein may function as an E3 ubiquitin-protein ligase, and facilitate the ubiquitination and degradation of syntaxin 1, which is an essential component of the neurotransmitter release machinery.	Protein degradation/ proteosome
SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2	Inhibits urokinase-type plasminogen activator. The monocyte derived PAI-2 is distinct from the endothelial cell-derived PAI-1.	Apoptosis
SIRT1	sirtuin 1	It is a member of the sirtuin family of proteins, homologs to the yeast Sir2 protein. Members of the sirtuin family are characterized by a sirtuin core domain and grouped into four classes. The functions of human sirtuins have not yet been determined; Studies suggest that the human sirtuins may function as intracellular regulatory proteins with mono-ADP-ribosyltransferase activity. The protein encoded by this gene is included in class I of the sirtuin family.	Unknown
SKI	v-ski sarcoma viral oncogene homolog (avian)	It is the nuclear protooncogene protein homolog of avian sarcoma viral (v-ski) oncogene. It functions as a repressor of TGF-beta signaling, and may play a role in neural tube development and muscle differentiation.	Cell Differentiation
SKIL	SKI-like oncogene	The protein encoded by this gene is a component of the SMAD pathway, which regulates cell growth and differentiation through transforming growth factor-beta (TGFB). In the absence of ligand, the encoded protein binds to the promoter region of TGFB-responsive genes and recruits a nuclear repressor complex. TGFB signaling causes SMAD3 to enter the nucleus and degrade this protein, allowing these genes to be activated.	Cell Differentiation
SKP2	S-phase kinase-associated protein 2, E3 ubiquitin protein ligase	It is a member of the F-box protein family which is characterized by an approximately 40 amino acid motif, the F-box. The F-box proteins constitute one of the four subunits of ubiquitin protein ligase complex called SCFs (SKP1-cullin-F-box), which function in phosphorylation-dependent ubiquitination. The F-box proteins are divided into 3 classes: Fbws containing WD-40 domains, Fbls containing leucine-rich repeats, and Fbxs containing either different protein-protein interaction modules or no recognizable motifs. The protein encoded by this gene belongs to the Fbls class; in addition to an F-box, this protein contains 10 tandem leucine-rich repeats. This protein is an essential element of the cyclin A-CDK2 S-phase kinase. It specifically recognizes phosphorylated cyclin-dependent kinase inhibitor 1B (CDKN1B, also referred to as p27 or KIP1) predominantly in S phase and interacts with S-phase kinase-associated protein 1 (SKP1 or p19). In addition, this gene is established as a protooncogene causally involved in the pathogenesis of lymphomas.	protein degradation/ proteosome
SMYD2	SET and MYND domain containing 2	SET domain-containing proteins, such as SMYD2, catalyze lysine methylation.	Chromatin/transcription
SNAPC1	small nuclear RNA activating	Part of the SNAPc complex required for the transcription of both RNA polymerase II and III small-nuclear RNA genes. Binds to the proximal	Chromatin/transcription

	complex, polypeptide 1, 43kDa	sequence element (PSE), a non-TATA-box basal promoter element common to these 2 types of genes. Recruits TBP and BRF2 to the U6 snRNA TATA box.	
SNAPC3	small nuclear RNA activating complex, polypeptide 3, 50kDa	Same as SNAPC1.	Chromatin/transcription
SNW1	SNW domain containing 1	This coactivator can bind to the ligand-binding domain of the vitamin D receptor and to retinoid receptors to enhance vitamin D-, retinoic acid-, estrogen-, and glucocorticoid-mediated gene expression. It can also function as a splicing factor by interacting with poly(A)-binding protein 2 to directly control the expression of muscle-specific genes at the transcriptional level. Finally, the protein may be involved in oncogenesis since it interacts with a region of SKI oncoproteins that is required for transforming activity.	RNA processing
SP1	Sp1 transcription factor	The protein is a zinc finger transcription factor that binds to GC-rich motifs of many promoters. The encoded protein is involved in many cellular processes, including cell differentiation, cell growth, apoptosis, immune responses, response to DNA damage, and chromatin remodeling.	Immune Response
STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)	The protein is a member of the STAT protein family. In response to cytokines and growth factors, STAT family members are phosphorylated by the receptor associated kinases, and then form homo- or heterodimers that translocate to the cell nucleus where they act as transcription activators. This protein is activated through phosphorylation in response to various cytokines and growth factors including IFNs, EGF, IL5, IL6, HGF, LIF and BMP2. This protein mediates the expression of a variety of genes in response to cell stimuli, and thus plays a key role in many cellular processes such as cell growth and apoptosis. The small GTPase Rac1 has been shown to bind and regulate the activity of this protein. PIAS3 protein is a specific inhibitor of this protein.	Immune Response
SUV39H1	suppressor of variegation 3-9 homolog 1 (Drosophila)	This gene is a member of the suppressor of variegation 3-9 homolog family and encodes a protein with a chromodomain and a C-terminal SET domain. This nuclear protein moves to the centromeres during mitosis and functions as a histone methyltransferase, methylating Lys-9 of histone H3. Overall, it plays a vital role in heterochromatin organization, chromosome segregation, and mitotic progression.	Chromatin/transcription
TAF1	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 250kDa	TFIID is composed of the TATA-binding protein (TBP) and a group of evolutionarily conserved proteins known as TBP-associated factors or TAFs. TAFs may participate in basal transcription, serve as coactivators, function in promoter recognition or modify general transcription factors (GTFs) to facilitate complex assembly and transcription initiation. It is the largest subunit of TFIID. This subunit binds to core promoter sequences encompassing the transcription start site. It also binds to activators and other transcriptional regulators, and these interactions affect the rate of transcription initiation. This subunit contains two independent protein kinase domains at the N and C-terminals, but also possesses acetyltransferase activity and can act as a ubiquitin-activating/conjugating enzyme. This gene is part of a complex transcriptional unit (TAF1/DYT3), wherein some products share exons with TAF1 as well as additional exons downstream.	G1/S and G2/M cell cycle progression/ meiosis
TAL1	T-cell acute lymphocytic leukemia 1	Implicated in the genesis of hemopoietic malignancies. It may play an important role in hemopoietic differentiation. Serves as a positive regulator of erythroid differentiation	Cell Differentiation
TBP	TATA box binding protein	It is TBP, the TATA-binding protein. A distinctive feature of TBP is a long string of glutamines in the N-terminus. This region of the protein modulates the DNA binding activity of the C terminus, and modulation of DNA binding affects the rate of transcription complex formation and initiation of transcription. The number of CAG repeats encoding the polyglutamine tract is usually 32-39, and expansion of the number of repeats increases the length of the polyglutamine string and is associated with spinocerebellar ataxia 17, a neurodegenerative disorder classified as a polyglutamine disease.	Chromatin/transcription
TCF3	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)	It is a member of the E protein (class I) family of helix-loop-helix transcription factors. E proteins activate transcription by binding to regulatory E-box sequences on target genes as heterodimers or homodimers, and are inhibited by heterodimerization with inhibitor of DNA-binding (class IV) helix-loop-helix proteins. E proteins play a critical role in lymphopoiesis, and the encoded protein is required for B and T lymphocyte development. Deletion of this gene or diminished activity of the encoded protein may play a role in lymphoid malignancies. This gene is also involved in several chromosomal translocations that are associated with lymphoid malignancies including pre-B-cell acute lymphoblastic leukemia (t(1;19), with PBX1), childhood leukemia (t(19;19), with TFPT) and acute leukemia (t(12;19), with ZNF384).	Chromatin/transcription
TFDP1	transcription factor Dp-1	It is a member of a family of transcription factors that heterodimerize with E2F proteins to enhance their DNA-binding activity and promote transcription from E2F target genes. The encoded protein functions as part of this complex to control the transcriptional activity of numerous genes involved in cell cycle progression from G1 to S phase.	G1/S and G2/M cell cycle progression/ meiosis
TFDP2	transcription factor Dp-2 (E2F dimerization partner 2)	The gene is a member of the transcription factor DP family. The encoded protein forms heterodimers with the E2F transcription factors resulting in transcriptional activation of cell cycle regulated genes.	G1/S and G2/M cell cycle progression/ meiosis
TGM2	transglutaminase 2 (C polypeptide, protein-glutamine-gamma-	Transglutaminases are enzymes that catalyze the crosslinking of proteins by epsilon-gamma glutamyl lysine isopeptide bonds. While the primary structure of transglutaminases is not conserved, they all have the same amino acid sequence at their active sites and their activity is calcium-dependent. The protein acts as a monomer, is induced by retinoic acid, and appears to be involved in apoptosis.	Apoptosis

	glutamyltransferase)		
THOC1	THO complex 1	Component of the THO subcomplex of the TREX complex. The TREX complex specifically associates with spliced mRNA and not with unspliced pre-mRNA. It is recruited to spliced mRNAs by a transcription-independent mechanism. Binds to mRNA upstream of the exon-junction complex (EJC) and is recruited in a splicing- and cap-dependent manner to a region near the 5' end of the mRNA where it functions in mRNA export.	RNA processing
TMPO	thymopoietin	The protein encoded by this gene resides in the nucleus and may play a role in the assembly of the nuclear lamina, and thus help maintain the structural organization of the nuclear envelope. It may function as a receptor for the attachment of lamin filaments to the inner nuclear membrane. Mutations in this gene are associated with dilated cardiomyopathy.	DNA replication/repair/HR/cohesion
TP53	tumor protein p53	It is tumor protein p53, which responds to diverse cellular stresses to regulate target genes that induce cell cycle arrest, apoptosis, senescence, DNA repair, or changes in metabolism. p53 is a DNA-binding protein containing transcription activation, DNA-binding, and oligomerization domains. It is postulated to bind to a p53-binding site and activate expression of downstream genes that inhibit growth and/or invasion, and thus function as a tumor suppressor. Mutants of p53 that frequently occur in a number of different human cancers fail to bind the consensus DNA binding site, and hence cause the loss of tumor suppressor activity. Alterations of this gene occur not only as somatic mutations in human malignancies, but also as germline mutations in some cancer-prone families with Li-Fraumeni syndrome.	Apoptosis
TP73	tumor protein p73	It is a member of the p53 family of transcription factors involved in cellular responses to stress and development. It maps to a region on chromosome 1p36 that is frequently deleted in neuroblastoma and other tumors, and thought to contain multiple tumor suppressor genes. The demonstration that this gene is monoallelically expressed (likely from the maternal allele), supports the notion that it is a candidate gene for neuroblastoma.	Apoptosis
TRAP1	TNF receptor-associated protein 1	HSP90 proteins are highly conserved molecular chaperones that have key roles in signal transduction, protein folding, protein degradation, and morphologic evolution. HSP90 proteins normally associate with other cochaperones and play important roles in folding newly synthesized proteins or stabilizing and refolding denatured proteins after stress. TRAP1 is a mitochondrial HSP90 protein.	protein folding/protein glycosylation/cell wall biogenesis&integrity
TRIM27	tripartite motif containing 27	It is a member of the tripartite motif (TRIM) family. The TRIM motif includes three zinc-binding domains, a RING, a B-box type 1 and a B-box type 2, and a coiled-coil region. This protein localizes to the nuclear matrix. It interacts with the enhancer of polycomb protein and represses gene transcription. It is also thought to be involved in the differentiation of male germ cells. Fusion of the N-terminus of this protein with the truncated C-terminus of the RET gene product has been shown to result in production of the ret transforming protein.	Apoptosis
TRIP11	thyroid hormone receptor interactor 11	This gene was identified based on the interaction of its protein product with thyroid hormone receptor beta. This protein is associated with the Golgi apparatus. The N-terminal region of the protein binds Golgi membranes and the C-terminal region binds the minus ends of microtubules; thus, the protein is thought to play a role in assembly and maintenance of the Golgi ribbon structure around the centrosome. Mutations in this gene cause achondrogenesis type IA.	Golgi/endosome/vacuole/sorting
UBC	ubiquitin C	This gene represents a ubiquitin gene, ubiquitin C. The encoded protein is a polyubiquitin precursor. Conjugation of ubiquitin monomers or polymers can lead to various effects within a cell, depending on the residues to which ubiquitin is conjugated. Ubiquitination has been associated with protein degradation, DNA repair, cell cycle regulation, kinase modification, endocytosis, and regulation of other cell signaling pathways.	Protein degradation/proteosome
UBR4	ubiquitin protein ligase E3 component n-recognin 4	The protein encoded by this gene is an E3 ubiquitin-protein ligase that interacts with the retinoblastoma-associated protein in the nucleus and with calcium-bound calmodulin in the cytoplasm. The encoded protein appears to be a cytoskeletal component in the cytoplasm and part of the chromatin scaffold in the nucleus. In addition, this protein is a target of the human papillomavirus type 16 E7 oncoprotein.	Protein degradation/proteosome
UBTF	upstream binding transcription factor, RNA polymerase I	It is a member of the HMG-box DNA-binding protein family. The encoded protein plays a critical role in ribosomal RNA transcription as a key component of the pre-initiation complex, mediating the recruitment of RNA polymerase I to rDNA promoter regions. The encoded protein may also play important roles in chromatin remodeling and pre-rRNA processing, and its activity is regulated by both phosphorylation and acetylation.	Chromatin/transcription
UHRF1	ubiquitin-like with PHD and ring finger domains 1	It is a member of a subfamily of RING-finger type E3 ubiquitin ligases. The protein binds to specific DNA sequences, and recruits a histone deacetylase to regulate gene expression. Its expression peaks at late G1 phase and continues during G2 and M phases of the cell cycle. It plays a major role in the G1/S transition by regulating topoisomerase IIalpha and retinoblastoma gene expression, and functions in the p53-dependent DNA damage checkpoint. Multiple transcript variants encoding different isoforms have been found for this gene.	Protein degradation/proteosome
UHRF2	ubiquitin-like with PHD and ring finger domains 2, E3 ubiquitin protein ligase	It is a nuclear protein which is involved in cell-cycle regulation. The encoded protein is a ubiquitin-ligase capable of ubiquinating PCNP (PEST-containing nuclear protein), and together they may play a role in tumorigenesis. The encoded protein contains an NIRF_N domain, a PHD finger, a set- and ring-associated (SRA) domain, and a RING finger domain and several of these domains have been shown to be essential for the regulation of cell proliferation. This protein may also have a role in intranuclear degradation of polyglutamine aggregates.	Protein degradation/proteosome
USP4	ubiquitin specific peptidase 4	The protein encoded by this gene is a protease that deubiquitinates target proteins such as ADORA2A and TRIM21. The encoded protein shuttles between the nucleus and cytoplasm and is involved in maintaining operational fidelity in the endoplasmic reticulum.	Protein degradation/proteosome

	(proto-oncogene)		
VDR	vitamin D (1,25-dihydroxyvitamin D3) receptor	It is the nuclear hormone receptor for vitamin D3. This receptor also functions as a receptor for the secondary bile acid lithocholic acid. The receptor belongs to the family of trans-acting transcriptional regulatory factors and shows sequence similarity to the steroid and thyroid hormone receptors. Downstream targets of this nuclear hormone receptor are principally involved in mineral metabolism though the receptor regulates a variety of other metabolic pathways, such as those involved in the immune response and cancer. Mutations in this gene are associated with type II vitamin D-resistant rickets. A single nucleotide polymorphism in the initiation codon results in an alternate translation start site three codons downstream.	Chromatin/transcription
YY1	YY1 transcription factor	YY1 is a ubiquitously distributed transcription factor belonging to the GLI-Kruppel class of zinc finger proteins. The protein is involved in repressing and activating a diverse number of promoters. YY1 may direct histone deacetylases and histone acetyltransferases to a promoter in order to activate or repress the promoter, thus implicating histone modification in the function of YY1.	Cell Differentiation
ZBTB7A	zinc finger and BTB domain containing 7A	Unknown	Unknown
AP1AR	adaptor-related protein complex 1 associated regulatory protein	Necessary for AP-1 dependent transport between the trans-Golgi network and endosomes. Regulates the membrane association of AP1G1/Gamma1-adaptin, one of the subunits of the AP-1 adapter complex. The direct interaction with AP1G1/Gamma1-adaptin attenuates the release of the AP-1 complex from membranes.	Golgi/endosome/vacuole/sorting
BDP1	B double prime 1, subunit of RNA polymerase III transcription initiation factor IIIB	The product of this gene is a subunit of the TFIIIB transcription initiation complex, which recruits RNA polymerase III to target promoters in order to initiate transcription. The encoded protein localizes to concentrated aggregates in the nucleus, and is required for transcription from all three types of polymerase III promoters. It is phosphorylated by casein kinase II during mitosis, resulting in its release from chromatin and suppression of polymerase III transcription.	Chromatin/transcription
BNC2	basonuclin 2	Probable transcription factor specific for skin keratinocytes. May play a role in the differentiation of spermatozoa and oocytes.	Cell Differentiation
BRAF	v-raf murine sarcoma viral oncogene homolog B1	It is a protein belonging to the raf/mil family of serine/threonine protein kinases. This protein plays a role in regulating the MAP kinase/ERKs signaling pathway, which affects cell division, differentiation, and secretion. Mutations in this gene are associated with cardiofaciocutaneous syndrome, a disease characterized by heart defects, mental retardation and a distinctive facial appearance. Mutations in this gene have also been associated with various cancers, including non-Hodgkin lymphoma, colorectal cancer, malignant melanoma, thyroid carcinoma, non-small cell lung carcinoma, and adenocarcinoma of lung.	signaling/stress response
CASP10	caspase 10, apoptosis-related cysteine peptidase	It is a member of the cysteine-aspartic acid protease (caspase) family. Caspases exist as inactive proenzymes which undergo proteolytic processing at conserved aspartic residues to produce two subunits, large and small, that dimerize to form theactive enzyme. This protein cleaves and activates caspases 3 and 7, and the protein itself is processed by caspase 8. Mutations in this gene are associated with type IIA autoimmune lymphoproliferative syndrome, non-Hodgkin lymphoma and gastric cancer.	Apoptosis
CASP2	caspase 2, apoptosis-related cysteine peptidase	It is a member of the cysteine-aspartic acid protease (caspase) family. The encoded protein may function in stress-induced cell death pathways, cell cycle maintenance, and the suppression of tumorigenesis. Increased expression of this gene may play a role in neurodegenerative disorders including Alzheimer's disease, Huntington's disease and temporal lobe epilepsy.	Apoptosis
CASP3	caspase 3, apoptosis-related cysteine peptidase	It is a protein which is a member of the caspase family. Caspases exist as inactive proenzymes which undergo proteolytic processing at conserved aspartic residues to produce two subunits, large and small, that dimerize to form the active enzyme. This protein cleaves and activates caspases 6, 7 and 9, and the protein itself is processed by caspases 8, 9 and 10. It is the predominant caspase involved in the cleavage of amyloid-beta 4A precursor protein, which is associated with neuronal death in Alzheimer's disease.	Apoptosis
CASP6	caspase 6, apoptosis-related cysteine peptidase	It is a protein which is a member of the caspase family. Caspases exist as inactive proenzymes which undergo proteolytic processing at conserved aspartic residues to produce two subunits, large and small, that dimerize to form the active enzyme. This protein is processed by caspases 7, 8 and 10, and is thought to function as a downstream enzyme in the caspase activation cascade.	Apoptosis
CASP7	caspase 7, apoptosis-related cysteine peptidase	It is a protein which is a member of the caspase family. Sequential activation of caspases plays a central role in the execution-phase of cell apoptosis. Caspases exist as inactive proenzymes which undergo proteolytic processing at conserved aspartic residues to produce two subunits, large and small, that dimerize to form the active enzyme. The precursor of this caspase is cleaved by caspase 3 and 10. It is activated upon cell death stimuli and induces apoptosis.	Apoptosis
CASP8	caspase 8, apoptosis-related cysteine peptidase	It is a member of the caspase family. Activation of caspases requires proteolytic processing at conserved internal aspartic residues to generate a heterodimeric enzyme consisting of the large and small subunits. This protein is involved in the programmed cell death induced by Fas and various apoptotic stimuli. The N-terminal FADD-like death effector domain of this protein suggests that it may interact with Fas-interacting protein FADD. This protein was detected in the insoluble fraction of the affected brain region from Huntington disease patients but not in those from normal controls, which implicated the role in neurodegenerative diseases.	Apoptosis
CASP9	caspase 9, apoptosis-	It is a member of caspase family. Sequential activation of caspases plays a central role in the execution-phase of cell apoptosis. Caspases exist as inactive proenzymes which undergo proteolytic processing at conserved aspartic residues to produce two subunits, large and small, that dimerize to form the	Apoptosis

	related cysteine peptidase	active enzyme. This protein is processed by caspase APAF1; this step is thought to be one of the earliest in the caspase activation cascade.	
CDK14	cyclin-dependent kinase 14	PFTK1 is a member of the CDC2-related protein kinase family.	G1/S and G2/M cell cycle progression/ meiosis
CDKN 1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	It is a potent cyclin-dependent kinase inhibitor. The encoded protein binds to and inhibits the activity of cyclin-CDK2 or -CDK4 complexes, and thus functions as a regulator of cell cycle progression at G1. The expression of this gene is tightly controlled by the tumor suppressor protein p53, through which this protein mediates the p53-dependent cell cycle G1 phase arrest in response to a variety of stress stimuli. This protein can interact with proliferating cell nuclear antigen (PCNA), a DNA polymerase accessory factor, and plays a regulatory role in S phase DNA replication and DNA damage repair. This protein was reported to be specifically cleaved by CASP3-like caspases, which thus leads to a dramatic activation of CDK2, and may be instrumental in the execution of apoptosis following caspase activation.	G1/S and G2/M cell cycle progression/ meiosis
CEBP D	CCAAT/enhancer binding protein (C/EBP), delta	The protein encoded by this intronless gene is a bZIP transcription factor which can bind as a homodimer to certain DNA regulatory regions. It can also form heterodimers with the related protein CEBP-alpha. The encoded protein is important in the regulation of genes involved in immune and inflammatory responses, and may be involved in the regulation of genes associated with activation and/or differentiation of macrophages.	Immune Response
CEBPE	CCAAT/enhancer binding protein (C/EBP), epsilon	The protein encoded by this gene is a bZIP transcription factor which can bind as a homodimer to certain DNA regulatory regions. It can also form heterodimers with the related protein CEBP-delta. The encoded protein may be essential for terminal differentiation and functional maturation of committed granulocyte progenitor cells. Mutations in this gene have been associated with Specific Granule Deficiency, a rare congenital disorder.	Immune Response
CHEK 1	checkpoint kinase 1	The protein encoded by this gene belongs to the Ser/Thr protein kinase family. It is required for checkpoint mediated cell cycle arrest in response to DNA damage or the presence of unreplicated DNA. This protein acts to integrate signals from ATM and ATR, two cell cycle proteins involved in DNA damage responses, that also associate with chromatin in meiotic prophase I. Phosphorylation of CDC25A protein phosphatase by this protein is required for cells to delay cell cycle progression in response to double-strand DNA breaks.	DNA replication/repair/HR/cohesion
CHEK 2	checkpoint kinase 2	In response to DNA damage and replication blocks, cell cycle progression is halted through the control of critical cell cycle regulators. The protein encoded by this gene is a cell cycle checkpoint regulator and putative tumor suppressor. It contains a forkhead-associated protein interaction domain essential for activation in response to DNA damage and is rapidly phosphorylated in response to replication blocks and DNA damage. When activated, the encoded protein is known to inhibit CDC25C phosphatase, preventing entry into mitosis, and has been shown to stabilize the tumor suppressor protein p53, leading to cell cycle arrest in G1. In addition, this protein interacts with and phosphorylates BRCA1, allowing BRCA1 to restore survival after DNA damage. Mutations in this gene have been linked with Li-Fraumeni syndrome, a highly penetrant familial cancer phenotype usually associated with inherited mutations in TP53. Also, mutations in this gene are thought to confer a predisposition to sarcomas, breast cancer, and brain tumors. This nuclear protein is a member of the CDS1 subfamily of serine/threonine protein kinases.	DNA replication/repair/HR/cohesion
CTBP1	C-terminal binding protein 1	It is a protein that binds to the C-terminus of adenovirus E1A proteins. This phosphoprotein is a transcriptional repressor and may play a role during cellular proliferation. This protein and the product of a second closely related gene, CTBP2, can dimerize. Both proteins can also interact with a polycomb group protein complex which participates in regulation of gene expression during development.	Cell Differentiation
DGKZ	diacylglycerol kinase, zeta	The protein belongs to the eukaryotic diacylglycerol kinase family. It may attenuate protein kinase C activity by regulating diacylglycerol levels in intracellular signaling cascade and signal transduction.	Metabolism/ mitochondria
E2F3	E2F transcription factor 3	Same as E2F1.	G1/S and G2/M cell cycle progression/ meiosis
ELF1	E74-like factor 1 (ets domain transcription factor)	It is an E26 transformation-specific related transcription factor. The protein is primarily expressed in lymphoid cells and acts as both an enhancer and a repressor to regulate transcription of various genes.	Chromatin/transcription
FOS	FBJ murine osteosarcoma viral oncogene homolog	The Fos gene family consists of 4 members: FOS, FOSB, FOSL1, and FOSL2. These genes encode leucine zipper proteins that can dimerize with proteins of the JUN family, thereby forming the transcription factor complex AP-1. As such, the FOS proteins have been implicated as regulators of cell proliferation, differentiation, and transformation. In some cases, expression of the FOS gene has also been associated with apoptotic cell death.	Cell Differentiation
GATA 1	GATA binding protein 1 (globin transcription factor 1)	It is a protein which belongs to the GATA family of transcription factors. The protein plays an important role in erythroid development by regulating the switch of fetal hemoglobin to adult hemoglobin. Mutations in this gene have been associated with X-linked dyserythropoietic anemia and thrombocytopenia.	Cell Differentiation
GSR	glutathione reductase	It is a member of the class-I pyridine nucleotide-disulfide oxidoreductase family. This enzyme is a homodimeric flavoprotein. It is a central enzyme of cellular antioxidant defense, and reduces oxidized glutathione disulfide (GSSG) to the sulfhydryl form GSH, which is an important cellular antioxidant. Rare mutations in this gene result in hereditary glutathione reductase deficiency.	Metabolism/ mitochondria

GTF3C 2	general transcription factor IIIC, polypeptide 2, beta 110kDa	Same as GTF3C1	Chromatin/tr anscription
HIF1A	hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)	It is the alpha subunit of transcription factor hypoxia-inducible factor-1 (HIF-1), which is a heterodimer composed of an alpha and a beta subunit. HIF-1 functions as a master regulator of cellular and systemic homeostatic response to hypoxia by activating transcription of many genes, including those involved in energy metabolism, angiogenesis, apoptosis, and other genes whose protein products increase oxygen delivery or facilitate metabolic adaptation to hypoxia. HIF-1 thus plays an essential role in embryonic vascularization, tumor angiogenesis and pathophysiology of ischemic disease.	Chromatin/tr anscription
HSPA8	heat shock 70kDa protein 8	It is a member of the heat shock protein 70 family, which contains both heat-inducible and constitutively expressed members. This protein belongs to the latter group, which are also referred to as heat-shock cognate proteins. It functions as a chaperone, and binds to nascent polypeptides to facilitate correct folding. It also functions as an ATPase in the disassembly of clathrin-coated vesicles during transport of membrane components through the cell.	Chromatin/tr anscription
ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	The protein encoded by this gene belongs to the inhibitor of DNA binding family, members of which are transcriptional regulators that contain a helix-loop-helix (HLH) domain but not a basic domain. Members of the inhibitor of DNA binding family inhibit the functions of basic helix-loop-helix transcription factors in a dominant-negative manner by suppressing their heterodimerization partners through the HLH domains.	Chromatin/tr anscription
INS	insulin	After removal of the precursor signal peptide, proinsulin is post-translationally cleaved into three peptides: the B chain and A chain peptides, which are covalently linked via two disulfide bonds to form insulin, and C-peptide. Binding of insulin to the insulin receptor (INSR) stimulates glucose uptake.	Metabolism/ mitochondria
LIN37	lin-37 homolog	It is a protein expressed in the eye.	G1/S and G2/M cell cycle progression/ meiosis
MAPK 9	mitogen-activated protein kinase 9	The protein encoded by this gene is a member of the MAP kinase family. MAP kinases act as an integration point for multiple biochemical signals, and are involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development. This kinase targets specific transcription factors, and thus mediates immediate-early gene expression in response to various cell stimuli. It is most closely related to MAPK8, both of which are involved in UV radiation induced apoptosis, thought to be related to the cytochrome c-mediated cell death pathway. This gene and MAPK8 are also known as c-Jun N-terminal kinases. This kinase blocks the ubiquitination of tumor suppressor p53, and thus it increases the stability of p53 in nonstressed cells. Studies of this gene's mouse counterpart suggest a key role in T-cell differentiation.	Cell Differentiatio n
MDM4	Mdm4 p53 binding protein homolog (mouse)	It is a nuclear protein that contains a p53 binding domain at the N-terminus and a RING finger domain at the C-terminus, and shows structural similarity to p53-binding protein MDM2. Both proteins bind the p53 tumor suppressor protein and inhibit its activity, and have been shown to be overexpressed in a variety of human cancers. However, unlike MDM2 which degrades p53, this protein inhibits p53 by binding its transcriptional activation domain. This protein also interacts with MDM2 protein via the RING finger domain, and inhibits the latter's degradation. So this protein can reverse MDM2-targeted degradation of p53, while maintaining suppression of p53 transactivation and apoptotic functions.	protein degradation/ proteosome
MRPS1 8B	mitochondrial ribosomal protein S18B	Mitochondrial ribosomes (mitoribosomes) consist of a small 28S subunit and a large 39S subunit. Among different species, the proteins comprising the mitoribosome differ greatly in sequence, and sometimes in biochemical properties, which prevents easy recognition by sequence homology. It is a 28S subunit protein that belongs to the ribosomal protein S18P family.	Metabolism/ mitochondria
NCL	nucleolin	Nucleolin (NCL), a eukaryotic nucleolar phosphoprotein, is involved in the synthesis and maturation of ribosomes. It is located mainly in dense fibrillar regions of the nucleolus. Human NCL gene consists of 14 exons with 13 introns and spans approximately 11kb. The intron 11 of the NCL gene encodes a small nucleolar RNA, termed U20.	RNA processing
NDC8 0	NDC80 kinetochore complex component homolog (S. cerevisiae)	It is a component of the NDC80 kinetochore complex. The encoded protein consists of an N-terminal microtubule binding domain and a C-terminal coiled-coiled domain that interacts with other components of the complex. This protein functions to organize and stabilize microtubule-kinetochore interactions and is required for proper chromosome segregation.	Chromosome segregation/k inetochore/sp indle/microtu bule
NEFM	neurofilament, medium polypeptide	Neurofilaments are type IV intermediate filament heteropolymers composed of light, medium, and heavy chains. Neurofilaments comprise the axoskeleton and functionally maintain neuronal caliber. They may also play a role in intracellular transport to axons and dendrites. It is the medium neurofilament protein. This protein is commonly used as a biomarker of neuronal damage.	Chromosome segregation/k inetochore/sp indle/microtu bule
PABP N1	poly(A) binding protein, nuclear 1	It is an abundant nuclear protein that binds with high affinity to nascent poly(A) tails. The protein is required for progressive and efficient polymerization of poly(A) tails at the 3' ends of eukaryotic transcripts and controls the size of the poly(A) tail to about 250 nt. At steady-state, this protein is localized in the nucleus whereas a different poly(A) binding protein is localized in the cytoplasm. This gene contains a GCG trinucleotide repeat at the 5' end of the coding region, and expansion of this repeat from the normal 6 copies to 8-13 copies leads to autosomal dominant oculopharyngeal muscular dystrophy (OPMD) disease.	RNA processing

PAX2	paired box 2	The central feature of this transcription factor is the conserved DNA-binding paired box domain. PAX2 is believed to be a target of transcriptional supression by the tumor supressor gene WT1. Mutations within PAX2 have been shown to result in optic nerve colobomas and renal hypoplasia.	Cell Differentiation
PAX5	paired box 5	It is a member of the paired box (PAX) family of transcription factors. The central feature of this gene family is a novel, highly conserved DNA-binding motif, known as the paired box. PAX proteins are important regulators in early development, and alterations in the expression of their genes are thought to contribute to neoplastic transformation. It is the B-cell lineage specific activator protein that is expressed at early, but not late stages of B-cell differentiation. Its expression has also been detected in developing CNS and testis and so the encoded protein may also play a role in neural development and spermatogenesis. This gene is located at 9p13, which is involved in t(9;14)(p13;q32) translocations recurring in small lymphocytic lymphomas of the plasmacytoid subtype, and in derived large-cell lymphomas. This translocation brings the potent E-mu enhancer of the IgH gene into close proximity of the PAX5 promoter, suggesting that the deregulation of transcription of this gene contributes to the pathogenesis of these lymphomas	Cell Differentiation
PCNA	proliferating cell nuclear antigen	The protein is found in the nucleus and is a cofactor of DNA polymerase delta. The protein acts as a homotrimer and helps increase the processivity of leading strand synthesis during DNA replication. In response to DNA damage, this protein is ubiquitinated and is involved in the RAD6-dependent DNA repair pathway.	DNA replication/repair/HR/cohesion
PIK3R1	phosphoinositide-3-kinase, regulatory subunit 1 (alpha)	Phosphatidylinositol 3-kinase phosphorylates the inositol ring of phosphatidylinositol at the 3-prime position. The enzyme comprises a 110 kD catalytic subunit and a regulatory subunit of either 85, 55, or 50 kD. It is the 85 kD regulatory subunit. Phosphatidylinositol 3-kinase plays an important role in the metabolic actions of insulin, and a mutation in this gene has been associated with insulin resistance.	Metabolism/ mitochondria
PON2	paraoxonase 2	It is a member of the paraoxonase gene family, which includes three known members located adjacent to each other on the long arm of chromosome 7. The encoded protein is ubiquitously expressed in human tissues, membrane-bound, and may act as a cellular antioxidant, protecting cells from oxidative stress. Hydrolytic activity against acylhomoserine lactones, important bacterial quorum-sensing mediators, suggests the encoded protein may also play a role in defense responses to pathogenic bacteria. Mutations in this gene may be associated with vascular disease and a number of quantitative phenotypes related to diabetes.	Metabolism/ mitochondria
PPIA	peptidylprolyl isomerase A (cyclophilin A)	It is a member of the peptidyl-prolyl cis-trans isomerase (PPIase) family. PPIases catalyze the cis-trans isomerization of proline imidic peptide bonds in oligopeptides and accelerate the folding of proteins. The encoded protein is a cyclosporin binding-protein and may play a role in cyclosporin A-mediated immunosuppression. The protein can also interact with several HIV proteins, including p55 gag, Vpr, and capsid protein, and has been shown to be necessary for the formation of infectious HIV virions.	Metabolism/ mitochondria
PPP1CB	protein phosphatase 1, catalytic subunit, beta isozyme	The protein is one of the three catalytic subunits of protein phosphatase 1 (PP1). PP1 is a serine/threonine specific protein phosphatase known to be involved in the regulation of a variety of cellular processes, such as cell division, glycogen metabolism, muscle contractility, protein synthesis, and HIV-1 viral transcription. Mouse studies suggest that PP1 functions as a suppressor of learning and memory.	Metabolism/ mitochondria
PPP1CC	protein phosphatase 1, catalytic subunit, gamma isozyme	The protein belongs to the protein phosphatase family, PP1 subfamily. PP1 is an ubiquitous serine/threonine phosphatase that regulates many cellular processes, including cell division. It is expressed in mammalian cells as three closely related isoforms, alpha, beta/delta and gamma, which have distinct localization patterns. It is the gamma isozyme.	Metabolism/ mitochondria
PRKCB	protein kinase C, beta	Protein kinase C (PKC) is a family of serine- and threonine-specific protein kinases that can be activated by calcium and second messenger diacylglycerol. PKC family members phosphorylate a wide variety of protein targets and are known to be involved in diverse cellular signaling pathways. PKC family members also serve as major receptors for phorbol esters, a class of tumor promoters. Each member of the PKC family has a specific expression profile and is believed to play a distinct role in cells. The protein encoded by this gene is one of the PKC family members. This protein kinase has been reported to be involved in many different cellular functions, such as B cell activation, apoptosis induction, endothelial cell proliferation, and intestinal sugar absorption. Studies in mice also suggest that this kinase may also regulate neuronal functions and correlate fear-induced conflict behavior after stress.	Metabolism/ mitochondria
PRMT2	protein arginine methyltransferase 2		Chromatin/transcription
PSMC4	proteasome (prosome, macropain) 26S subunit, ATPase, 4	The 26S proteasome is a multicatalytic proteinase complex with a highly ordered structure composed of 2 complexes, a 20S core and a 19S regulator. The 20S core is composed of 4 rings of 28 non-identical subunits; 2 rings are composed of 7 alpha subunits and 2 rings are composed of 7 beta subunits. The 19S regulator is composed of a base, which contains 6 ATPase subunits and 2 non-ATPase subunits, and a lid, which contains up to 10 non-ATPase subunits. Proteasomes are distributed throughout eukaryotic cells at a high concentration and cleave peptides in an ATP/ubiquitin-dependent process in a non-lysosomal pathway. An essential function of a modified proteasome, the immunoproteasome, is the processing of class I MHC peptides. It is one of the ATPase subunits, a member of the triple-A family of ATPases which have a chaperone-like activity. This subunit has been shown to interact with an orphan member of the nuclear hormone receptor superfamily highly expressed in liver, and with gankyrin, a liver oncoprotein.	protein degradation/ proteosome
PURA	purine-rich element binding protein A	This gene product is a sequence-specific, single-stranded DNA-binding protein. It binds preferentially to the single strand of the purine-rich element termed PUR, which is present at origins of replication and in gene flanking regions in a variety of eukaryotes from yeasts through humans. Thus, it is implicated in the control of both DNA replication and transcription. Deletion	Chromatin/transcription

		of this gene has been associated with myelodysplastic syndrome and acute myelogenous leukemia.	
RBAK	RB-associated KRAB zinc finger	It is a nuclear protein which interacts with the tumor suppressor retinoblastoma 1. The two interacting proteins are thought to act as a transcriptional repressor for promoters which are activated by the E2F1 transcription factor. This protein contains a Kruppel-associated box (KRAB), which is a transcriptional repressor motif.	Chromatin/transcription
RUNX2	runt-related transcription factor 2	The protein is a member of the RUNX family of transcription factors and has Runt DNA-binding domain. This protein is essential for osteoblastic differentiation and skeletal morphogenesis and acts as a scaffold for nucleic acids and regulatory factors involved in skeletal gene expression. The protein can bind DNA both as a monomer or, with more affinity, as a subunit of a heterodimeric complex. Mutations in this gene have been associated with the bone development disorder cleidocranial dysplasia (CCD).	Cell Differentiation
SEPT4	septin 4	The protein is a member of the septin family of nucleotide binding proteins, originally described in yeast as cell division cycle regulatory proteins. Septins are highly conserved in yeast, Drosophila, and mouse, and appear to regulate cytoskeletal organization. Disruption of septin function disturbs cytokinesis and results in large multinucleate or polyploid cells. This gene is highly expressed in brain and heart.	Chromosome segregation/kinetochore/spindle/microtubule
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	The protein encoded by this gene is a member of the SWI/SNF family of proteins. Members of this family have helicase and ATPase activities and are thought to regulate transcription of certain genes by altering the chromatin structure around those genes. The encoded protein is part of the large ATP-dependent chromatin remodeling complex SNF/SWI, which is required for transcriptional activation of genes normally repressed by chromatin. In addition, this protein can bind BRCA1, as well as regulate the expression of the tumorigenic protein CD44.	Chromatin/transcription
SMARCB1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1	The protein is part of a complex that relieves repressive chromatin structures, allowing the transcriptional machinery to access its targets more effectively. The encoded nuclear protein may also bind to and enhance the DNA joining activity of HIV-1 integrase. This gene has been found to be a tumor suppressor, and mutations in it have been associated with malignant rhabdoid tumors.	Chromatin/transcription
SP3	Sp3 transcription factor	This gene belongs to a family of Sp1 related genes that encode transcription factors that regulate transcription by binding to consensus GC- and GT-box regulatory elements in target genes. This protein contains a zinc finger DNA-binding domain and several transactivation domains, and has been reported to function as a bifunctional transcription factor that either stimulates or represses the transcription of numerous genes.	Chromatin/transcription
SPI1	spleen focus forming virus (SFFV) proviral integration oncogene spi1	It is an ETS-domain transcription factor that activates gene expression during myeloid and B-lymphoid cell development. The nuclear protein binds to a purine-rich sequence known as the PU-box found near the promoters of target genes, and regulates their expression in coordination with other transcription factors and cofactors. The protein can also regulate alternative splicing of target genes.	Chromatin/transcription
SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	The protein encoded by this gene is a transcriptional activator that binds to the PU-box (5'-GAGGAA-3') and acts as a lymphoid-specific enhancer.	Chromatin/transcription
SUV420H1	suppressor of variegation 4-20 homolog 1	SUV420H1 (MIM 610881) function as histone methyltransferases that specifically trimethylate nucleosomal histone H4 on lysine-20 (K20).	Chromatin/transcription
SUV420H2	suppressor of variegation 4-20 homolog 2	SUV420H2 and the related enzyme SUV420H1 (MIM 610881) function as histone methyltransferases that specifically trimethylate nucleosomal histone H4 on lysine-20 (K20).	Chromatin/transcription
TFAP2A	transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)	The protein is a transcription factor that binds the consensus sequence 5'-GCCNNNGGC-3'. The protein functions as either a homodimer or as a heterodimer with similar family members. This protein activates the transcription of some genes while inhibiting the transcription of others. Defects in this gene are a cause of branchiooculofacial syndrome (BOFS).	Cell polarity/morphogenesis
TOP2A	topoisomerase (DNA) II alpha 170kDa	It is a DNA topoisomerase, an enzyme that controls and alters the topologic states of DNA during transcription. This nuclear enzyme is involved in processes such as chromosome condensation, chromatid separation, and the relief of torsional stress that occurs during DNA transcription and replication. It catalyzes the transient breaking and rejoining of two strands of duplex DNA which allows the strands to pass through one another, thus altering the topology of DNA. Two forms of this enzyme exist as likely products of a gene duplication event. The gene encoding this form, alpha, is localized to chromosome 17 and the beta gene is localized to chromosome 3. The gene encoding this enzyme functions as the target for several anticancer agents and a variety of mutations in this gene have been associated with the development of drug resistance. Reduced activity of this enzyme may also play a role in ataxia-telangiectasia.	DNA replication/repair/HR/cohesion
VHL	von Hippel-Lindau tumor suppressor, E3 ubiquitin protein ligase	A germline mutation of this gene is the basis of familial inheritance of VHL syndrome. The protein encoded by this gene is a component of the protein complex that includes elongin B, elongin C, and cullin-2, and possesses ubiquitin ligase E3 activity. This protein is involved in the ubiquitination and degradation of hypoxia-inducible-factor (HIF), which is a transcription factor that plays a central role in the regulation of gene expression by oxygen. RNA polymerase II subunit POLR2G/RPB7 is also reported to be a target of this protein.	Protein degradation/proteosome

ZBTB16	zinc finger and BTB domain containing 16	The protein is a member of the Krueppel C2H2-type zinc-finger protein family and encodes a zinc finger transcription factor that contains nine Kruppel-type zinc finger domains at the carboxyl terminus. This protein is located in the nucleus, is involved in cell cycle progression, and interacts with a histone deacetylase. Specific instances of aberrant gene rearrangement at this locus have been associated with acute promyelocytic leukemia (APL).	Protein degradation/ proteosome

Table 5: Interactors of human p107 protein

Gene Name	Protein Name	Description
AATF	apoptosis antagonizing transcription factor	*See table 4*
ARID4A	AT rich interactive domain 4A (RBP1-like)	*See table 4*
BEGAIN	brain-enriched guanylate kinase-associated homolog	May sustain the structure of the postsynaptic density (PSD).
BRCA1	breast cancer 1, early onset	*See table 4*
BRF1	BRF1 homolog, subunit of RNA polymerase III transcription initiation factor IIIB (S. cerevisiae)	*See table 4*
CCNA1	cyclin A1	*See table 4*
CCNA2	cyclin A2	*See table 4*
CCNE1	cyclin E1	*See table 4*
CDK2	cyclin-dependent kinase 2	*See table 4*
CDKN1C	cyclin-dependent kinase inhibitor 1C (p57, Kip2)	This gene is imprinted, with preferential expression of the maternal allele. The encoded protein is a tight-binding, strong inhibitor of several G1 cyclin/Cdk complexes and a negative regulator of cell proliferation. Mutations in this gene are implicated in sporadic cancers and Beckwith-Wiedemann syndorome, suggesting that this gene is a tumor suppressor candidate.
CREG1	cellular repressor of E1A-stimulated genes 1	*See table 4*
DHX30	DEAH (Asp-Glu-Ala-His) box polypeptide 30	DEAD box proteins, characterized by the conserved motif Asp-Glu-Ala-Asp (DEAD), are putative RNA helicases. They are implicated in a number of cellular processes involving alteration of RNA secondary structure such as translation initiation, nuclear and mitochondrial splicing, and ribosome and spliceosome assembly. Based on their distribution patterns, some members of this DEAD box protein family are believed to be involved in embryogenesis, spermatogenesis, and cellular growth and division. This gene encodes a member of this family. The encoded protein has 97% sequence identity with the mouse HELG protein.
E2F1	E2F transcription factor 1	*See table 4*
E2F3	E2F transcription factor 3	*See table 4*
E2F4	E2F transcription factor 4, p107/p130-binding	*See table 4*
E2F5	E2F transcription factor 5, p130-binding	This protein is differentially phosphorylated and is expressed in a wide variety of human tissues. It has higher identity to E2F4 than to other family members. Both this protein and E2F4 interact with tumor suppressor proteins p130 and p107, but not with pRB.
EMD	emerin	Emerin is a serine-rich nuclear membrane protein and a member of the nuclear lamina-associated protein family. It mediates membrane anchorage to the cytoskeleton. Dreifuss-Emery muscular dystrophy is an X-linked inherited degenerative myopathy resulting from mutation in the emerin gene.
HDAC1	histone deacetylase 1	*See table 4*
HDAC2	histone deacetylase 2	*See table 4*
HDAC3	histone deacetylase 3	*See table 4*
IRF3	interferon regulatory factor 3	*See table 4*
KDM5A	lysine (K)-specific demethylase 5A	*See table 4*
LIN37	lin-37 homolog (*C. elegans*)	*See table 4*
LIN54	lin-54 homolog (*C. elegans*)	*See table 4*
LIN9	lin-9 homolog (*C. elegans*)	*See table 4*
MAPK6	mitogen-activated protein kinase 6	The protein encoded by this gene is a member of the Ser/Thr protein kinase family, and is most closely related to mitogen-activated protein kinases (MAP kinases). MAP kinases also known as extracellular signal-regulated kinases (ERKs), are activated through protein phosphorylation cascades and act as integration points for multiple biochemical signals. This kinase is localized in the nucleus, and has been reported to be activated in fibroblasts upon treatment with serum or phorbol esters.
MCM7	minichromosome maintenance complex component 7	*See table 4*
MYBL2	v-myb myeloblastosis viral oncogene homolog (avian)-like 2	The protein encoded by this gene, a member of the MYB family of transcription factor genes, is a nuclear protein involved in cell cycle progression. The encoded protein is phosphorylated by cyclin A/cyclin-dependent kinase 2 during the S-phase of the cell cycle and possesses both activator and repressor activities. It has been shown to activate the cell division cycle 2, cyclin D1, and insulin-like growth factor-binding protein 5 genes.
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	*See table 4*
NR2E3	nuclear receptor subfamily 2, group E, member 3	This protein is part of a large family of nuclear receptor transcription factors involved in signaling pathways. Nuclear receptors have been shown to regulate pathways involved in embryonic development, as well as in maintenance of proper cell function in adults. Members of this family are characterized by discrete domains that function in DNA and ligand binding. This gene encodes a retinal nuclear receptor that is a ligand-dependent transcription factor.

PHB	prohibitin	*See table 4*
PPP1CA	protein phosphatase 1, catalytic subunit, alpha isozyme	*See table 4*
PPP2R3A	protein phosphatase 2, regulatory subunit B'', alpha	This gene encodes one of the regulatory subunits of the protein phosphatase 2. Protein phosphatase 2 (formerly named type 2A) is one of the four major Ser/Thr phosphatases and is implicated in the negative control of cell growth and division. Protein phosphatase 2 holoenzymes are heterotrimeric proteins composed of a structural subunit A, a catalytic subunit C, and a regulatory subunit B. The regulatory subunit is encoded by a diverse set of genes that have been grouped into the B/PR55, B'/PR61, and B''/PR72 families. These different regulatory subunits confer distinct enzymatic specificities and intracellular localizations to the holozenzyme. The product of this gene belongs to the B'' family. The B'' family has been further divided into subfamilies. The product of this gene belongs to the alpha subfamily of regulatory subunit B''. Alternative splicing results in multiple transcript variants encoding different isoforms.
RB1	retinoblastoma 1	The protein encoded by this gene is a negative regulator of the cell cycle and was the first tumor suppressor gene found. The encoded protein also stabilizes constitutive heterochromatin to maintain the overall chromatin structure. The active, hypophosphorylated form of the protein binds transcription factor E2F1. Defects in this gene are a cause of childhood cancer retinoblastoma (RB), bladder cancer, and osteogenic sarcoma.
RBBP8	retinoblastoma binding protein 8	*See table 4*
RBBP9	retinoblastoma binding protein 9	*See table 4*
RBL2	retinoblastoma-like 2 (p130)	*See table 4*
SMAD3	SMAD family member 3	The protein belongs to the SMAD, a family of proteins similar to the gene products of the Drosophila gene 'mothers against decapentaplegic' (Mad) and the *C. elegans* gene Sma. SMAD proteins are signal transducers and transcriptional modulators that mediate multiple signaling pathways. This protein functions as a transcriptional modulator activated by transforming growth factor-beta and is thought to play a role in the regulation of carcinogenesis.
SMAD4	SMAD family member 4	Smad proteins are phosphorylated and activated by transmembrane serine-threonine receptor kinases in response to TGF-beta signaling. The product of this gene forms homomeric complexes and heteromeric complexes with other activated Smad proteins, which then accumulate in the nucleus and regulate the transcription of target genes. This protein binds to DNA and recognizes an 8-bp palindromic sequence (GTCTAGAC) called the Smad-binding element (SBE). The Smad proteins are subject to complex regulation by post-translational modifications. Mutations or deletions in this gene have been shown to result in pancreatic cancer, juvenile polyposis syndrome, and hereditary hemorrhagic telangiectasia syndrome.
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	*See table 4*
SNW1	SNW domain containing 1	*See table 4*
SUMO2	SMT3 suppressor of mif two 3 homolog 2 (S. cerevisiae)	The protein is a member of the SUMO protein family. It functions in a manner similar to ubiquitin in that it is bound to target proteins as part of a post-translational modification system. However, unlike ubiquitin which targets proteins for degradation, this protein is involved in a variety of cellular processes, such as nuclear transport, transcriptional regulation, apoptosis, and protein stability. It is not active until the last two amino acids of the carboxy-terminus have been cleaved off. Numerous pseudogenes have been reported for this gene.
SUMO3	SMT3 suppressor of mif two 3 homolog 3 (S. cerevisiae)	SUMO proteins, such as SUMO3, and ubiquitin posttranslationally modify numerous cellular proteins and affect their metabolism and function. However, unlike ubiquitination, which targets proteins for degradation, sumoylation participates in a number of cellular processes, such as nuclear transport, transcriptional regulation, apoptosis, and protein stability.
SUV39H1	suppressor of variegation 3-9 homolog 1 (Drosophila)	*See table 4*
TAF1	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 250kDa	*See table 4*
TFDP1	transcription factor Dp-1	*See table 4*
TFDP2	transcription factor Dp-2 (E2F dimerization partner 2)	*See table 4*
TOP1	topoisomerase (DNA) I	DNA topoisomerase controls and alters the topologic states of DNA during transcription. This enzyme catalyzes the transient breaking and rejoining of a single strand of DNA which allows the strands to pass through one another, thus altering the topology of DNA.
UBC	ubiquitin C	*See table 4*
USP4	ubiquitin specific peptidase 4 (proto-oncogene)	*See table 4*

Table 6: Interactors of human p130 protein

Gene Name	Protein Name	Description
AATF	apoptosis antagonizing transcription factor	*See table 4*
BRCA1	breast cancer 1, early onset	*See table 4*
BRF1	BRF1 homolog, subunit of RNA polymerase III transcription initiation factor IIIB (S. cerevisiae)	*See table 4*
CCNA1	cyclin A1	*See table 4*
CCNA2	cyclin A2	*See table 4*
CCNE1	cyclin E1	*See table 4*
CDK2	cyclin-dependent kinase 2	*See table 4*
CUL1	cullin 1	Core component of multiple cullin-RING-based SCF (SKP1-CUL1-F-box protein) E3 ubiquitin-protein ligase complexes, which mediate the ubiquitination of proteins involved in cell cycle progression, signal transduction and transcription. In the SCF complex, serves as a rigid scaffold that organizes the SKP1-F-box protein and RBX1 subunits. May contribute to catalysis through positioning of the substrate and the ubiquitin-conjugating enzyme. The E3 ubiquitin-protein ligase activity of the complex is dependent on the neddylation of the cullin subunit and is inhibited by the association of the deneddylated cullin subunit with TIP120A/CAND1. The functional specificity of the SCF complex depends on the F-box protein as substrate recognition component. SCF(BTRC) and SCF(FBXW11) direct ubiquitination of CTNNB1 and participate in Wnt signaling.
E2F4	E2F transcription factor 4, p107/p130-binding	*See table 4*
E2F5	E2F transcription factor 5, p130-binding	This protein is differentially phosphorylated and is expressed in a wide variety of human tissues. It has higher identity to E2F4 than to other family members. Both this protein and E2F4 interact with tumor suppressor proteins p130 and p107, but not with pRB.
ELAVL1	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 1 (Hu antigen R)	The protein is a member of the ELAVL protein family. This encoded protein contains 3 RNA-binding domains and binds cis-acting AU-rich elements. It destabilizes mRNAs and thereby regulates gene expression.
EZH2	enhancer of zeste homolog 2 (Drosophila)	The protein is a member of the Polycomb-group (PcG) family. PcG family members form multimeric protein complexes, which are involved in maintaining the transcriptional repressive state of genes over successive cell generations. This protein associates with the embryonic ectoderm development protein, the VAV1 oncoprotein, and the X-linked nuclear protein. This protein may play a role in the hematopoietic and central nervous systems. Multiple alternatively splcied transcript variants encoding distinct isoforms have been identified for this gene.
HBP1	HMG-box transcription factor 1	*See table 4*
HDAC1	histone deacetylase 1	*See table 4*
ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein	*See table 4*
IRF3	interferon regulatory factor 3	*See table 4*
LIN37	lin-37 homolog (C. elegans)	*See table 4*
LIN52	lin-52 homolog (C. elegans)	LIN52 is a component of the LIN, or DREAM, complex, an essential regulator of cell cycle genes
LIN54	lin-54 homolog (C. elegans)	*See table 4*
LIN9	lin-9 homolog (C. elegans)	*See table 4*
MCM7	minichromosome maintenance complex component 7	*See table 4*
PCBD1	pterin-4 alpha-carbinolamine dehydratase/dimerization cofactor of hepatocyte nuclear factor 1 alpha	This gene encodes pterin-4 alpha-carbinolamine dehydratase, an enzyme involved in phenylalanine hydroxylation. A deficiency of this enzyme leads to hyperphenylalaninemia. The enzyme regulates the homodimerization of the transcription factor hepatocyte nuclear factor 1 (HNF1).
PHB	prohibitin	*See table 4*
PPP2CA	protein phosphatase 2, catalytic subunit, alpha isozyme	This gene encodes the phosphatase 2A catalytic subunit. Protein phosphatase 2A is one of the four major Ser/Thr phosphatases, and it is implicated in the negative control of cell growth and division. It consists of a common heteromeric core enzyme, which is composed of a catalytic subunit and a constant regulatory subunit that associates with a variety of regulatory subunits. This gene encodes an alpha isoform of the catalytic subunit.
RAF1	v-raf-1 murine leukemia viral oncogene homolog 1	*See table 4*

RB1	retinoblastoma 1	The protein is a negative regulator of the cell cycle and was the first tumor suppressor gene found. The encoded protein also stabilizes constitutive heterochromatin to maintain the overall chromatin structure. The active, hypophosphorylated form of the protein binds transcription factor E2F1. Defects in this gene are a cause of childhood cancer retinoblastoma (RB), bladder cancer, and osteogenic sarcoma.
RBBP4	retinoblastoma binding protein 4	*See table 4*
RBBP8	retinoblastoma binding protein 8	*See table 4*
RBBP9	retinoblastoma binding protein 9	*See table 4*
RBL1	retinoblastoma-like 1 (p107)	*See table 4*
SERPINB2	serpin peptidase inhibitor, clade B (ovalbumin), member 2	*See table 4*
SKP1	S-phase kinase-associated protein 1	This gene encodes a component of SCF complexes, which are composed of this protein, cullin 1, a ring-box protein, and one member of the F-box family of proteins. This protein binds directly to the F-box motif found in F-box proteins. SCF complexes are involved in the regulated ubiquitination of specific protein substrates, which targets them for degradation by the proteosome. Specific F-box proteins recognize different target protein(s), and many specific SCF substrates have been identified including regulators of cell cycle progression and development. Studies have also characterized the protein as an RNA polymerase II elongation factor.
SKP2	S-phase kinase-associated protein 2, E3 ubiquitin protein ligase	*See table 4*
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	*See table 4*
SNW1	SNW domain containing 1	*See table 4*
SUV39H1	suppressor of variegation 3-9 homolog 1 (Drosophila)	*See table 4*
TAF1	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 250kDa	*See table 4*
TFDP1	transcription factor Dp-1	*See table 4*
TFDP2	transcription factor Dp-2 (E2F dimerization partner 2)	*See table 4*
TOP1	topoisomerase (DNA) I	This gene encodes a DNA topoisomerase, an enzyme that controls and alters the topologic states of DNA during transcription. This enzyme catalyzes the transient breaking and rejoining of a single strand of DNA which allows the strands to pass through one another, thus altering the topology of DNA. This gene is localized to chromosome 20 and has pseudogenes which reside on chromosomes 1 and 22.
TP63	tumor protein p63	This gene encodes a member of the p53 family of transcription factors. An animal model, p63 -/- mice, has been useful in defining the role this protein plays in the development and maintenance of stratified epithelial tissues. p63 -/- mice have several developmental defects which include the lack of limbs and other tissues, such as teeth and mammary glands, which develop as a result of interactions between mesenchyme and epithelium. Mutations in this gene are associated with ectodermal dysplasia, and cleft lip/palate syndrome 3 (EEC3); split-hand/foot malformation 4 (SHFM4); ankyloblepharon-ectodermal defects-cleft lip/palate; ADULT syndrome (acro-dermato-ungual-lacrimal-tooth); limb-mammary syndrome; Rap-Hodgkin syndrome (RHS); and orofacial cleft 8.
UBC	ubiquitin C	*See table 4*
USP4	ubiquitin specific peptidase 4 (proto-oncogene)	*See table 4*
XBP1	X-box binding protein 1	The transcription factor regulates MHC class II genes by binding to a promoter element referred to as an X box. This gene product is a bZIP protein, which was also identified as a cellular transcription factor that binds to an enhancer in the promoter of the T cell leukemia virus type 1 promoter. It may increase expression of viral proteins by acting as the DNA binding partner of a viral transactivator. It has been found that upon accumulation of unfolded proteins in the endoplasmic reticulum (ER), the mRNA of this gene is processed to an active form by an unconventional splicing mechanism that is mediated by the endonuclease inositol-requiring enzyme 1 (IRE1).

Table 7: Interactors of Whi5 protein

Gene Name	Genetic/Physical Interactors (G/P)	Protein Name	Description	Functional classification in Costanzo et al. 2010
ADA2	G	Transcriptional adapter 2	Transcription coactivator, component of the ADA and SAGA transcriptional adaptor/HAT (histone acetyltransferase) complexes	chromatin/transcription
AIM31	G	Altered inheritance of mitochondria protein 31, mitochondrial	Putative protein of unknown function; GFP-fusion protein localizes to mitochondria; may interact with respiratory chain complexes III or IV; null mutant is viable and displays reduced frequency of mitochondrial genome loss	unknown
APA1	G	AP4A phosphorylase	AP4A phosphorylase; bifunctional diadenosine 5',5'''-P1,P4-tetraphosphate phosphorylase and ADP sulfurylase involved in catabolism of bis(5'-nucleosidyl) tetraphosphates; catalyzes phosphorolysis of dinucleoside oligophosphates, cleaving the substrates' alpha/beta-anhydride bond and introducing Pi into the beta-position of the corresponding NDP formed; APA1 and APA2 are paralogs arising from whole genome duplication; protein abundance increases under DNA replication stress	metabolism/mitochondria
APL3	G	Clathrin Adaptor Protein complex Large chain AP-2 complex subunit alpha	Alpha-adaptin, large subunit of the clathrin associated protein complex (AP-2); involved in vesicle mediated transport	cell polarity/morphogenesis
ASK10	G	Activator of Skn7	Component of RNA polymerase II holoenzyme; phosphorylated in response to oxidative stress; has a role in destruction of Ssn8p; proposed to function in activation of the glycerol channel Fps1p; ASK10 has a paralog, RGC1, that arose from the whole genome duplication	chromatin/transcription
ATG1	P	Autophagy related Serine/threonine-protein kinase	Protein ser/thr kinase required for vesicle formation in autophagy and the cytoplasm-to-vacuole targeting (Cvt) pathway; structurally required for phagophore assembly site formation; during autophagy forms a complex with Atg13p and Atg17p	Autophagy
BCK2	G	Bypass of C Kinase	Protein rich in serine and threonine residues involved in protein kinase C signaling pathway, which controls cell integrity; overproduction suppresses pkc1 mutations	G1/S and G2/M cell cycle progression/meiosis;signaling/stress response
CAJ1	G	Protein CAJ1	Nuclear type II J heat shock protein of the E. coli dnaJ family, contains a leucine zipper-like motif, binds to non-native substrates for presentation to Ssa3p, may function during protein translocation, assembly and disassembly	unknown
CBC2	G	Nuclear cap-binding protein subunit 2	Small subunit of the heterodimeric cap binding complex that also contains Sto1p, component of the spliceosomal commitment complex; interacts with Npl3p, possibly to package mRNA for export from the nucleus; contains an RNA-binding motif	RNA processing
CCR4	G	Glucose-repressible alcohol dehydrogenase transcriptional effector	Component of the CCR4-NOT transcriptional complex, which is involved in regulation of gene expression; component of the major cytoplasmic deadenylase, which is involved in mRNA poly(A) tail shortening	chromatin/transcription;RNA processing
CDC28	G/P	Cyclin-dependent kinase 1	Catalytic subunit of the main cell cycle cyclin-dependent kinase (CDK); alternately associates with G1 cyclins (CLNs) and G2/M cyclins (CLBs) which direct the CDK to specific substrates; involved in modulating membrane trafficking dynamics; protein abundance increases in response to DNA replication stress	G1/S and G2/M cell cycle progression/meiosis;signaling/stress response
CDH1	G	CDC20 Homolog APC/C activator	Cell-cycle regulated activator of the anaphase-promoting complex/cyclosome (APC/C), which directs ubiquitination of cyclins resulting in mitotic exit; targets the APC/C to specific substrates including Cdc20p, Ase1p, Cin8p and Fin1p	protein degradation/proteosome;chromosome segregation/kinetochore/spindle/microtubule
CKB2	G	Casein Kinase II Beta' subunit	Beta' regulatory subunit of casein kinase 2 (CK2), a Ser/Thr protein kinase with roles in cell growth and	signaling/stress response

			proliferation; CK2, comprised of CKA1, CKA2, CKB1 and CKB2, has many substrates including transcription factors and all RNA polymerase	
CLN1	G	G1/S-specific cyclin 1	G1 cyclin involved in regulation of the cell cycle; activates Cdc28p kinase to promote the G1 to S phase transition; late G1 specific expression depends on transcription factor complexes, MBF (Swi6p-Mbp1p) and SBF (Swi6p-Swi4p)	G1/S and G2/M cell cycle progression/ meiosis;signaling/stress response
CLN2	G	G1/S-specific cyclin 2	G1 cyclin involved in regulation of the cell cycle; activates Cdc28p kinase to promote the G1 to S phase transition; late G1 specific expression depends on transcription factor complexes, MBF (Swi6p-Mbp1p) and SBF (Swi6p-Swi4p)	G1/S and G2/M cell cycle progression/ meiosis;signaling/stress response
CLN3	G	G1/S-specific cyclin 3	G1 cyclin involved in cell cycle progression; activates Cdc28p kinase to promote the G1 to S phase transition; plays a role in regulating transcription of the other G1 cyclins, CLN1 and CLN2; regulated by phosphorylation and proteolysis	G1/S and G2/M cell cycle progression/ meiosis;signaling/stress response
COX10	G	Protoheme IX farnesyltransferase, mitochondrial	Heme A:farnesyltransferase, catalyzes the first step in the conversion of protoheme to the heme A prosthetic group required for cytochrome c oxidase activity; human ortholog is associated with mitochondrial disorders	metabolism/ mitochondria
CPR7	G	Cyclosporin-sensitive Proline Rotamase	Peptidyl-prolyl cis-trans isomerase (cyclophilin), catalyzes the cis-trans isomerization of peptide bonds N-terminal to proline residues; binds to Hsp82p and contributes to chaperone activity	ER<->Golgi traffic
CSM3	G	Chromosome Segregation in Meiosis protein 3	Replication fork associated factor, required for stable replication fork pausing; component of the DNA replication checkpoint pathway; required for accurate chromosome segregation during meiosis	DNA replication/repair/HR/cohesion
CTF8	G	Chromosome Transmission Fidelity protein 8	Subunit of a complex with Ctf18p that shares some subunits with Replication Factor C and is required for sister chromatid cohesion	DNA replication/repair/HR/cohesion
DUN1	G	DNA damage response protein kinase	Cell-cycle checkpoint serine-threonine kinase required for DNA damage-induced transcription of certain target genes, phosphorylation of Rad55p and Sml1p, and transient G2/M arrest after DNA damage; also regulates postreplicative DNA repair	DNA replication/repair/HR/cohesion
EAF1	G	ESA1p-Associated Factor	Component of the NuA4 histone acetyltransferase complex; acts as a platform for assembly of NuA4 subunits into the native complex; required for initiation of pre-meiotic DNA replication, likely due to its requirement for expression of IME1	chromatin/transcription
ECM8	G	ExtraCellular Mutant 8	Non-essential protein of unknown function	unknown
ELM1	G	Elongated Morphology 1	Serine/threonine protein kinase that regulates cellular morphogenesis, septin behavior, and cytokinesis; required for the regulation of other kinases; forms part of the bud neck ring	cell polarity/morphogenesis;chromosome segregation/kinetochore/spindle/microtubule
EMI1	G	Early Meiotic Induction protein 1	Non-essential protein required for transcriptional induction of the early meiotic-specific transcription factor IME1, also required for sporulation; contains twin cysteine-x9-cysteine motifs	unknown
EMP46	G	Protein EMP46	Integral membrane component of endoplasmic reticulum-derived COPII-coated vesicles, which function in ER to Golgi transport	ER<->Golgi traffic
ENT5	G	Epsin N-Terminal homology protein 5	Protein containing an N-terminal epsin-like domain involved in clathrin recruitment and traffic between the Golgi and endosomes; associates with the clathrin adaptor Gga2p, clathrin adaptor complex AP-1, and clathrin	Golgi/endosome/vacuole/sorting
ESA1	G/P	Histone acetyltransferase	Catalytic subunit of the histone acetyltransferase complex (NuA4) that acetylates four conserved internal lysines of histone H4 N-terminal tail; required for cell cycle progression and transcriptional silencing at the rDNA locus	chromatin/transcription
FMC1	G	ATP synthase assembly factor FMC1, mitochondrial	Mitochondrial matrix protein, required for assembly or stability at high temperature of the F1 sector of mitochondrial F1F0 ATP synthase; null mutant temperature sensitive growth on glycerol is suppressed by multicopy expression of Odc1p	drug/ion transport;metabolism/mitochondria
FYV7	G	rRNA-processing protein	Essential protein required for maturation of 18S rRNA; required for survival upon exposure to K1 killer toxin	ribosome/translation

GPX2	G	Glutathione Peroxidase 2	Phospholipid hydroperoxide glutathione peroxidase; protects cells from phospholipid hydroperoxides and nonphospholipid peroxides during oxidative stress; induced by glucose starvation; protein abundance increases in response to DNA replication stress	metabolism/ mitochondria
GSH1	G	Glutamate--cysteine ligase	Gamma glutamylcysteine synthetase; catalyzes the first step in glutathione (GSH) biosynthesis; expression induced by oxidants, cadmium, and mercury; protein abundance increases in response to DNA replication stress	metabolism/ mitochondria
HMT1	G	HnRNP Methyl Transferase	Nuclear SAM-dependent mono- and asymmetric arginine dimethylating methyltransferase that modifies hnRNPs, including Npl3p and Hrp1p, affecting their activity and nuclear export; methylates U1 snRNP protein Snp1p and ribosomal protein Rps2p	ribosome/translation;nuclear-cytoplasic transport;RNA processing
HOS1	P	Histone deacetylase	Class I histone deacetylase (HDAC) family member that deacetylates Smc3p on lysine residues at anaphase onset; has sequence similarity to Hda1p, Rpd3p, Hos2p, and Hos3p; interacts with the Tup1p-Ssn6p corepressor complex	chromatin/transcription
HOS3	P	Histone deacetylase	Trichostatin A-insensitive homodimeric histone deacetylase (HDAC) with specificity in vitro for histones H3, H4, H2A, and H2B; similar to Hda1p, Rpd3p, Hos1p, and Hos2p; deletion results in increased histone acetylation at rDNA repeats	chromatin/transcription
HSL1	G	Probable serine/threonine-protein kinase	Nim1p-related protein kinase that regulates the morphogenesis and septin checkpoints; associates with the assembled septin filament; required along with Hsl7p for bud neck recruitment, phosphorylation, and degradation of Swe1p	cell polarity/morphogenesis;G1/S and G2/M cell cycle progression/ meiosis
HSP82	P	ATP-dependent molecular chaperone	Hsp90 chaperone; required for pheromone signaling and negative regulation of Hsf1p; docks with Tom70p for mitochondrial preprotein delivery; promotes telomerase DNA binding and nucleotide addition; interacts with Cns1p, Cpr6p, Cpr7p, Sti1p; protein abundance increases in response to DNA replication stress	signaling/stress response
HST3	G	Homolog of SIR Two (SIR2) NAD-dependent histone deacetylase	Member of the Sir2 family of NAD(+)-dependent protein deacetylases; involved along with Hst4p in telomeric silencing, cell cycle progression, radiation resistance, genomic stability and short-chain fatty acid metabolism	DNA replication/repair/HR/cohesion
HTA1	G	Histone H2 A	Histone H2A, core histone protein required for chromatin assembly and chromosome function; one of two nearly identical subtypes (see also HTA2); DNA damage-dependent phosphorylation by Mec1p facilitates DNA repair; acetylated by Nat4p	chromatin/transcription
HXT17	G	Hexose Transporter	Hexose transporter, up-regulated in media containing raffinose and galactose at pH 7.7 versus pH 4.7, repressed by high levels of glucose	drug/ion transport;metabolism/mitochondria
IKI3	G	Elongator complex protein 1	Subunit of Elongator complex, which is required for modification of wobble nucleosides in tRNA; maintains structural integrity of Elongator; homolog of human IKAP, mutations in which cause familial dysautonomia (FD)	ribosome/translation
IME2	G	Serine/threonine protein kinase (Inducer of Meiosis)	Serine/threonine protein kinase involved in activation of meiosis, associates with Ime1p and mediates its stability, activates Ndt80p; IME2 expression is positively regulated by Ime1p	G1/S and G2/M cell cycle progression/ meiosis
ISW2	G	ISWI chromatin-remodeling complex ATPase	ATP-dependent DNA translocase involved in chromatin remodeling; ATPase component that, with Itc1p, forms a complex required for repression of a-specific genes, INO1, and early meiotic genes during mitotic growth	chromatin/transcription
KEL1	G	Kelch repeat-containing protein 1	Protein required for proper cell fusion and cell morphology; functions in a complex with Kel2p to negatively regulate mitotic exit, interacts with Tem1p and Lte1p; localizes to regions of polarized growth; potential Cdc28p substrate	chromosome segregation/kinetochore/spindle/microtubule
KRE28	G		Subunit of a kinetochore-microtubule binding complex with Spc105p that bridges centromeric heterochromatin and kinetochore MAPs and motors, and is also required for sister chromatid bi-orientation and kinetochore binding of SAC components	chromosome segregation/kinetochore/spindle/microtubule
LTE1	G	Guanine nucleotide exchange factor	Protein similar to GDP/GTP exchange factors but without detectable GEF activity; required for asymmetric localization of Bfa1p at daughter-directed spindle pole bodies and for mitotic exit at low temperatures	chromosome segregation/kinetochore/spindle/microtubule

MAD2	G	Mitotic spindle checkpoint component	Component of the spindle-assembly checkpoint complex; delays the onset of anaphase in cells with defects in mitotic spindle assembly; forms a complex with Mad1p; regulates APC/C activity during prometaphase and metaphase of meiosis I	chromosome segregation/kinetochore/spindle/microtubule
MBF1	G	Multiprotein Bridging Factor 1	Transcriptional coactivator; bridges the DNA-binding region of Gcn4p and TATA-binding protein Spt15p; suppressor of frameshift mutations; protein abundance increases in response to DNA replication stress	metabolism/mitochondria; chromatin/transcription
MET12	G	Methylenetetrahydrofolate reductase 1	Protein with methylenetetrahydrofolate reductase (MTHFR) activity in vitro; null mutant has no phenotype and is prototrophic for methionine; MET13 encodes major isozyme of MTHFR	metabolism/mitochondria
MFB1	G	Mitochondria-associated F-Box protein	Mitochondria-associated F-box protein involved in maintenance of normal mitochondrial morphology; interacts with Skp1p through the F-box motif; preferentially localizes to the mother cell during budding	metabolism/mitochondria
MNN11	G	Probable alpha-1,6-mannosyltransferase	Subunit of a Golgi mannosyltransferase complex that also contains Anp1p, Mnn9p, Mnn10p, and Hoc1p, and mediates elongation of the polysaccharide mannan backbone; has homology to Mnn10p	protein folding/protein glycosylation/cell wall biogenesis&integrity
MPS3	G	MonoPolar Spindle	Nuclear envelope protein required for SPB duplication and nuclear fusion; localizes to the SPB half bridge and at telomeres during meiosis; required with Ndj1p and Csm4p for meiotic bouquet formation and telomere-led rapid prophase movement	chromosome segregation/kinetochore/spindle/microtubule
MSC7	G	Putative aldehyde dehydrogenase-like protein	Protein of unknown function, green fluorescent protein (GFP)-fusion protein localizes to the endoplasmic reticulum; msc7 mutants are defective in directing meiotic recombination events to homologous chromatids	unknown
MSG5	G	Tyrosine-protein phosphatase (Multicopy Suppressor of GPA1)	Dual-specificity protein phosphatase; exists in 2 isoforms; required for maintenance of a low level of signaling through the cell integrity pathway, adaptive response to pheromone; regulates and is regulated by Slt2p; dephosphorylates Fus3p	protein folding/protein glycosylation/cell wall biogenesis&integrity
MSI1	G	Chromatin assembly factor 1 subunit p50 (Multicopy Suppressor of IRA1)	Subunit of chromatin assembly factor I (CAF-1); chromatin assembly by CAF-1 is important for multiple processes including silencing at telomeres, mating type loci, and rDNA; maintenance of kinetochore structure; deactivation of the DNA damage checkpoint after DNA repair; and chromatin dynamics during transcription; Msi1p localizes to both nucleus and cytoplasm and has an independent role as a negative regulator of the RAS/cAMP pathway via sequestration of Npr1p kinase	chromatin/transcription
MSN5	P	Multicopy suppressor of SNF1 mutation	Karyopherin involved in nuclear import and export of proteins, including import of replication protein A and export of Swi6p, Far1p, and Pho4p; required for re-export of mature tRNAs after their retrograde import from the cytoplasm	G1/S and G2/M cell cycle progression/meiosis;nuclear-cytoplasic transport
NPT1	G	Nicotinate Phosphoribosyl Transferase	Nicotinate phosphoribosyltransferase, acts in the salvage pathway of NAD+ biosynthesis; required for silencing at rDNA and telomeres and has a role in silencing at mating-type loci; localized to the nucleus	cell polarity/morphogenesis;chromatin/transcription
OM45	G	Mitochondrial outer membrane protein	Mitochondrial outer membrane protein of unknown function; major constituent of the outer membrane, located on the outer (cytosolic) face; protein abundance increases in response to DNA replication stress	metabolism/mitochondria
PCI8	G	Proteasome-COP9 signalosome (CSN)-eIF3	Possible shared subunit of Cop9 signalosome (CSN) and eIF3, binds eIF3b subunit Prt1p, has possible dual functions in transcriptional and translational control, contains a PCI (Proteasome-COP9 signalosome (CSN)-eIF3) domain	protein degradation/proteosome
PCL1	G	Pho85 Cyclin 1	Cyclin, interacts with cyclin-dependent kinase Pho85p; member of the Pcl1,2-like subfamily, involved in the regulation of polarized growth and morphogenesis and progression through the cell cycle; localizes to sites of polarized cell growth	cell polarity/morphogenesis;G1/S and G2/M cell cycle progression/meiosis;signaling/stress response
PCL9	G/P	Pho85 Cyclin 9	Cyclin, forms a functional kinase complex with Pho85p cyclin-dependent kinase (Cdk), expressed in late M/early G1 phase, activated by Swi5p	G1/S and G2/M cell

				cycle progression/ meiosis;signaling/stress response
PEP1	G	Carboxypeptidase Y-deficient (Vacuolar protein sorting/targeting protein VPS10)	Type I transmembrane sorting receptor for multiple vacuolar hydrolases; cycles between the late-Golgi and prevacuolar endosome-like compartments	Golgi/endosome/vacuole/sorting
PEX30	G	Peroxisomal membrane protein	Peroxisomal integral membrane protein, involved in negative regulation of peroxisome number; partially functionally redundant with Pex31p; genetic interactions suggest action at a step downstream of steps mediated by Pex28p and Pex29p	Unknown
PHO85	G/P	Cyclin-dependent protein kinase	Cyclin-dependent kinase, with ten cyclin partners; involved in regulating the cellular response to nutrient levels and environmental conditions and progression through the cell cycle	G1/S and G2/M cell cycle progression/ meiosis;chromatin/transcription
PKP2	P	Pyruvate dehydrogenase kinase 2, mitochondrial	Mitochondrial protein kinase that negatively regulates activity of the pyruvate dehydrogenase complex by phosphorylating the ser-133 residue of the Pda1p subunit; acts in concert with kinase Pkp1p and phosphatases Ptc5p and Ptc6p	metabolism/ mitochondria
PLB2	G	Lysophospholipase 2	Phospholipase B (lysophospholipase) involved in phospholipid metabolism; displays transacylase activity in vitro; overproduction confers resistance to lysophosphatidylcholine	lipid/sterol/fatty acid biosynth
POB3	G	FACT complex subunit	Subunit of the heterodimeric FACT complex (Spt16p-Pob3p); FACT associates with chromatin via interaction with Nhp6Ap and Nhp6Bp, and reorganizes nucleosomes to facilitate access to DNA by RNA and DNA polymerases; protein abundance increases in response to DNA replication stress	chromatin/transcription
POL32	G	DNA polymerase delta subunit 3	Third subunit of DNA polymerase delta, involved in chromosomal DNA replication; required for error-prone DNA synthesis in the presence of DNA damage and processivity; interacts with Hys2p, PCNA (Pol30p), and Pol1p	DNA replication/repair/HR/cohesion
PPH21	G	Serine/threonine-protein phosphatase PP2A-1 catalytic subunit	Catalytic subunit of protein phosphatase 2A (PP2A), functionally redundant with Pph22p; methylated at C terminus; forms alternate complexes with several regulatory subunits; involved in signal transduction and regulation of mitosis	signaling/stress response
PRM7	G	Pheromone-Regulated Membrane protein	Pheromone-regulated protein, predicted to have one transmembrane segment; promoter contains Gcn4p binding elements	unknown
PRP4	G	U4/U6 small nuclear ribonucleoprotein (Pre-mRNA Processing)	Splicing factor, component of the U4/U6-U5 snRNP complex	RNA processing
PRY1	G	Pathogen Related in Yeast	Sterol binding protein involved in the export of acetylated sterols; secreted glycoprotein and member of the CAP protein superfamily (cysteine-rich secretory proteins (CRISP), antigen 5, and pathogenesis related 1 proteins); sterol export function is redundant with that of PRY2; may be involved in detoxification of hydrophobic compounds	unknown
PSR1	G	Phosphatase	Plasma membrane associated protein phosphatase involved in the general stress response; required along with binding partner Whi2p for full activation of STRE-mediated gene expression, possibly through dephosphorylation of Msn2p	signaling/stress response
PTC1	G	Phosphatase type Two C	Type 2C protein phosphatase (PP2C); dephosphorylates Hog1p, inactivating osmosensing MAPK cascade; involved in Fus3p activation during pheromone response; deletion affects precursor tRNA splicing, mitochondrial inheritance, and sporulation	signaling/stress response
PTK2	P	Putative serine/Threonine protein Kinase	Putative serine/threonine protein kinase involved in regulation of ion transport across plasma membrane; enhances spermine uptake	drug/ion transport;signaling/stress response
PTP3	G	Protein Tyrosine Phosphatase 3	Phosphotyrosine-specific protein phosphatase involved in the inactivation of mitogen-activated protein kinase (MAPK) during osmolarity sensing; dephosporylates Hog1p MAPK and regulates its localization; localized to the cytoplasm	protein folding/protein glycosylation/ cell wall

				biogenesis&integrity;cell polarity/morphogenesis;metabolism/mitochondria;signaling/stress response
PYC2	G	Pyruvate Carboxylase 2	Pyruvate carboxylase isoform, cytoplasmic enzyme that converts pyruvate to oxaloacetate; highly similar to isoform Pyc1p but differentially regulated; mutations in the human homolog are associated with lactic acidosis	metabolism/mitochondria
QCR2	G	Ubiquinol-Cytochrome C oxidoReductase	Subunit 2 of the ubiquinol cytochrome-c reductase complex, which is a component of the mitochondrial inner membrane electron transport chain; phosphorylated; transcription is regulated by Hap1p, Hap2p/Hap3p, and heme	metabolism/mitochondria
RAD27	G	Flap endonuclease 1	5' to 3' exonuclease, 5' flap endonuclease, required for Okazaki fragment processing and maturation as well as for long-patch base-excision repair; member of the S. pombe RAD2/FEN1 family	DNA replication/repair/HR/cohesion
RAD52	G	DNA repair and recombination protein RAD52	Protein that stimulates strand exchange by facilitating Rad51p binding to single-stranded DNA; anneals complementary single-stranded DNA; involved in the repair of double-strand breaks in DNA during vegetative growth and meiosis	DNA replication/repair/HR/cohesion
RAD53	G	Serine/threonine-protein kinase RAD53	Protein kinase, required for cell-cycle arrest in response to DNA damage; activated by trans autophosphorylation when interacting with hyperphosphorylated Rad9p; also interacts with ARS1 and plays a role in initiation of DNA replication	DNA replication/repair/HR/cohesion
RAD55	G	DNA repair protein RAD55	Protein that stimulates strand exchange by stabilizing the binding of Rad51p to single-stranded DNA; involved in the recombinational repair of double-strand breaks in DNA during vegetative growth and meiosis; forms heterodimer with Rad57p	DNA replication/repair/HR/cohesion
RAD6	G	Radiation sensitive	Ubiquitin-conjugating enzyme (E2), involved in postreplication repair (as a heterodimer with Rad18p), DSBR and checkpoint control (as a heterodimer with Bre1p), ubiquitin-mediated N-end rule protein degradation (as a heterodimer with Ubr1p	chromosome segregation/kinetochore/spindle/microtubule
RAP1	G	DNA-binding protein RAP1(Repressor Activator Protein)	Essential DNA-binding transcription regulator that binds at many loci; involved in either transcription activation or repression, chromatin silencing, and telomere length maintenance; conserved protein with an N-terminal BRCT domain, a central region with homology to the Myb DNA binding domain, and a C-terminal Rap1-specific protein-interaction domain (RCT domain)	unknown
RGI2	G	Respiratory growth induced	Protein of unknown function involved in energy metabolism under respiratory conditions; expression induced under carbon limitation and repressed under high glucose	unknown
RIM1	G	Single-stranded DNA-binding protein RIM1, mitochondrial	Single-stranded DNA-binding protein essential for mitochondrial genome maintenance; involved in mitochondrial DNA replication	metabolism/mitochondria
RIM15	G	Serine/threonine-protein kinase RIM15	Glucose-repressible protein kinase involved in signal transduction during cell proliferation in response to nutrients, specifically the establishment of stationary phase; identified as a regulator of IME2; substrate of Pho80p-Pho85p kinase	metabolism/mitochondria;signaling/stress response
RKM3	G	Ribosomal lysine (K) Methyltransferase 3	Ribosomal lysine methyltransferase specific for monomethylation of Rpl42ap and Rpl42bp (lysine 40); nuclear SET domain containing protein	unknown
RLF2	G	Rap1 protein Localization Factor	Largest subunit (p90) of the Chromatin Assembly Complex (CAF-1); chromatin assembly by CAF-1 is important for multiple processes including silencing at telomeres, mating type loci, and rDNA; maintenance of kinetochore structure; deactivation of the DNA damage checkpoint after DNA repair; and chromatin dynamics during transcription	chromatin/transcription
RLI1	G	Translation initiation factor (RNase L Inhibitor)	Essential iron-sulfur protein required for ribosome biogenesis and translation initiation and termination; facilitates binding of a multifactor complex (MFC) of initiation factors to the small ribosomal subunit; predicted ABC family ATPase	ribosome/translation
RML2	G	54S ribosomal protein RML2, mitochondrial	Mitochondrial ribosomal protein of the large subunit, has similarity to E. coli L2 ribosomal protein; fat21 mutant allele causes inability to utilize oleate and may interfere with activity of the Adr1p transcription factor	metabolism/mitochondria;ribosome/translation;lipid/sterol/fatty acid biosynth

RPD3	P	Histone deacetylase	Histone deacetylase; regulates transcription, silencing, and other processes by influencing chromatin remodeling; forms at least two different complexes which have distinct functions and members	chromatin/transcription
RPL14A	G	60S ribosomal protein L14-A	Ribosomal 60S subunit protein L14A; N-terminally acetylated; homologous to mammalian ribosomal protein L14, no bacterial homolog; RPL14A has a paralog, RPL14B, that arose from the whole genome duplication	ribosome/translation
RPL22A	G	60S ribosomal protein L22-A	Ribosomal 60S subunit protein L22A; required for the oxidative stress response in yeast; homologous to mammalian ribosomal protein L22, no bacterial homolog; RPL22A has a paralog, RPL22B, that arose from the whole genome duplication	ribosome/translation
RPL37A	G	60S ribosomal protein L37-A	Ribosomal 60S subunit protein L37A; homologous to mammalian ribosomal protein L37, no bacterial homolog; RPL37A has a paralog, RPL37B, that arose from the whole genome duplication	ribosome/translation
RPS0B	G	40S ribosomal protein S0-B	Protein component of the small (40S) ribosomal subunit; RPS0B has a paralog, RPS0A, that arose from the whole genome duplication; required for maturation of 18S rRNA along with Rps0Ap; deletion of either RPS0 gene reduces growth rate, deletion of both genes is lethal; homologous to human ribosomal protein SA and bacterial S2	ribosome/translation
RRF1	G	Ribosome Recycling Factor	Mitochondrial ribosome recycling factor, essential for mitochondrial protein synthesis and for the maintenance of the respiratory function of mitochondria	metabolism/mitochondria;ribosome/translation
RRM3	G	ATP-dependent helicase (rDNA Recombination Mutation)	DNA helicase involved in rDNA replication and Ty1 transposition; relieves replication fork pauses at telomeric regions; structurally and functionally related to Pif1p	DNA replication/repair/HR/cohesion
RSC2	G	Chromatin structure-remodeling complex subunit	Component of the RSC chromatin remodeling complex; required for expression of mid-late sporulation-specific genes; involved in telomere maintenance	chromatin/transcription
RTC2	G	Restriction of Telomere Capping	Putative vacuolar membrane transporter for cationic amino acids; likely contributes to amino acid homeostasis by exporting cationic amino acids from the vacuole; positive regulation by Lys14p suggests that lysine may be the primary substrate; member of the PQ-loop family, with seven transmembrane domains; similar to mammalian PQLC2 vacuolar transporter	unknown
RTS1	G	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform	B-type regulatory subunit of protein phosphatase 2A (PP2A); Rts1p and Cdc55p are alternative regulatory subunits for PP2A; PP2A-Rts1p protects cohesin when recruited by Sgo1p to the pericentromere; highly enriched at centromeres in absence of Cdc55p; homolog of the mammalian B' subunit of PP2A	chromosome segregation/kinetochore/spindle/microtubule
RUD3	G	GRIP domain-containing protein	Golgi matrix protein involved in the structural organization of the cis-Golgi; interacts genetically with COG3 and USO1	ER<->Golgi traffic
SCO1	G	Suppressor of Cytochrome Oxidase deficiency	Copper-binding protein of the mitochondrial inner membrane, required for cytochrome c oxidase activity and respiration; may function to deliver copper to cytochrome c oxidase; has similarity to thioredoxins	metabolism/mitochondria
SCS7	G	Ceramide very long chain fatty acid hydroxylase (Suppressor of Ca2+ Sensitivity)	Sphingolipid alpha-hydroxylase, functions in the alpha-hydroxylation of sphingolipid-associated very long chain fatty acids, has both cytochrome b5-like and hydroxylase/desaturase domains, not essential for growth	lipid/sterol/fatty acid biosynth
SHE9	G	Sensitivity to High Expression protein 9	Mitochondrial inner membrane protein required for normal mitochondrial morphology, may be involved in fission of the inner membrane; forms a homo-oligomeric complex	metabolism/mitochondria
SHR3	G	Secretory component protein	Endoplasmic reticulum packaging chaperone, required for incorporation of amino acid permeases into COPII coated vesicles for transport to the cell surface	ER<->Golgi traffic
SIP4	G	SNF1-Interacting Protein	C6 zinc cluster transcriptional activator that binds to the carbon source-responsive element (CSRE) of gluconeogenic genes; involved in the positive regulation of gluconeogenesis; regulated by Snf1p protein kinase; localized to the nucleus	metabolism/mitochondria
SIW14	G	Tyrosine-protein phosphatase (Synthetic Interaction with Whi2)	Tyrosine phosphatase that plays a role in actin filament organization and endocytosis; localized to the cytoplasm	cell polarity/morphogenesis;Golgi/endosome/vacuole/sorting;signaling/stress response
SKM1	G	STE20/PAK homologous Kinase related to Morphogenesis	Member of the PAK family of serine/threonine protein kinases with similarity to Ste20p and Cla4p; involved in down-regulation of sterol uptake; proposed to be a downstream effector of Cdc42p during polarized growth	

SNO1	G	Probable glutamine amidotransferase	Protein of unconfirmed function, involved in pyridoxine metabolism; expression is induced during stationary phase; forms a putative glutamine amidotransferase complex with Snz1p, with Sno1p serving as the glutaminase	metabolism/ mitochondria
SPS4	G	Sporulation Specific protein 4	Protein whose expression is induced during sporulation; not required for sporulation; heterologous expression in E. coli induces the SOS response that senses DNA damage	G1/S and G2/M cell cycle progression/ meiosis
SPT16	G	FACT complex subunit	Subunit of the heterodimeric FACT complex (Spt16p-Pob3p), which associates with chromatin via interaction with Nhp6Ap and Nhp6Bp, and reorganizes nucleosomes to facilitate access to DNA by RNA and DNA polymerases	chromatin/tra nscription
SPT21	G	Protein SPT21	Protein with a role in transcriptional silencing; required for normal transcription at several loci including HTA2-HTB2 and HHF2-HHT2, but not required at the other histone loci; functionally related to Spt10p	chromatin/tra nscription
STB1	G	Sin Three Binding protein	Protein with a role in regulation of MBF-specific transcription at Start, phosphorylated by Cln-Cdc28p kinases in vitro; unphosphorylated form binds Swi6p and binding is required for Stb1p function; expression is cell-cycle regulated	G1/S and G2/M cell cycle progression/ meiosis;chrom atin/transcript ion
STE50	G	Protein STE50	Protein involved in mating response, invasive/filamentous growth, and osmotolerance, acts as an adaptor that links G protein-associated Cdc42p-Ste20p complex to the effector Ste11p to modulate signal transduction	cell polarity/morp hogenesis;sign aling/stress response
SUV3	G	ATP-dependent RNA helicase SUV3, mitochondrial	ATP-dependent RNA helicase, component of the mitochondrial degradosome along with the RNase Dss1p; the degradosome associates with the ribosome and mediates RNA turnover; also required during splicing of the COX1 AI5_beta intron	metabolism/ mitochondria;r ibosome/trans lation;RNA processing
SWD3	G	Set1c, WD40 repeat protein	Essential subunit of the COMPASS (Set1C) complex, which methylates histone H3 on lysine 4 and is required in transcriptional silencing near telomeres; WD40 beta propeller superfamily member and ortholog of mammalian WDR5	chromatin/tra nscription
SWI4	P	Regulatory protein SWI4	DNA binding component of the SBF complex (Swi4p-Swi6p), a transcriptional activator that in concert with MBF (Mbp1-Swi6p) regulates late G1-specific transcription of targets including cyclins and genes required for DNA synthesis and repair	G1/S and G2/M cell cycle progression/ meiosis;chrom atin/transcript ion
SWI6	G/P	Regulatory protein SWI6	Transcription cofactor; forms complexes with Swi4p and Mbp1p to regulate transcription at the G1/S transition; involved in meiotic gene expression; also binds Stb1p to regulate transcription at START; cell wall stress induces phosphorylation by Mpk1p, which regulates Swi6p localization; required for the unfolded protein response, independently of its known transcriptional coactivators	G1/S and G2/M cell cycle progression/ meiosis;chrom atin/transcript ion
TAF1	G	TATA binding protein-Associated Factor	TFIID subunit (145 kDa), involved in RNA polymerase II transcription initiation; possesses in vitro histone acetyltransferase activity but its role in vivo appears to be minor; involved in promoter binding and G1/S progression	chromatin/tra nscription
TAF8	G	TATA binding protein-Associated Factor	TFIID subunit (65 kDa), involved in RNA polymerase II transcription initiation	chromatin/tra nscription
TEL1	G	Serine/threonine-protein kinase (Telomere maintenance)	Protein kinase primarily involved in telomere length regulation; contributes to cell cycle checkpoint control in response to DNA damage; functionally redundant with Mec1p; regulates P-body formation induced by replication stress; homolog of human ataxia-telangiectasia mutated (ATM) gene, the gene responsible for ataxia telangiectasia (AT) (OMIM 607585)	DNA replication/re pair/HR/cohe sion
TPK1	P	Takashi's Protein Kinase	cAMP-dependent protein kinase catalytic subunit; promotes vegetative growth in response to nutrients via the Ras-cAMP signaling pathway; inhibited by regulatory subunit Bcy1p in the absence of cAMP; partially redundant with Tpk2p and Tpk3p	signaling/stre ss response
TPS1	G	Trehalose-6-Phosphate Synthase	Synthase subunit of trehalose-6-P synthase/phosphatase complex; synthesizes the storage carbohydrate trehalose; also found in a monomeric form; expression is induced by the stress response and repressed by the Ras-cAMP pathway; protein abundance increases in response to DNA replication stress	metabolism/ mitochondria

UPS3	G	Protein UPS3, mitochondrial	Mitochondrial protein of unknown function; similar to Ups1p and Ups2p which are involved in regulation of mitochondrial cardiolipin and phosphatidylethanolamine levels; null is viable but interacts synthetically with ups1 and ups2 mutations	metabolism/ mitochondria
VAC17	G	Vacuole-related protein 17	Phosphoprotein involved in vacuole inheritance; degraded in late M phase of the cell cycle; acts as a vacuole-specific receptor for myosin Myo2p	Golgi/endosome/vacuole/sorting
VIP1	G	Inositol hexakisphosphate and diphosphoinositol-pentakisphosphate kinase	Inositol hexakisphosphate (IP6) and inositol heptakisphosphate (IP7) kinase; IP7 production is important for phosphate signaling; involved in cortical actin cytoskeleton function, and invasive pseudohyphal growth analogous to S. pombe asp1	Golgi/endosome/vacuole/sorting
VMS1	G	VCP/Cdc48-associated Mitochondrial Stress-responsive	Component of a Cdc48p-complex involved in protein quality control; exhibits cytosolic and ER-membrane localization, with Cdc48p, during normal growth, and contributes to ER-associated degradation (ERAD) of specific substrates at a step after their ubiquitination; forms a mitochondrially-associated complex with Cdc48p and Npl4p under oxidative stress that is required for ubiquitin-mediated mitochondria-associated protein degradation (MAD); conserved in C. elegans and humans	unknown
VPS41	G	Vacuolar protein sorting-associated protein 41	Vacuolar membrane protein that is a subunit of the homotypic vacuole fusion and vacuole protein sorting (HOPS) complex; essential for membrane docking and fusion at the Golgi-to-endosome and endosome-to-vacuole stages of protein transport	Golgi/endosome/vacuole/sorting
YCK1	P	Yeast Casein Kinase I homologue	Palmitoylated plasma membrane-bound casein kinase I isoform; shares redundant functions with Yck2p in morphogenesis, proper septin assembly, endocytic trafficking; provides an essential function overlapping with that of Yck2p	cell polarity/morphogenesis
YND1	G	Yeast Nucleoside Diphosphatase	Apyrase with wide substrate specificity, helps prevent inhibition of glycosylation by hydrolyzing nucleoside tri- and diphosphates that inhibit glycotransferases; partially redundant with Gda1p; mediates adenovirus E4orf4-induced toxicity	protein folding/protein glycosylation/ cell wall biogenesis&integrity
YPT31	G	GTP-binding protein YPT31/YPT8	Rab family GTPase, very similar to Ypt32p; involved in the exocytic pathway; mediates intra-Golgi traffic or the budding of post-Golgi vesicles from the trans-Golgi	cell polarity/morphogenesis
ZDS1	G	Protein ZDS1	Protein with a role in regulating Swe1p-dependent polarized growth; involved in maintaining Cdc55p in the cytoplasm where it promotes mitotic entry; involved in mitotic exit through Cdc14p regulation; interacts with silencing proteins at the telomere; has a role in Bcy1p localization; implicated in mRNA nuclear export	G1/S and G2/M cell cycle progression/ meiosis;chromosome segregation/kinetochore/spindle/microtubule
ZIP1	G	Synaptonemal complex protein ZIP1	Transverse filament protein of the synaptonemal complex; required for normal levels of meiotic recombination and pairing between homologous chromosome during meiosis; potential Cdc28p substrate	G1/S and G2/M cell cycle progression/ meiosis;chromosome segregation/kinetochore/spindle/microtubule
YBR235W /VHC1	G	Vacuolar protein Homologous to CCC family	Vacuolar membrane cationÂ¿chloride cotransporter (CCC); likely mediates K+ and Cl- cotransport into the vacuole; has a role in potassium homeostasis and salt tolerance; similar to mammalian electroneutral Na(+)-(K+)-C1- cotransporter family	unknown
YGR125W	G	Uncharacterized vacuolar membrane protein	Putative protein of unknown function; deletion mutant has decreased rapamycin resistance but normal wormannin resistance; green fluorescent protein (GFP)-fusion protein localizes to the vacuole	unknown
YJL160C	G	Cell wall protein PIR5	Putative protein of unknown function; member of the PIR (proteins with internal repeats) family of cell wall proteins; non-essential gene that is required for sporulation; mRNA is weakly cell cycle regulated, peaking in mitosis	unknown
YJR011C	G	Uncharacterized protein	Putative protein of unknown function; GFP-fusion protein expression is induced in response to the DNA-damaging agent MMS	unknown

YKL069W	G	Free methionine-R-sulfoxide reductase	Methionine-R-sulfoxide reductase, reduces the R enantiomer of free Met-SO, in contrast to Ycl033Cp which reduces Met-R-SO in a peptide linkage; has a role in protection against oxidative stress	nuclear-cytoplasic transport
YLR407W	G	Uncharacterized protein	Putative protein of unknown function; null mutant displays elongated buds and a large fraction of budded cells have only one nucleus	unknown
YMR160W	G	Uncharacterized protein	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes to the membrane of the vacuole; mutant has enhanced sensitivity to overexpression of mutant huntingtin; YMR160W is not an essential gene	unknown
YMR291W /TDA1	G	Serine/threonine-protein kinase TDA1	Protein kinase of unknown cellular role; green fluorescent protein (GFP)-fusion protein localizes to the cytoplasm and nucleus; null mutant is sensitive to expression of the top1-T722A allele; not an essential gene	unknown
YPL109C	G	ABC1 family protein YPL109C, mitochondrial	Putative protein of unknown function; the authentic, non-tagged protein is detected in highly purified mitochondria in high-throughput studies	unknown

Table 8: List of homologs of physical interactors of Whi5

Physical Interactors	NAME	Homologs in human (Results from Homologene)	Homologs in human (Results from Biomart)	Homologs in human (Results from BLAST)	Homolog s from pRb Network
ATG1	Serine/threonine-protein kinase ATG1	ULK1, ULK2	ULK3 ULK1 ULK2 STK33 STK35	Isoform of Serine/threonine-protein kinase, ULK1, ULK2, ULK3, 31-39%	Not Present in pRb interactors pRb inactivation contributes to autophagy directly, by deregulating expression of autophagy genes LC3, ATG1, DRAM.
CDC28	Cyclin-dependent kinase 1	CDK1, CDK2	CDK3	Cyclin-dependent kinase, CDK1, CDK2, CDK3, CDK5, 55-64%	Present in pRb ineractors
ESA1	Histone acetyltransferase ESA1	KAT5, MYST association factor 6 Not Present in pRb ineractors	KAT5 KAT8 KAT6A, B KAT7	Isoform 2 of Histone acetyltransferase, KAT5, 46%	Not Present in pRb ineractors KAT2B is Present in pRb ineractors
HOS1	Histone deacetylase HOS1	No	HDAC11	Histone deacetylase, HDAC4, HDAC5, HDAC5, 25%	HDAC is doing similar function.
HOS3	Histone deacetylase HOS3	No		Histone deacetylase, HDAC1, HDAC2, 38%	HDAC
HSP82	ATP-dependent molecular chaperone HSP82	HSP90 Not Present in pRb ineractors	HSP90AB1 HSP90AA1	Heat shock protein, HSP90AB1, 62%	HSPA8 or HSP73 hsc73 may act exclusively as the molecular chaperone for nonphosphorylated pRb. As a result, hsc73 may function as a molecular stabilizer of nonphosphorylated pRb.
MSN5	Protein MSN5	No	XPO5	Exportin-5, XPO5, 21%	Not Present in pRb ineractors
PCL9	PHO85 cyclin-9	No		Cyclin-Y-like protein 2, CCNYL2, 235	Not Present in pRb ineractors
PHO85	Cyclin-dependent protein kinase PHO85	CDK5	CDK16 CDK16 CDK18 CDK14 CDK17	Cyclin-dependent kinase, CDK2, CDK3, CDK5, 54-56%	Present in pRb ineractors p25-Cdk5 complex phosphorylates Rb directly without the need for any intermediary kinase activity.
PKP2	Pyruvate dehydrogenase kinase 2, mitochondrial	PKP2=plakophilin 2 ??	PDK2 BCKDK PDK1 PDK3 PDK4	Pyruvate dehydrogenase Kinase, PDK4, PDK1, 23%	Not Present in pRb ineractors May be PDK family E2F1 induces PDK4 transcription and blunts glucose oxidation (Hsieh et al, 2008)
PTK2	Serine/threonine-protein kinase PTK2/STK2	PTK2		Isoform 16 of Serine/threonine-protein kinase, MARK2, 23%	Not Present in pRb ineractors

RAD53	Serine/threonine-protein kinase RAD53	CHEK2	PHKG1 PHKG2	Isoform 9 of Serine/threonine-protein kinase Chk2, CHEK2, 33%	Present in pRb ineractors
RPD3	Histone deacetylase RPD3	HDAC1,2,3,8	HDAC2	Histone deacetylase, HDAC1, HDAC2, HDAC3, HDAC8, 43-63%	Present in pRb ineractors
SWI4	Regulatory protein SWI4	No		Dentin sialophosphoprotein, DSPP, 19%	Not Present in pRb ineractors E2F family doing same function
SWI6	Regulatory protein SWI6	No		Isoform 6 of A-kinase anchor protein 9, AKAP9, 20%	Not Present in pRb ineractors E2F family doing same function
TPK1	cAMP-dependent protein kinase type 1	TPK1	PRKACA PRKACG PRKACB	cAMP-dependent protein kinase catalytic subunit beta, PRKACB, 49%	Not Present in pRb ineractors
YCK1	Casein kinase I homolog 1	CSNK1G3	VRK1 VRK2 VRK3	Casein kinase I isoform gamma, CSNK1G1, CSNK1G2, CSNK1G3, 47-55%	Not Present in pRb ineractors

Table 9: Whi5 phosphorylation sites

Kinase presented in this table interact with Whi5 (See Table 8). Phosphorylation sites are predicted by GPS2.1. As GPS 2.1 include almost human only protein kinases, the phosphorylation sites are predicted based human homolog of yeast kinases.

Experimental Phosphosites

Kinase (homolog in human)	**No of sites**	**Sites+**	**Experimental/Computational**
CDK1	12	T5, T47, T57, S59, S62, S88, T143, S154, S156, S161, T215, S262	Experimental
Non CDK1	6	S113, S115, S149, S276, T281, S288	Experimental
??	7	T67, S78, T79, T80, T164, T284, T290	Experimental (Large scale)

Computational Phosphosites

Kinase (homolog in human)	**No of sites**	**Sites+**	**Experimental/Computational**
ATG1		Not found in GPS2.1	
CDC28 (CDC2/CDK2)	11	**T5, T47, T57, S59, S62, T143, S154, S156, S161, T215, S262**	Computational (High threshold)
PHO85 (CDK5)	6	**T5, S59, S62, T143, S156, T215**	Computational (High threshold)
PKP2 (PDK1)	6	T43**, T181,** T201, T232, T275, **T290**	Computational (Medium threshold)
PTK2		Not found in GPS2.1	
RAD53 (RAD53/Chek2)	5	T47, **S78**, **S149,** S169, T170	Computational (High threshold)
TPK1 (PKA)	6	T11, S12, **S78, T79**, S169, T170	Computational (High threshold)
YCK1 (CK1)	12	S2, S9, S12, **S113, S115, S276**, **T281**, **T284**, **S288, T290**, T294, S295	Computational (High threshold)

+Bold sites are experimental phosphorylation sites

Whi5 phosphorylation Sites predicted By NetPhosYeast 1.0 Server

S2, **T5**, S9, T11, S12, **T47, T57, S59, S62**, S69, **S78, T79**, S81, **S88, S113, S115, T143,** S146, **S149, S154, S156, S161,** S169, T170, S179, **T215,** S227, S233, **S262,** T275, **S276, T281, T284, S288, T290,** S295. **Total 36 sites (22 experimental)**

Table 10: Different level of Whi5 interactors

1st Level	2nd Level	3rd Level
ATG1	ENT5	
	SPT16	ASK10, HSL1, RIM1, IME2
	VIP1	RPS0B
	ELM1	
CDC28	BCK2	
	CDH1	HSL1, IME2
	CLN1	
	CLN2	
	CLN3	
	KEL1	
	LTE1	
	PTP3	
	RLF2	
	RTS1	
	STB1	
	VAC17	
	YMR291W	
	ZDS1	
	ZIP1	
ESA1	EAF1	
	HTA1	ADA2, RIM1,
	RAP1	CCR4, TAF1, RPL37A
	SPT16	ASK10, HSL1, RIM1, IME2
HOS1		
HOS3		
HSP82	CPR7	
	HXT17	
	RAD52	RIM1
	RKM3	
	RPL22A	
	VAC17	
	YPT31	
MSN5	CDH1	HSL1, IME2
PCL9	QCR2	AIM31
	RSC2	
PHO85	APA1	
	CAJ1	
	CLN1	
	CSM3	VPS41
	MBF1	
	MET12	
	PCL1	
	PTC1	HSL1, IME2
	RGI2	HSL1
	RIM15	TPS1
	VIP1	RPS0B
PKP2		
PTK2	ENT5	
	KEL1	
	NPT1	
	VIP1	RPS0B
	YMR291W	
RAD53	DUN1	TEL1
	RAD55	
RPD3	CKB2	
	CPR7	
SWI4	BCK2	
	POB3	
	SPT16	ASK10, HSL1, RIM1, IME2
SWI6	CDH1	HSL1, IME2

	CLN3	
	POB3	
	SKM1	
	SPT16	ASK10, HSL1, RIM1, IME2
	STB1	
	TAF8	TAF1
TPK1	ENT5	
	MET12	
	RIM15	TPS1
	RUD3	
	SPT16	ASK10, HSL1, RIM1, IME2
	VIP1	RPS0B
YCK1	CDH1	HSL1
	DUN1	TEL1
	ELM1	
	ENT5	
	PPH21	
	PYC2	
	RLI1	
	SPT16	ASK10, HSL1, RIM1, IME2
	STE50	
	VIP1	RPS0B

Table 11: List of post-translational modification (PTM) sites of Whi5 and pRb

Protein	**Whi5**	**pRb**			
PTM Type	Cdk Phosphorylation	Cdk Phosphorylation	Methylation	Acetylation	SUMOylation
1	T5	S230	K810	K873	K720
2	T47	S249	K860	K874	
3	T57	T252	K873		
4	S59	T356			
5	S62	T373			
6	S88	S567			
7	T143	S608			
8	S154	S612			
9	S156	S780			
10	S161	S788			
11	T215	S795			
12	S262	S807			
13		S811			
14		T821			
15		T826			

Table 12: GO term enrichment of Whi5 interactors

Biological Process

GO id	GO name	adjusted-P
GO:0051726	regulation of cell cycle	5.42E-11
GO:0006468	protein phosphorylation	1.46E-09
GO:0051325	interphase	4.71E-08
GO:0006464	protein modification process	4.93E-08
GO:0007049	cell cycle	5.33E-08
GO:0048519	negative regulation of biological process	1.04E-07
GO:0022402	cell cycle process	1.13E-07
GO:0060255	regulation of macromolecule metabolic process	1.65E-07
GO:0000083	regulation of transcription involved in G1/S phase of mitotic cell cycle	2.67E-07
GO:0006793	phosphorus metabolic process	2.95E-07
GO:0006796	phosphate metabolic process	2.95E-07
GO:0051329	interphase of mitotic cell cycle	3.59E-07
GO:0000082	G1/S transition of mitotic cell cycle	6.69E-07
GO:0048523	negative regulation of cellular process	1.17E-06
GO:0022403	cell cycle phase	1.32E-06
GO:0010564	regulation of cell cycle process	1.34E-06
GO:0050789	regulation of biological process	1.48E-06
GO:0016310	phosphorylation	2.36E-06
GO:0019222	regulation of metabolic process	3.11E-06
GO:0050794	regulation of cellular process	3.19E-06
GO:0065007	biological regulation	4.03E-06
GO:0043412	macromolecule modification	9.65E-06
GO:0045786	negative regulation of cell cycle	1.96E-05
GO:0080090	regulation of primary metabolic process	2.12E-05
GO:0031323	regulation of cellular metabolic process	2.59E-05
GO:0007050	cell cycle arrest	1.47E-04
GO:0000278	mitotic cell cycle	1.91E-04
GO:2000112	regulation of cellular macromolecule biosynthetic process	2.74E-04
GO:0010556	regulation of macromolecule biosynthetic process	2.94E-04
GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	3.08E-04
GO:0051171	regulation of nitrogen compound metabolic process	3.42E-04
GO:0006355	regulation of transcription, DNA-dependent	4.56E-04
GO:0051276	chromosome organization	4.90E-04
GO:0006366	transcription from RNA polymerase II promoter	5.64E-04
GO:0006950	response to stress	5.99E-04
GO:0009889	regulation of biosynthetic process	6.10E-04
GO:0031326	regulation of cellular biosynthetic process	6.10E-04
GO:0000075	cell cycle checkpoint	6.12E-04
GO:0071156	regulation of cell cycle arrest	6.12E-04
GO:0051252	regulation of RNA metabolic process	8.74E-04
GO:0009893	positive regulation of metabolic process	0.001048
GO:0031399	regulation of protein modification process	0.001156
GO:0006325	chromatin organization	0.001177
GO:0006261	DNA-dependent DNA replication	0.001291
GO:0071900	regulation of protein serine/threonine kinase activity	0.001378
GO:0001932	regulation of protein phosphorylation	0.001728
GO:0010468	regulation of gene expression	0.001819
GO:0006357	regulation of transcription from RNA polymerase II promoter	0.002118
GO:0019220	regulation of phosphate metabolic process	0.002528
GO:0051174	regulation of phosphorus metabolic process	0.002528
GO:0042325	regulation of phosphorylation	0.003133
GO:0031325	positive regulation of cellular metabolic process	0.003248
GO:0006260	DNA replication	0.003514
GO:0048518	positive regulation of biological process	0.003831
GO:0006351	transcription, DNA-dependent	0.003849
GO:0032774	RNA biosynthetic process	0.003973
GO:0007346	regulation of mitotic cell cycle	0.004494
GO:0045859	regulation of protein kinase activity	0.005434
GO:0033554	cellular response to stress	0.006262
GO:0050896	response to stimulus	0.008357
GO:0043549	regulation of kinase activity	0.008581
GO:0006259	DNA metabolic process	0.008619
GO:0048522	positive regulation of cellular process	0.009554
GO:0045934	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0.01043

GO:0051172	negative regulation of nitrogen compound metabolic process	0.01043
GO:0051338	regulation of transferase activity	0.011437
GO:0010605	negative regulation of macromolecule metabolic process	0.01749
GO:0010558	negative regulation of macromolecule biosynthetic process	0.017832
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	0.017832
GO:0043170	macromolecule metabolic process	0.024393
GO:0006348	chromatin silencing at telomere	0.025776
GO:0010604	positive regulation of macromolecule metabolic process	0.02843
GO:0044260	cellular macromolecule metabolic process	0.028798
GO:0009890	negative regulation of biosynthetic process	0.041866
GO:0031327	negative regulation of cellular biosynthetic process	0.041866
GO:0090329	regulation of DNA-dependent DNA replication	0.041976
GO:0006974	response to DNA damage stimulus	0.047288
GO:0010629	negative regulation of gene expression	0.053839
GO:0009892	negative regulation of metabolic process	0.061735
GO:0006996	organelle organization	0.062529
GO:0034401	regulation of transcription by chromatin organization	0.065235
GO:0051321	meiotic cell cycle	0.070456
GO:0051716	cellular response to stimulus	0.08005
GO:0006342	chromatin silencing	0.086266
GO:0040029	regulation of gene expression, epigenetic	0.086266
GO:0045814	negative regulation of gene expression, epigenetic	0.086266
GO:0044267	cellular protein metabolic process	0.103682
GO:0016568	chromatin modification	0.110313
GO:0016458	gene silencing	0.114555
GO:0006275	regulation of DNA replication	0.114594
GO:0000183	chromatin silencing at rDNA	0.115955
GO:0000279	M phase	0.125522
GO:0045892	negative regulation of transcription, DNA-dependent	0.142866
GO:0051253	negative regulation of RNA metabolic process	0.149992
GO:0007093	mitotic cell cycle checkpoint	0.151097
GO:0031324	negative regulation of cellular metabolic process	0.151956
GO:0032878	regulation of establishment or maintenance of cell polarity	0.192311
GO:0016569	covalent chromatin modification	0.213116
GO:0016570	histone modification	0.213116
GO:0007126	meiosis	0.228095
GO:0051327	M phase of meiotic cell cycle	0.228095
GO:0090304	nucleic acid metabolic process	0.290501
GO:2000602	regulation of interphase of mitotic cell cycle	0.349616
GO:0019538	protein metabolic process	0.358501
GO:0031331	positive regulation of cellular catabolic process	0.364043
GO:0048478	replication fork protection	0.37453
GO:0045935	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0.377309
GO:0051173	positive regulation of nitrogen compound metabolic process	0.394074
GO:0071824	protein-DNA complex subunit organization	0.415412
GO:0009891	positive regulation of biosynthetic process	0.459444
GO:0031328	positive regulation of cellular biosynthetic process	0.459444
GO:0016043	cellular component organization	0.482921
GO:0065004	protein-DNA complex assembly	0.48555
GO:0000079	regulation of cyclin-dependent protein kinase activity	0.505837
GO:0010948	negative regulation of cell cycle process	0.557776
GO:0006281	DNA repair	0.733641
GO:0051052	regulation of DNA metabolic process	0.745754
GO:0090068	positive regulation of cell cycle process	0.808535
GO:0009987	cellular process	0.836662
GO:0071842	cellular component organization at cellular level	0.865325
GO:0070933	histone H4 deacetylation	0.877091
GO:0071511	inactivation of MAPK activity involved in conjugation with cellular fusion	0.877091
GO:0016311	dephosphorylation	0.877253
GO:0065009	regulation of molecular function	0.908991
GO:0051254	positive regulation of RNA metabolic process	0.916311
GO:0000188	inactivation of MAPK activity	0.993544
GO:0043407	negative regulation of MAP kinase activity	0.993544
GO:0051302	regulation of cell division	0.993544
GO:0071901	negative regulation of protein serine/threonine kinase activity	0.993544
GO:2000104	negative regulation of DNA-dependent DNA replication	0.993544

Cellular Component

GO id	GO name	adjusted-P
GO:0005634	nucleus	0.012589
GO:0000307	cyclin-dependent protein kinase holoenzyme complex	0.04808

GO:0033309	SBF transcription complex	0.079208
GO:0005694	chromosome	0.094417
GO:0043226	organelle	0.138068
GO:0043229	intracellular organelle	0.138068
GO:0043234	protein complex	0.148843
GO:0044427	chromosomal part	0.179654
GO:0043227	membrane-bounded organelle	0.253857
GO:0043231	intracellular membrane-bounded organelle	0.253857
GO:0044454	nuclear chromosome part	0.411481
GO:0044424	intracellular part	0.458662
GO:0000228	nuclear chromosome	0.478186
GO:0005622	intracellular	0.626387
GO:0035101	FACT complex	0.877091

Molecular Function

GO id	GO name	adjusted-P
GO:0004672	protein kinase activity	7.79E-05
GO:0004674	protein serine/threonine kinase activity	2.74E-04
GO:0016773	phosphotransferase activity, alcohol group as acceptor	0.011603
GO:0016301	kinase activity	0.01315
GO:0019887	protein kinase regulator activity	0.114594
GO:0019207	kinase regulator activity	0.151097
GO:0034739	histone deacetylase activity (H3-K16 specific)	0.192311
GO:0004721	phosphoprotein phosphatase activity	0.28635
GO:0016772	transferase activity, transferring phosphorus-containing groups	0.339772
GO:0031078	histone deacetylase activity (H3-K14 specific)	0.37453
GO:0032041	NAD-dependent histone deacetylase activity (H3-K14 specific)	0.37453
GO:0032129	histone deacetylase activity (H3-K9 specific)	0.37453
GO:0046969	NAD-dependent histone deacetylase activity (H3-K9 specific)	0.37453
GO:0046970	NAD-dependent histone deacetylase activity (H4-K16 specific)	0.37453
GO:0016538	cyclin-dependent protein kinase regulator activity	0.422036
GO:0004407	histone deacetylase activity	0.482921
GO:0033558	protein deacetylase activity	0.482921
GO:0005524	ATP binding	0.489247
GO:0032559	adenyl ribonucleotide binding	0.514536
GO:0030554	adenyl nucleotide binding	0.569266
GO:0017136	NAD-dependent histone deacetylase activity	0.993544
GO:0034979	NAD-dependent protein deacetylase activity	0.993544

Table 13: GO term enrichment of Rb interactors

Biological Process

GO id	GO name	adjusted -P
GO:0060255	regulation of macromolecule metabolic process	6.99E-96
GO:0019222	regulation of metabolic process	4.86E-93
GO:0080090	regulation of primary metabolic process	1.27E-89
GO:0006351	transcription, DNA-dependent	6.96E-89
GO:0031323	regulation of cellular metabolic process	8.67E-89
GO:0032774	RNA biosynthetic process	7.72E-88
GO:2000112	regulation of cellular macromolecule biosynthetic process	3.47E-87
GO:0010468	regulation of gene expression	7.26E-87
GO:0006355	regulation of transcription, DNA-dependent	2.53E-86
GO:0010556	regulation of macromolecule biosynthetic process	3.68E-86
GO:0010467	gene expression	8.68E-85
GO:0051252	regulation of RNA metabolic process	8.95E-85
GO:0044260	cellular macromolecule metabolic process	1.09E-84
GO:0031326	regulation of cellular biosynthetic process	1.60E-84
GO:0009889	regulation of biosynthetic process	4.75E-84
GO:0043170	macromolecule metabolic process	1.80E-82
GO:0051171	regulation of nitrogen compound metabolic process	2.46E-82
GO:0034645	cellular macromolecule biosynthetic process	2.47E-81
GO:0016070	RNA metabolic process	4.74E-81
GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	7.55E-81
GO:0009059	macromolecule biosynthetic process	5.39E-80
GO:0090304	nucleic acid metabolic process	1.32E-78
GO:0006366	transcription from RNA polymerase II promoter	6.17E-78
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1.72E-72
GO:0050794	regulation of cellular process	2.19E-72
GO:0048523	negative regulation of cellular process	8.10E-71
GO:0050789	regulation of biological process	8.48E-69
GO:0044249	cellular biosynthetic process	9.14E-69
GO:0006357	regulation of transcription from RNA polymerase II promoter	2.34E-68
GO:0044238	primary metabolic process	5.00E-68
GO:0007049	cell cycle	2.56E-67
GO:0009058	biosynthetic process	6.33E-67
GO:0034641	cellular nitrogen compound metabolic process	1.98E-66
GO:0048522	positive regulation of cellular process	2.84E-66
GO:0048519	negative regulation of biological process	1.07E-65
GO:0044237	cellular metabolic process	1.41E-65
GO:0010604	positive regulation of macromolecule metabolic process	1.05E-64
GO:0006807	nitrogen compound metabolic process	1.72E-64
GO:0065007	biological regulation	2.13E-64
GO:0031325	positive regulation of cellular metabolic process	2.45E-63
GO:0009893	positive regulation of metabolic process	4.28E-63
GO:0010605	negative regulation of macromolecule metabolic process	7.54E-63
GO:0008152	metabolic process	1.78E-62
GO:0009892	negative regulation of metabolic process	8.31E-61
GO:0048518	positive regulation of biological process	1.09E-60
GO:0031324	negative regulation of cellular metabolic process	4.04E-60
GO:0051726	regulation of cell cycle	2.68E-59
GO:0010558	negative regulation of macromolecule biosynthetic process	2.78E-57
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	6.01E-57
GO:0009890	negative regulation of biosynthetic process	6.51E-57
GO:0010629	negative regulation of gene expression	1.09E-56
GO:0010628	positive regulation of gene expression	1.09E-56
GO:0045892	negative regulation of transcription, DNA-dependent	3.22E-56
GO:0031327	negative regulation of cellular biosynthetic process	4.15E-56
GO:0051253	negative regulation of RNA metabolic process	5.70E-55
GO:0045934	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	4.55E-54
GO:0051172	negative regulation of nitrogen compound metabolic process	8.89E-54
GO:0045893	positive regulation of transcription, DNA-dependent	4.68E-53
GO:0051254	positive regulation of RNA metabolic process	5.60E-52
GO:0045935	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	6.39E-52
GO:0010557	positive regulation of macromolecule biosynthetic process	8.71E-52
GO:0000278	mitotic cell cycle	1.22E-51
GO:0051173	positive regulation of nitrogen compound metabolic process	2.38E-51

GO:0031328	positive regulation of cellular biosynthetic process	9.40E-50
GO:0009891	positive regulation of biosynthetic process	2.50E-49
GO:0051276	chromosome organization	4.93E-48
GO:0016568	chromatin modification	1.46E-47
GO:0022402	cell cycle process	3.58E-46
GO:0006325	chromatin organization	3.28E-45
GO:0008283	cell proliferation	1.34E-44
GO:0043412	macromolecule modification	2.02E-43
GO:0051329	interphase of mitotic cell cycle	2.13E-43
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	2.53E-43
GO:0006464	protein modification process	3.03E-43
GO:0051325	interphase	4.68E-43
GO:0009987	cellular process	5.90E-40
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	3.14E-39
GO:0006996	organelle organization	5.93E-39
GO:0022403	cell cycle phase	8.75E-39
GO:0033554	cellular response to stress	8.93E-38
GO:0045786	negative regulation of cell cycle	9.74E-38
GO:0071840	cellular component organization or biogenesis	2.25E-37
GO:0016043	cellular component organization	9.38E-37
GO:0006950	response to stress	1.82E-36
GO:0051716	cellular response to stimulus	9.30E-36
GO:0042981	regulation of apoptosis	1.22E-35
GO:0044267	cellular protein metabolic process	1.42E-35
GO:0043067	regulation of programmed cell death	2.43E-35
GO:0006915	apoptosis	2.84E-35
GO:0012501	programmed cell death	6.98E-35
GO:0010941	regulation of cell death	1.71E-34
GO:0019538	protein metabolic process	6.31E-34
GO:0032502	developmental process	1.80E-33
GO:0016569	covalent chromatin modification	2.03E-33
GO:0007050	cell cycle arrest	2.33E-33
GO:0008219	cell death	2.93E-33
GO:0016265	death	3.09E-33
GO:0065009	regulation of molecular function	6.39E-33
GO:0006974	response to DNA damage stimulus	1.06E-32
GO:0042127	regulation of cell proliferation	2.44E-32
GO:0016570	histone modification	2.93E-32
GO:0010564	regulation of cell cycle process	7.08E-32
GO:0071841	cellular component organization or biogenesis at cellular level	2.27E-31
GO:0050896	response to stimulus	5.10E-31
GO:0071842	cellular component organization at cellular level	1.05E-30
GO:0008150	biological_process	1.21E-30
GO:0071156	regulation of cell cycle arrest	5.42E-30
GO:0051246	regulation of protein metabolic process	5.86E-30
GO:0032501	multicellular organismal process	1.74E-29
GO:0022414	reproductive process	1.14E-28
GO:0022415	viral reproductive process	1.32E-28
GO:0000003	reproduction	1.37E-28
GO:0007275	multicellular organismal development	5.20E-28
GO:0048513	organ development	1.39E-27
GO:0000082	G1/S transition of mitotic cell cycle	2.80E-27
GO:0044403	symbiosis, encompassing mutualism through parasitism	3.80E-27
GO:0044419	interspecies interaction between organisms	3.80E-27
GO:0048856	anatomical structure development	5.93E-27
GO:0032268	regulation of cellular protein metabolic process	7.91E-27
GO:0031399	regulation of protein modification process	3.14E-26
GO:0045595	regulation of cell differentiation	3.54E-26
GO:0016032	viral reproduction	5.07E-26
GO:0019048	virus-host interaction	9.99E-26
GO:0030154	cell differentiation	1.02E-25
GO:0048731	system development	1.26E-25
GO:0007346	regulation of mitotic cell cycle	1.31E-25
GO:0050793	regulation of developmental process	4.85E-25
GO:0006259	DNA metabolic process	8.67E-25
GO:0000080	G1 phase of mitotic cell cycle	9.35E-25
GO:0051318	G1 phase	2.20E-24
GO:0050790	regulation of catalytic activity	4.85E-24
GO:0048869	cellular developmental process	5.24E-24
GO:0042221	response to chemical stimulus	5.43E-24
GO:0051094	positive regulation of developmental process	5.78E-24
GO:0051701	interaction with host	1.31E-23
GO:0006367	transcription initiation from RNA polymerase II promoter	2.26E-23

GO:0000075	cell cycle checkpoint	3.71E-23
GO:0006352	transcription initiation, DNA-dependent	6.34E-23
GO:0044093	positive regulation of molecular function	1.21E-22
GO:0051704	multi-organism process	4.93E-22
GO:0000077	DNA damage checkpoint	1.35E-21
GO:0031570	DNA integrity checkpoint	3.71E-21
GO:0051301	cell division	4.91E-21
GO:0007167	enzyme linked receptor protein signaling pathway	4.92E-21
GO:0010033	response to organic substance	5.09E-21
GO:0065008	regulation of biological quality	5.63E-21
GO:0006793	phosphorus metabolic process	5.88E-21
GO:0006796	phosphate metabolic process	5.88E-21
GO:0030099	myeloid cell differentiation	8.39E-21
GO:0016310	phosphorylation	1.17E-20
GO:0007165	signal transduction	1.20E-20
GO:0070887	cellular response to chemical stimulus	2.57E-20
GO:0023052	signaling	8.12E-20
GO:0006468	protein phosphorylation	1.10E-19
GO:0043068	positive regulation of programmed cell death	2.27E-19
GO:0010942	positive regulation of cell death	5.25E-19
GO:0051239	regulation of multicellular organismal process	1.53E-18
GO:0042770	signal transduction in response to DNA damage	1.67E-18
GO:0043065	positive regulation of apoptosis	1.87E-18
GO:2000602	regulation of interphase of mitotic cell cycle	1.87E-18
GO:2000026	regulation of multicellular organismal development	2.63E-18
GO:0007093	mitotic cell cycle checkpoint	1.13E-17
GO:0043066	negative regulation of apoptosis	1.33E-17
GO:0043069	negative regulation of programmed cell death	2.05E-17
GO:0032270	positive regulation of cellular protein metabolic process	2.23E-17
GO:0043085	positive regulation of catalytic activity	2.98E-17
GO:0045597	positive regulation of cell differentiation	4.02E-17
GO:0008284	positive regulation of cell proliferation	4.32E-17
GO:0045637	regulation of myeloid cell differentiation	5.02E-17
GO:0007166	cell surface receptor linked signaling pathway	8.13E-17
GO:0060548	negative regulation of cell death	9.29E-17
GO:0051247	positive regulation of protein metabolic process	1.25E-16
GO:0030330	DNA damage response, signal transduction by p53 class mediator	4.34E-16
GO:0032446	protein modification by small protein conjugation	5.89E-16
GO:0048583	regulation of response to stimulus	5.91E-16
GO:0072331	signal transduction by p53 class mediator	8.16E-16
GO:0009653	anatomical structure morphogenesis	8.30E-16
GO:0007569	cell aging	8.39E-16
GO:0007179	transforming growth factor beta receptor signaling pathway	8.62E-16
GO:0051128	regulation of cellular component organization	1.17E-15
GO:0031401	positive regulation of protein modification process	1.97E-15
GO:0002376	immune system process	2.27E-15
GO:0030097	hemopoiesis	2.42E-15
GO:0009790	embryo development	4.98E-15
GO:0002682	regulation of immune system process	7.07E-15
GO:0070647	protein modification by small protein conjugation or removal	9.72E-15
GO:0006260	DNA replication	9.94E-15
GO:0008285	negative regulation of cell proliferation	1.09E-14
GO:0048534	hemopoietic or lymphoid organ development	1.28E-14
GO:0016567	protein ubiquitination	1.38E-14
GO:0042325	regulation of phosphorylation	1.56E-14
GO:0051338	regulation of transferase activity	1.68E-14
GO:0009719	response to endogenous stimulus	1.79E-14
GO:0018205	peptidyl-lysine modification	1.87E-14
GO:2000045	regulation of G1/S transition of mitotic cell cycle	1.91E-14
GO:0051320	S phase	2.29E-14
GO:0009314	response to radiation	7.04E-14
GO:0048585	negative regulation of response to stimulus	7.97E-14
GO:0002520	immune system development	8.56E-14
GO:0044092	negative regulation of molecular function	1.08E-13
GO:0051248	negative regulation of protein metabolic process	1.25E-13
GO:0000084	S phase of mitotic cell cycle	1.33E-13
GO:0045646	regulation of erythrocyte differentiation	2.21E-13
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	2.46E-13
GO:0019220	regulation of phosphate metabolic process	2.54E-13
GO:0051174	regulation of phosphorus metabolic process	2.54E-13
GO:0031400	negative regulation of protein modification process	3.99E-13
GO:0043549	regulation of kinase activity	6.00E-13
GO:0032269	negative regulation of cellular protein metabolic process	8.29E-13

GO:0071158	positive regulation of cell cycle arrest	8.41E-13
GO:0009966	regulation of signal transduction	9.56E-13
GO:0006281	DNA repair	1.16E-12
GO:0009725	response to hormone stimulus	1.72E-12
GO:0009628	response to abiotic stimulus	1.89E-12
GO:0080134	regulation of response to stress	2.64E-12
GO:0033044	regulation of chromosome organization	2.97E-12
GO:0018193	peptidyl-amino acid modification	2.98E-12
GO:0071310	cellular response to organic substance	5.82E-12
GO:0009968	negative regulation of signal transduction	6.15E-12
GO:0002573	myeloid leukocyte differentiation	6.54E-12
GO:0030522	intracellular receptor mediated signaling pathway	7.54E-12
GO:0010646	regulation of cell communication	8.20E-12
GO:0001932	regulation of protein phosphorylation	9.39E-12
GO:0023051	regulation of signaling	9.76E-12
GO:0051090	regulation of sequence-specific DNA binding transcription factor activity	1.01E-11
GO:0031056	regulation of histone modification	1.16E-11
GO:0031571	mitotic cell cycle G1/S transition DNA damage checkpoint	1.16E-11
GO:0007568	aging	1.22E-11
GO:0006917	induction of apoptosis	1.63E-11
GO:0012502	induction of programmed cell death	2.06E-11
GO:0023057	negative regulation of signaling	3.39E-11
GO:0007154	cell communication	3.79E-11
GO:0010648	negative regulation of cell communication	3.88E-11
GO:0090068	positive regulation of cell cycle process	6.15E-11
GO:0031575	mitotic cell cycle G1/S transition checkpoint	6.64E-11
GO:0071779	G1/S transition checkpoint	8.66E-11
GO:0000086	G2/M transition of mitotic cell cycle	1.02E-10
GO:0045859	regulation of protein kinase activity	1.24E-10
GO:0000079	regulation of cyclin-dependent protein kinase activity	1.46E-10
GO:0040008	regulation of growth	1.49E-10
GO:0009611	response to wounding	2.25E-10
GO:0000083	regulation of transcription involved in G1/S phase of mitotic cell cycle	3.10E-10
GO:0009411	response to UV	4.29E-10
GO:0030218	erythrocyte differentiation	4.29E-10
GO:0035556	intracellular signal transduction	5.51E-10
GO:0033043	regulation of organelle organization	5.64E-10
GO:0007399	nervous system development	5.85E-10
GO:0006383	transcription from RNA polymerase III promoter	6.76E-10
GO:0048468	cell development	7.67E-10
GO:0034101	erythrocyte homeostasis	8.86E-10
GO:0009057	macromolecule catabolic process	9.77E-10
GO:0009416	response to light stimulus	1.12E-09
GO:0051129	negative regulation of cellular component organization	1.45E-09
GO:0002521	leukocyte differentiation	1.45E-09
GO:0043086	negative regulation of catalytic activity	1.51E-09
GO:0071478	cellular response to radiation	1.74E-09
GO:0043525	positive regulation of neuron apoptosis	1.88E-09
GO:0051052	regulation of DNA metabolic process	1.93E-09
GO:0090398	cellular senescence	2.27E-09
GO:0009887	organ morphogenesis	2.32E-09
GO:0031145	anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	2.83E-09
GO:0051093	negative regulation of developmental process	3.07E-09
GO:0072401	signal transduction involved in DNA integrity checkpoint	4.71E-09
GO:0072422	signal transduction involved in DNA damage checkpoint	4.71E-09
GO:0018394	peptidyl-lysine acetylation	4.91E-09
GO:0031396	regulation of protein ubiquitination	5.10E-09
GO:0043414	macromolecule methylation	5.10E-09
GO:0006338	chromatin remodeling	5.38E-09
GO:0072395	signal transduction involved in cell cycle checkpoint	5.45E-09
GO:0006730	one-carbon metabolic process	5.63E-09
GO:0071495	cellular response to endogenous stimulus	8.01E-09
GO:0045639	positive regulation of myeloid cell differentiation	8.31E-09
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	8.58E-09
GO:0032259	methylation	1.02E-08
GO:0040007	growth	1.07E-08
GO:0043921	modulation by host of viral transcription	1.10E-08
GO:0052472	modulation by host of symbiont transcription	1.10E-08
GO:0000280	nuclear division	1.34E-08
GO:0007067	mitosis	1.34E-08
GO:0071214	cellular response to abiotic stimulus	1.46E-08
GO:0050792	regulation of viral reproduction	1.47E-08

GO:0007423	sensory organ development	1.50E-08
GO:0048524	positive regulation of viral reproduction	1.61E-08
GO:0071453	cellular response to oxygen levels	1.61E-08
GO:0090342	regulation of cell aging	1.63E-08
GO:0044265	cellular macromolecule catabolic process	1.67E-08
GO:0006473	protein acetylation	2.10E-08
GO:0048872	homeostasis of number of cells	2.14E-08
GO:0052312	modulation of transcription in other organism involved in symbiotic interaction	2.38E-08
GO:0044257	cellular protein catabolic process	2.54E-08
GO:0000087	M phase of mitotic cell cycle	2.58E-08
GO:0030163	protein catabolic process	2.71E-08
GO:0043161	proteasomal ubiquitin-dependent protein catabolic process	3.02E-08
GO:0072358	cardiovascular system development	3.21E-08
GO:0072359	circulatory system development	3.21E-08
GO:0048011	nerve growth factor receptor signaling pathway	3.63E-08
GO:0001654	eye development	3.67E-08
GO:0048545	response to steroid hormone stimulus	3.85E-08
GO:0010035	response to inorganic substance	4.06E-08
GO:0048285	organelle fission	4.36E-08
GO:0010498	proteasomal protein catabolic process	4.65E-08
GO:0030518	steroid hormone receptor signaling pathway	5.08E-08
GO:0060284	regulation of cell development	5.23E-08
GO:0016573	histone acetylation	6.07E-08
GO:0045648	positive regulation of erythrocyte differentiation	6.61E-08
GO:0008361	regulation of cell size	8.33E-08
GO:0018393	internal peptidyl-lysine acetylation	8.59E-08
GO:0042592	homeostatic process	8.76E-08
GO:0071900	regulation of protein serine/threonine kinase activity	8.83E-08
GO:0010639	negative regulation of organelle organization	9.27E-08
GO:0030521	androgen receptor signaling pathway	9.88E-08
GO:0051603	proteolysis involved in cellular protein catabolic process	1.20E-07
GO:0051851	modification by host of symbiont morphology or physiology	1.21E-07
GO:0010212	response to ionizing radiation	1.22E-07
GO:0046782	regulation of viral transcription	1.30E-07
GO:0006475	internal protein amino acid acetylation	1.41E-07
GO:0006977	DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	1.48E-07
GO:0072413	signal transduction involved in mitotic cell cycle checkpoint	1.48E-07
GO:0072431	signal transduction involved in mitotic cell cycle G1/S transition DNA damage checkpoint	1.48E-07
GO:0072474	signal transduction involved in mitotic cell cycle G1/S checkpoint	1.48E-07
GO:0072404	signal transduction involved in G1/S transition checkpoint	1.69E-07
GO:0031397	negative regulation of protein ubiquitination	1.71E-07
GO:0043523	regulation of neuron apoptosis	1.74E-07
GO:0048568	embryonic organ development	1.89E-07
GO:0006511	ubiquitin-dependent protein catabolic process	1.91E-07
GO:0006476	protein deacetylation	1.92E-07
GO:0009888	tissue development	2.06E-07
GO:0010038	response to metal ion	2.28E-07
GO:0019941	modification-dependent protein catabolic process	2.29E-07
GO:0043543	protein acylation	2.35E-07
GO:0071456	cellular response to hypoxia	2.47E-07
GO:0043632	modification-dependent macromolecule catabolic process	2.48E-07
GO:0035601	protein deacylation	2.79E-07
GO:0000279	M phase	3.04E-07
GO:0016575	histone deacetylation	3.50E-07
GO:0051439	regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	4.00E-07
GO:0031576	G2/M transition checkpoint	4.12E-07
GO:0051702	interaction with symbiont	4.45E-07
GO:0043010	camera-type eye development	5.21E-07
GO:2000241	regulation of reproductive process	5.22E-07
GO:0019058	viral infectious cycle	5.83E-07
GO:0051402	neuron apoptosis	5.83E-07
GO:0032870	cellular response to hormone stimulus	6.37E-07
GO:0032535	regulation of cellular component size	6.41E-07
GO:0044248	cellular catabolic process	6.91E-07
GO:0070997	neuron death	7.11E-07
GO:0007243	intracellular protein kinase cascade	7.40E-07
GO:0016571	histone methylation	9.74E-07
GO:0045596	negative regulation of cell differentiation	1.05E-06
GO:0008629	induction of apoptosis by intracellular signals	1.08E-06
GO:0034968	histone lysine methylation	1.17E-06
GO:0050878	regulation of body fluid levels	1.30E-06

GO:0031057	negative regulation of histone modification	1.66E-06
GO:0090066	regulation of anatomical structure size	1.89E-06
GO:0016049	cell growth	2.15E-06
GO:0008630	DNA damage response, signal transduction resulting in induction of apoptosis	2.30E-06
GO:0043281	regulation of caspase activity	2.31E-06
GO:0048646	anatomical structure formation involved in morphogenesis	2.33E-06
GO:0031572	G2/M transition DNA damage checkpoint	2.38E-06
GO:0043280	positive regulation of caspase activity	2.42E-06
GO:0080135	regulation of cellular response to stress	2.74E-06
GO:0010001	glial cell differentiation	2.96E-06
GO:0006479	protein methylation	3.02E-06
GO:0008213	protein alkylation	3.02E-06
GO:0043923	positive regulation by host of viral transcription	3.24E-06
GO:0006461	protein complex assembly	3.63E-06
GO:0048699	generation of neurons	3.68E-06
GO:0010948	negative regulation of cell cycle process	3.74E-06
GO:0070271	protein complex biogenesis	3.76E-06
GO:0010952	positive regulation of peptidase activity	4.01E-06
GO:0034644	cellular response to UV	4.04E-06
GO:0040029	regulation of gene expression, epigenetic	4.42E-06
GO:0051438	regulation of ubiquitin-protein ligase activity	4.51E-06
GO:0007417	central nervous system development	4.59E-06
GO:0035821	modification of morphology or physiology of other organism	4.75E-06
GO:0051817	modification of morphology or physiology of other organism involved in symbiotic interaction	4.75E-06
GO:2000243	positive regulation of reproductive process	4.92E-06
GO:0052548	regulation of endopeptidase activity	5.09E-06
GO:0045787	positive regulation of cell cycle	5.27E-06
GO:0048598	embryonic morphogenesis	5.30E-06
GO:0032504	multicellular organism reproduction	5.39E-06
GO:0048609	multicellular organismal reproductive process	5.39E-06
GO:0044085	cellular component biogenesis	5.93E-06
GO:0051437	positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	5.96E-06
GO:0043009	chordate embryonic development	6.31E-06
GO:0051340	regulation of ligase activity	6.34E-06
GO:0052547	regulation of peptidase activity	7.01E-06
GO:0006302	double-strand break repair	8.37E-06
GO:0009792	embryo development ending in birth or egg hatching	8.41E-06
GO:0042060	wound healing	8.67E-06
GO:0006508	proteolysis	8.72E-06
GO:0045165	cell fate commitment	8.97E-06
GO:0001775	cell activation	8.98E-06
GO:0071482	cellular response to light stimulus	8.98E-06
GO:0051403	stress-activated MAPK cascade	9.20E-06
GO:0007596	blood coagulation	9.28E-06
GO:0031398	positive regulation of protein ubiquitination	1.01E-05
GO:0071496	cellular response to external stimulus	1.03E-05
GO:0007599	hemostasis	1.08E-05
GO:0051130	positive regulation of cellular component organization	1.18E-05
GO:0033143	regulation of steroid hormone receptor signaling pathway	1.21E-05
GO:0042063	gliogenesis	1.25E-05
GO:0006978	DNA damage response, signal transduction by p53 class mediator resulting in transcription of p21 class mediator	1.37E-05
GO:0050817	coagulation	1.38E-05
GO:0051960	regulation of nervous system development	1.58E-05
GO:0051443	positive regulation of ubiquitin-protein ligase activity	1.87E-05
GO:0007507	heart development	1.98E-05
GO:0042772	DNA damage response, signal transduction resulting in transcription	2.04E-05
GO:0032844	regulation of homeostatic process	2.17E-05
GO:0022008	neurogenesis	2.21E-05
GO:0051573	negative regulation of histone H3-K9 methylation	2.32E-05
GO:0043933	macromolecular complex subunit organization	2.37E-05
GO:0050767	regulation of neurogenesis	2.44E-05
GO:0051101	regulation of DNA binding	2.71E-05
GO:0051351	positive regulation of ligase activity	2.71E-05
GO:0050434	positive regulation of viral transcription	2.72E-05
GO:0030219	megakaryocyte differentiation	3.17E-05
GO:0065003	macromolecular complex assembly	3.17E-05
GO:0048584	positive regulation of response to stimulus	3.53E-05
GO:0071822	protein complex subunit organization	3.77E-05
GO:0031331	positive regulation of cellular catabolic process	4.09E-05
GO:0001558	regulation of cell growth	4.14E-05
GO:0006913	nucleocytoplasmic transport	4.16E-05

GO:0051169	nuclear transport	4.51E-05
GO:0030225	macrophage differentiation	4.60E-05
GO:2000756	regulation of peptidyl-lysine acetylation	4.60E-05
GO:0048145	regulation of fibroblast proliferation	4.60E-05
GO:0009056	catabolic process	5.40E-05
GO:0009896	positive regulation of catabolic process	5.60E-05
GO:0048144	fibroblast proliferation	5.91E-05
GO:0009605	response to external stimulus	6.00E-05
GO:0043627	response to estrogen stimulus	6.61E-05
GO:0006261	DNA-dependent DNA replication	6.84E-05
GO:0006310	DNA recombination	7.11E-05
GO:0034599	cellular response to oxidative stress	7.27E-05
GO:0007420	brain development	8.20E-05
GO:0035666	TRIF-dependent toll-like receptor signaling pathway	9.71E-05
GO:0051436	negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	9.71E-05
GO:0048610	cellular process involved in reproduction	1.05E-04
GO:0043470	regulation of carbohydrate catabolic process	1.06E-04
GO:0043471	regulation of cellular carbohydrate catabolic process	1.06E-04
GO:0045814	negative regulation of gene expression, epigenetic	1.06E-04
GO:0022607	cellular component assembly	1.32E-04
GO:0071897	DNA biosynthetic process	1.44E-04
GO:0042493	response to drug	1.54E-04
GO:0001701	in utero embryonic development	1.54E-04
GO:0019080	viral genome expression	1.56E-04
GO:0019083	viral transcription	1.56E-04
GO:0002756	MyD88-independent toll-like receptor signaling pathway	1.61E-04
GO:0034138	toll-like receptor 3 signaling pathway	1.77E-04
GO:0009894	regulation of catabolic process	1.87E-04
GO:0090343	positive regulation of cell aging	1.90E-04
GO:0090344	negative regulation of cell aging	1.90E-04
GO:0060765	regulation of androgen receptor signaling pathway	2.19E-04
GO:0050673	epithelial cell proliferation	2.19E-04
GO:0048592	eye morphogenesis	2.33E-04
GO:0031058	positive regulation of histone modification	2.51E-04
GO:0051352	negative regulation of ligase activity	2.56E-04
GO:0051444	negative regulation of ubiquitin-protein ligase activity	2.56E-04
GO:0042327	positive regulation of phosphorylation	2.57E-04
GO:0070482	response to oxygen levels	2.83E-04
GO:0030512	negative regulation of transforming growth factor beta receptor signaling pathway	3.06E-04
GO:0045444	fat cell differentiation	3.14E-04
GO:0051570	regulation of histone H3-K9 methylation	3.15E-04
GO:2000377	regulation of reactive oxygen species metabolic process	3.34E-04
GO:0010562	positive regulation of phosphorus metabolic process	3.51E-04
GO:0045937	positive regulation of phosphate metabolic process	3.51E-04
GO:0051348	negative regulation of transferase activity	3.71E-04
GO:0006921	cellular component disassembly involved in apoptosis	3.96E-04
GO:0008063	Toll signaling pathway	3.96E-04
GO:0043966	histone H3 acetylation	4.18E-04
GO:0031061	negative regulation of histone methylation	4.92E-04
GO:0010608	posttranscriptional regulation of gene expression	4.97E-04
GO:0017015	regulation of transforming growth factor beta receptor signaling pathway	5.21E-04
GO:0071241	cellular response to inorganic substance	5.21E-04
GO:0030855	epithelial cell differentiation	5.52E-04
GO:0051384	response to glucocorticoid stimulus	5.77E-04
GO:0006354	transcription elongation, DNA-dependent	5.87E-04
GO:0045321	leukocyte activation	6.10E-04
GO:0032869	cellular response to insulin stimulus	6.82E-04
GO:0051336	regulation of hydrolase activity	7.23E-04
GO:0031060	regulation of histone methylation	7.47E-04
GO:0031960	response to corticosteroid stimulus	8.14E-04
GO:0030278	regulation of ossification	8.45E-04
GO:0001890	placenta development	8.57E-04
GO:0034142	toll-like receptor 4 signaling pathway	8.72E-04
GO:0042771	DNA damage response, signal transduction by p53 class mediator resulting in induction of apoptosis	8.92E-04
GO:0043516	regulation of DNA damage response, signal transduction by p53 class mediator	8.92E-04
GO:0048146	positive regulation of fibroblast proliferation	9.23E-04
GO:0032879	regulation of localization	9.38E-04
GO:0030308	negative regulation of cell growth	9.79E-04
GO:0002757	immune response-activating signal transduction	0.001062
GO:0051054	positive regulation of DNA metabolic process	0.001094
GO:0000209	protein polyubiquitination	0.001115
GO:0001934	positive regulation of protein phosphorylation	0.001129

GO:0006275	regulation of DNA replication	0.001218
GO:0010720	positive regulation of cell development	0.001259
GO:0007283	spermatogenesis	0.001287
GO:0048232	male gamete generation	0.001315
GO:0032070	regulation of deoxyribonuclease activity	0.001326
GO:0031668	cellular response to extracellular stimulus	0.001399
GO:0002764	immune response-regulating signaling pathway	0.001439
GO:0006979	response to oxidative stress	0.00144
GO:0043518	negative regulation of DNA damage response, signal transduction by p53 class mediator	0.001457
GO:0060766	negative regulation of androgen receptor signaling pathway	0.001457
GO:0035065	regulation of histone acetylation	0.001458
GO:0051345	positive regulation of hydrolase activity	0.001489
GO:0071375	cellular response to peptide hormone stimulus	0.001515
GO:0032868	response to insulin stimulus	0.001555
GO:0045088	regulation of innate immune response	0.001603
GO:0045860	positive regulation of protein kinase activity	0.001652
GO:0045792	negative regulation of cell size	0.001694
GO:0051568	histone H3-K4 methylation	0.0017
GO:0010627	regulation of intracellular protein kinase cascade	0.001753
GO:0006919	activation of caspase activity	0.00187
GO:0051091	positive regulation of sequence-specific DNA binding transcription factor activity	0.001913
GO:0002221	pattern recognition receptor signaling pathway	0.001932
GO:0005981	regulation of glycogen catabolic process	0.00197
GO:0071780	mitotic cell cycle G2/M transition checkpoint	0.00197
GO:0090399	replicative senescence	0.00197
GO:0010332	response to gamma radiation	0.002013
GO:0048708	astrocyte differentiation	0.002013
GO:0048863	stem cell differentiation	0.002028
GO:0030182	neuron differentiation	0.002141
GO:0010906	regulation of glucose metabolic process	0.002265
GO:0035162	embryonic hemopoiesis	0.002274
GO:0072332	signal transduction by p53 class mediator resulting in induction of apoptosis	0.002274
GO:0002758	innate immune response-activating signal transduction	0.002464
GO:0034504	protein localization to nucleus	0.002624
GO:0034130	toll-like receptor 1 signaling pathway	0.002627
GO:0050769	positive regulation of neurogenesis	0.002824
GO:0001836	release of cytochrome c from mitochondria	0.002865
GO:0002761	regulation of myeloid leukocyte differentiation	0.002902
GO:0071248	cellular response to metal ion	0.002902
GO:0033674	positive regulation of kinase activity	0.002948
GO:0001503	ossification	0.002967
GO:0031329	regulation of cellular catabolic process	0.003005
GO:0060429	epithelium development	0.003064
GO:0034097	response to cytokine stimulus	0.003165
GO:0090239	regulation of histone H4 acetylation	0.003255
GO:0090101	negative regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	0.003271
GO:0045737	positive regulation of cyclin-dependent protein kinase activity	0.003381
GO:0051567	histone H3-K9 methylation	0.003381
GO:0030851	granulocyte differentiation	0.003413
GO:0016458	gene silencing	0.003469
GO:0031098	stress-activated protein kinase signaling cascade	0.003496
GO:0009607	response to biotic stimulus	0.003574
GO:0051347	positive regulation of transferase activity	0.003706
GO:0045926	negative regulation of growth	0.003716
GO:0048593	camera-type eye morphogenesis	0.003732
GO:0050678	regulation of epithelial cell proliferation	0.003832
GO:0014013	regulation of gliogenesis	0.003894
GO:0048660	regulation of smooth muscle cell proliferation	0.003894
GO:0043467	regulation of generation of precursor metabolites and energy	0.003975
GO:0002218	activation of innate immune response	0.004078
GO:0010869	regulation of receptor biosynthetic process	0.004309
GO:0043434	response to peptide hormone stimulus	0.004384
GO:0000165	MAPKKK cascade	0.004466
GO:0034134	toll-like receptor 2 signaling pathway	0.004655
GO:0033002	muscle cell proliferation	0.00486
GO:0001892	embryonic placenta development	0.004882
GO:0048659	smooth muscle cell proliferation	0.004882
GO:0031347	regulation of defense response	0.005108
GO:0051569	regulation of histone H3-K4 methylation	0.005422
GO:0051149	positive regulation of muscle cell differentiation	0.005857
GO:0001889	liver development	0.006276

GO:0031647	regulation of protein stability	0.006419
GO:0006606	protein import into nucleus	0.006525
GO:0000216	M/G1 transition of mitotic cell cycle	0.006783
GO:0002755	MyD88-dependent toll-like receptor signaling pathway	0.006783
GO:0061008	hepaticobiliary system development	0.006808
GO:0001666	response to hypoxia	0.006872
GO:0007219	Notch signaling pathway	0.007093
GO:0051170	nuclear import	0.007099
GO:0010675	regulation of cellular carbohydrate metabolic process	0.007123
GO:0046649	lymphocyte activation	0.007341
GO:0006109	regulation of carbohydrate metabolic process	0.007505
GO:0009991	response to extracellular stimulus	0.007535
GO:0042176	regulation of protein catabolic process	0.008108
GO:0006270	DNA-dependent DNA replication initiation	0.008694
GO:0043392	negative regulation of DNA binding	0.008694
GO:0042113	B cell activation	0.008939
GO:0008637	apoptotic mitochondrial changes	0.008993
GO:0045638	negative regulation of myeloid cell differentiation	0.008993
GO:0030162	regulation of proteolysis	0.009533
GO:0019953	sexual reproduction	0.009643
GO:0050872	white fat cell differentiation	0.010029
GO:0032355	response to estradiol stimulus	0.010539
GO:0071479	cellular response to ionizing radiation	0.010688
GO:0009304	tRNA transcription	0.011088
GO:0010243	response to organic nitrogen	0.011157
GO:0035295	tube development	0.011235
GO:0043967	histone H4 acetylation	0.011814
GO:2000145	regulation of cell motility	0.012811
GO:0030900	forebrain development	0.012915
GO:0033365	protein localization to organelle	0.013
GO:0006282	regulation of DNA repair	0.013033
GO:0045089	positive regulation of innate immune response	0.014236
GO:0035019	somatic stem cell maintenance	0.014357
GO:0032800	receptor biosynthetic process	0.014389
GO:0021700	developmental maturation	0.014721
GO:0006469	negative regulation of protein kinase activity	0.014756
GO:0002224	toll-like receptor signaling pathway	0.014882
GO:0001824	blastocyst development	0.015185
GO:0019827	stem cell maintenance	0.017202
GO:0050776	regulation of immune response	0.019206
GO:0051098	regulation of binding	0.019234
GO:0006952	defense response	0.019243
GO:0045685	regulation of glial cell differentiation	0.019407
GO:0033144	negative regulation of steroid hormone receptor signaling pathway	0.02002
GO:0031349	positive regulation of defense response	0.020951
GO:0051270	regulation of cellular component movement	0.021757
GO:0033673	negative regulation of kinase activity	0.0218
GO:0050679	positive regulation of epithelial cell proliferation	0.022264
GO:0048864	stem cell development	0.023127
GO:0019079	viral genome replication	0.024578
GO:0048661	positive regulation of smooth muscle cell proliferation	0.024578
GO:0031669	cellular response to nutrient levels	0.024976
GO:0045090	retroviral genome replication	0.026152
GO:0010638	positive regulation of organelle organization	0.026237
GO:0040012	regulation of locomotion	0.028014
GO:0048015	phosphatidylinositol-mediated signaling	0.028965
GO:0048017	inositol lipid-mediated signaling	0.028965
GO:0051049	regulation of transport	0.030372
GO:0070979	protein K11-linked ubiquitination	0.031271
GO:0048596	embryonic camera-type eye morphogenesis	0.03143
GO:0090092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	0.032492
GO:0010741	negative regulation of intracellular protein kinase cascade	0.033756
GO:0060216	definitive hemopoiesis	0.035897
GO:0007098	centrosome cycle	0.036812
GO:0008635	activation of caspase activity by cytochrome c	0.036949
GO:0034770	histone H4-K20 methylation	0.036949
GO:0042789	mRNA transcription from RNA polymerase II promoter	0.036949
GO:0045821	positive regulation of glycolysis	0.036949
GO:0051123	RNA polymerase II transcriptional preinitiation complex assembly	0.036949
GO:0051571	positive regulation of histone H3-K4 methylation	0.036949
GO:0070897	DNA-dependent transcriptional preinitiation complex assembly	0.036949
GO:0090400	stress-induced premature senescence	0.036949

GO:2000757	negative regulation of peptidyl-lysine acetylation	0.036949
GO:2000772	regulation of cellular senescence	0.036949
GO:0007276	gamete generation	0.040263
GO:0002684	positive regulation of immune system process	0.040982
GO:0022603	regulation of anatomical structure morphogenesis	0.042096
GO:0048511	rhythmic process	0.043484
GO:0051259	protein oligomerization	0.044744
GO:0006305	DNA alkylation	0.046015
GO:0006306	DNA methylation	0.046015
GO:0002088	lens development in camera-type eye	0.046412
GO:0009303	rRNA transcription	0.046435
GO:0010907	positive regulation of glucose metabolic process	0.046435
GO:0006928	cellular component movement	0.048393
GO:0000085	G2 phase of mitotic cell cycle	0.050352
GO:0006975	DNA damage induced protein phosphorylation	0.050352
GO:0007095	mitotic cell cycle G2/M transition DNA damage checkpoint	0.050352
GO:0010870	positive regulation of receptor biosynthetic process	0.050352
GO:0051319	G2 phase	0.050352
GO:2000036	regulation of stem cell maintenance	0.050352
GO:0008543	fibroblast growth factor receptor signaling pathway	0.051394
GO:0033500	carbohydrate homeostasis	0.051394
GO:0042593	glucose homeostasis	0.051394
GO:0048469	cell maturation	0.051394
GO:0008406	gonad development	0.051738
GO:0030224	monocyte differentiation	0.05239
GO:0043388	positive regulation of DNA binding	0.05239
GO:0048048	embryonic eye morphogenesis	0.052944
GO:0001944	vasculature development	0.053333
GO:0009967	positive regulation of signal transduction	0.055231
GO:0000724	double-strand break repair via homologous recombination	0.056797
GO:0051641	cellular localization	0.057323
GO:0005980	glycogen catabolic process	0.059033
GO:0071103	DNA conformation change	0.059937
GO:0000725	recombinational repair	0.060844
GO:0030168	platelet activation	0.061366
GO:0017038	protein import	0.064129
GO:0071901	negative regulation of protein serine/threonine kinase activity	0.064917
GO:0071824	protein-DNA complex subunit organization	0.065016
GO:0072593	reactive oxygen species metabolic process	0.066173
GO:0009251	glucan catabolic process	0.066183
GO:0044247	cellular polysaccharide catabolic process	0.066183
GO:0051353	positive regulation of oxidoreductase activity	0.066183
GO:0043922	negative regulation by host of viral transcription	0.066226
GO:0030221	basophil differentiation	0.067852
GO:0090241	negative regulation of histone H4 acetylation	0.067852
GO:0006342	chromatin silencing	0.07391
GO:0010165	response to X-ray	0.07391
GO:0048732	gland development	0.077569
GO:0031100	organ regeneration	0.077828
GO:0000723	telomere maintenance	0.079186
GO:0006271	DNA strand elongation involved in DNA replication	0.08239
GO:0010389	regulation of G2/M transition of mitotic cell cycle	0.08239
GO:0032200	telomere organization	0.08442
GO:0007088	regulation of mitosis	0.085042
GO:0051783	regulation of nuclear division	0.085042
GO:0045346	regulation of MHC class II biosynthetic process	0.085319
GO:0031667	response to nutrient levels	0.086448
GO:0051707	response to other organism	0.092595
GO:0043254	regulation of protein complex assembly	0.093084
GO:0031076	embryonic camera-type eye development	0.095681
GO:0044281	small molecule metabolic process	0.099309
GO:0010565	regulation of cellular ketone metabolic process	0.099461
GO:0051591	response to cAMP	0.101279
GO:0010833	telomere maintenance via telomere lengthening	0.101454
GO:0061180	mammary gland epithelium development	0.105709
GO:0045471	response to ethanol	0.10619
GO:0009299	mRNA transcription	0.107743
GO:0045342	MHC class II biosynthetic process	0.107743
GO:0046886	positive regulation of hormone biosynthetic process	0.107743
GO:0002253	activation of immune response	0.107852
GO:0061061	muscle structure development	0.108825
GO:0031099	regeneration	0.109074
GO:0050778	positive regulation of immune response	0.110754

GO:0010226	response to lithium ion	0.111862
GO:0022616	DNA strand elongation	0.111862
GO:0051412	response to corticosterone stimulus	0.111862
GO:0043433	negative regulation of sequence-specific DNA binding transcription factor activity	0.114705
GO:0045137	development of primary sexual characteristics	0.116568
GO:0048562	embryonic organ morphogenesis	0.119163
GO:0010647	positive regulation of cell communication	0.119214
GO:0032496	response to lipopolysaccharide	0.121768
GO:0023056	positive regulation of signaling	0.12202
GO:0006955	immune response	0.122188
GO:0000226	microtubule cytoskeleton organization	0.123108
GO:0006312	mitotic recombination	0.123108
GO:0010039	response to iron ion	0.123108
GO:0045600	positive regulation of fat cell differentiation	0.123108
GO:0019216	regulation of lipid metabolic process	0.123903
GO:0048709	oligodendrocyte differentiation	0.129613
GO:0031062	positive regulation of histone methylation	0.13329
GO:0032897	negative regulation of viral transcription	0.13329
GO:0001568	blood vessel development	0.138973
GO:0042692	muscle cell differentiation	0.139671
GO:0032881	regulation of polysaccharide metabolic process	0.162642
GO:0070873	regulation of glycogen metabolic process	0.162642
GO:0018022	peptidyl-lysine methylation	0.162818
GO:0060260	regulation of transcription initiation from RNA polymerase II promoter	0.162818
GO:0071850	mitotic cell cycle arrest	0.162818
GO:2000378	negative regulation of reactive oxygen species metabolic process	0.162818
GO:0002526	acute inflammatory response	0.164408
GO:0071845	cellular component disassembly at cellular level	0.167193
GO:0051385	response to mineralocorticoid stimulus	0.177245
GO:0048870	cell motility	0.178205
GO:0051674	localization of cell	0.178205
GO:0022411	cellular component disassembly	0.18143
GO:0010676	positive regulation of cellular carbohydrate metabolic process	0.193056
GO:0045913	positive regulation of carbohydrate metabolic process	0.193056
GO:0007389	pattern specification process	0.195836
GO:0070076	histone lysine demethylation	0.196156
GO:0002237	response to molecule of bacterial origin	0.196727
GO:0006344	maintenance of chromatin silencing	0.197492
GO:0021537	telencephalon development	0.197492
GO:0021603	cranial nerve formation	0.197492
GO:0032071	regulation of endodeoxyribonuclease activity	0.197492
GO:0051097	negative regulation of helicase activity	0.197492
GO:2000773	negative regulation of cellular senescence	0.197492
GO:0051100	negative regulation of binding	0.198361
GO:0006954	inflammatory response	0.212506
GO:0033036	macromolecule localization	0.214877
GO:0045839	negative regulation of mitosis	0.227229
GO:0051784	negative regulation of nuclear division	0.227229
GO:0006289	nucleotide-excision repair	0.231437
GO:0006368	transcription elongation from RNA polymerase II promoter	0.231437
GO:0008286	insulin receptor signaling pathway	0.231438
GO:0071902	positive regulation of protein serine/threonine kinase activity	0.232658
GO:0032352	positive regulation of hormone metabolic process	0.233011
GO:0043353	enucleate erythrocyte differentiation	0.233011
GO:0006333	chromatin assembly or disassembly	0.23785
GO:0045862	positive regulation of proteolysis	0.243059
GO:0007005	mitochondrion organization	0.254777
GO:0051147	regulation of muscle cell differentiation	0.25521
GO:0033574	response to testosterone stimulus	0.265809
GO:0006304	DNA modification	0.268043
GO:0045664	regulation of neuron differentiation	0.274122
GO:0006359	regulation of transcription from RNA polymerase III promoter	0.274339
GO:0010390	histone monoubiquitination	0.274339
GO:0030949	positive regulation of vascular endothelial growth factor receptor signaling pathway	0.274339
GO:0031065	positive regulation of histone deacetylation	0.274339
GO:0061029	eyelid development in camera-type eye	0.274339
GO:2000142	regulation of transcription initiation, DNA-dependent	0.274339
GO:0045087	innate immune response	0.280189
GO:0007548	sex differentiation	0.284003
GO:0045687	positive regulation of glial cell differentiation	0.286002
GO:0060749	mammary gland alveolus development	0.286002
GO:0061377	mammary gland lobule development	0.286002

GO:0003006	developmental process involved in reproduction	0.297444
GO:0048010	vascular endothelial growth factor receptor signaling pathway	0.308521
GO:0006301	postreplication repair	0.32051
GO:0016577	histone demethylation	0.32051
GO:0009266	response to temperature stimulus	0.324116
GO:0045667	regulation of osteoblast differentiation	0.324116
GO:0032880	regulation of protein localization	0.329993
GO:0051240	positive regulation of multicellular organismal process	0.334947
GO:0008104	protein localization	0.337283
GO:0051297	centrosome organization	0.338293
GO:0006605	protein targeting	0.348237
GO:0045598	regulation of fat cell differentiation	0.354177
GO:0006482	protein demethylation	0.370962
GO:0008214	protein dealkylation	0.370962
GO:0033158	regulation of protein import into nucleus, translocation	0.370962
GO:0045922	negative regulation of fatty acid metabolic process	0.370962
GO:0071285	cellular response to lithium ion	0.370962
GO:0006006	glucose metabolic process	0.374359
GO:0060041	retina development in camera-type eye	0.381854
GO:0001835	blastocyst hatching	0.385183
GO:0034773	histone H4-K20 trimethylation	0.385183
GO:0035188	hatching	0.385183
GO:0044026	DNA hypermethylation	0.385183
GO:0044027	hypermethylation of CpG island	0.385183
GO:0071684	organism emergence from protective structure	0.385183
GO:2000774	positive regulation of cellular senescence	0.385183
GO:0031023	microtubule organizing center organization	0.403128
GO:0009913	epidermal cell differentiation	0.405823
GO:0032436	positive regulation of proteasomal ubiquitin-dependent protein catabolic process	0.407892
GO:0048608	reproductive structure development	0.42292
GO:0031063	regulation of histone deacetylation	0.424444
GO:0032459	regulation of protein oligomerization	0.424444
GO:0043618	regulation of transcription from RNA polymerase II promoter in response to stress	0.424444
GO:0045736	negative regulation of cyclin-dependent protein kinase activity	0.424444
GO:0060644	mammary gland epithelial cell differentiation	0.424444
GO:2000736	regulation of stem cell differentiation	0.424444
GO:0070201	regulation of establishment of protein localization	0.429652
GO:0000060	protein import into nucleus, translocation	0.435808
GO:0005977	glycogen metabolic process	0.438363
GO:2000177	regulation of neural precursor cell proliferation	0.438363
GO:0060341	regulation of cellular localization	0.446491
GO:0006073	cellular glucan metabolic process	0.457482
GO:0044042	glucan metabolic process	0.457482
GO:0006336	DNA replication-independent nucleosome assembly	0.484211
GO:0034080	CenH3-containing nucleosome assembly at centromere	0.484211
GO:0034724	DNA replication-independent nucleosome organization	0.484211
GO:0060395	SMAD protein signal transduction	0.484211
GO:0016071	mRNA metabolic process	0.486056
GO:0006521	regulation of cellular amino acid metabolic process	0.529416
GO:0048713	regulation of oligodendrocyte differentiation	0.529416
GO:0035239	tube morphogenesis	0.530803
GO:0001779	natural killer cell differentiation	0.550045
GO:0032201	telomere maintenance via semi-conservative replication	0.550045
GO:0071157	negative regulation of cell cycle arrest	0.550045
GO:0030183	B cell differentiation	0.562651
GO:0051649	establishment of localization in cell	0.56824
GO:0045732	positive regulation of protein catabolic process	0.585692
GO:0045995	regulation of embryonic development	0.585692
GO:0006110	regulation of glycolysis	0.621277
GO:0006361	transcription initiation from RNA polymerase I promoter	0.621277
GO:0090312	positive regulation of protein deacetylation	0.621277
GO:0061030	epithelial cell differentiation involved in mammary gland alveolus development	0.630571
GO:0046677	response to antibiotic	0.635734
GO:0031055	chromatin remodeling at centromere	0.69737
GO:0003151	outflow tract morphogenesis	0.71573
GO:0014015	positive regulation of gliogenesis	0.71573
GO:0045740	positive regulation of DNA replication	0.71573
GO:0000722	telomere maintenance via recombination	0.77927
GO:0019433	triglyceride catabolic process	0.77927
GO:0043620	regulation of DNA-dependent transcription in response to stress	0.77927
GO:0046885	regulation of hormone biosynthetic process	0.77927
GO:0006997	nucleus organization	0.80152
GO:0018107	peptidyl-threonine phosphorylation	0.80152

GO:0071383	cellular response to steroid hormone stimulus	0.80152
GO:0030098	lymphocyte differentiation	0.804387
GO:0007249	I-kappaB kinase/NF-kappaB cascade	0.808347
GO:0046777	protein autophosphorylation	0.808347
GO:0051179	localization	0.81824
GO:0033273	response to vitamin	0.826112
GO:0042110	T cell activation	0.827189
GO:0050864	regulation of B cell activation	0.854843
GO:0006278	RNA-dependent DNA replication	0.865745
GO:0070303	negative regulation of stress-activated protein kinase signaling cascade	0.865745
GO:0002683	negative regulation of immune system process	0.90119
GO:0000302	response to reactive oxygen species	0.913287
GO:0007173	epidermal growth factor receptor signaling pathway	0.913287
GO:0051099	positive regulation of binding	0.918534
GO:0032873	negative regulation of stress-activated MAPK cascade	0.930108
GO:0042791	5S class rRNA transcription from RNA polymerase III type 1 promoter	0.930108
GO:0042797	tRNA transcription from RNA polymerase III promoter	0.930108
GO:0045347	negative regulation of MHC class II biosynthetic process	0.930108
GO:0046628	positive regulation of insulin receptor signaling pathway	0.930108
GO:0006953	acute-phase response	0.941532
GO:0018210	peptidyl-threonine modification	0.941532
GO:0043486	histone exchange	0.958155
GO:0060249	anatomical structure homeostasis	0.978948
GO:0016055	Wnt receptor signaling pathway	0.985954
GO:0021543	pallium development	0.9864
GO:0034613	cellular protein localization	0.987784
GO:0006284	base-excision repair	0.991478
GO:0030888	regulation of B cell proliferation	0.991478

Cellular Component

GO id	GO name	adjusted-P
GO:0031981	nuclear lumen	2.56E-120
GO:0044428	nuclear part	3.08E-112
GO:0070013	intracellular organelle lumen	3.90E-111
GO:0043233	organelle lumen	2.68E-110
GO:0031974	membrane-enclosed lumen	2.89E-109
GO:0005654	nucleoplasm	9.90E-106
GO:0005634	nucleus	1.25E-102
GO:0043231	intracellular membrane-bounded organelle	4.22E-72
GO:0043227	membrane-bounded organelle	7.95E-72
GO:0044446	intracellular organelle part	1.32E-65
GO:0044422	organelle part	1.86E-64
GO:0043229	intracellular organelle	6.35E-62
GO:0043226	organelle	1.14E-61
GO:0044424	intracellular part	7.73E-59
GO:0044451	nucleoplasm part	2.22E-52
GO:0005622	intracellular	1.01E-49
GO:0044427	chromosomal part	1.24E-42
GO:0005694	chromosome	3.40E-42
GO:0000785	chromatin	1.49E-37
GO:0043234	protein complex	2.10E-36
GO:0032991	macromolecular complex	1.44E-35
GO:0005667	transcription factor complex	3.55E-34
GO:0000228	nuclear chromosome	7.35E-34
GO:0043228	non-membrane-bounded organelle	4.12E-33
GO:0043232	intracellular non-membrane-bounded organelle	4.12E-33
GO:0044454	nuclear chromosome part	1.37E-31
GO:0044464	cell part	1.82E-27
GO:0005623	cell	1.83E-27
GO:0000790	nuclear chromatin	2.80E-25
GO:0005575	cellular_component	2.89E-22
GO:0005829	cytosol	3.19E-21
GO:0005737	cytoplasm	1.88E-17
GO:0017053	transcriptional repressor complex	1.40E-15
GO:0005730	nucleolus	4.73E-15
GO:0000307	cyclin-dependent protein kinase holoenzyme complex	6.64E-14
GO:0035097	histone methyltransferase complex	1.78E-12
GO:0034708	methyltransferase complex	3.65E-12
GO:0016585	chromatin remodeling complex	8.14E-12
GO:0016604	nuclear body	4.18E-10
GO:0016363	nuclear matrix	6.66E-10
GO:0000792	heterochromatin	8.27E-10

GO:0034399	nuclear periphery	1.89E-09
GO:0016605	PML body	1.97E-09
GO:0005657	replication fork	2.18E-07
GO:0000118	histone deacetylase complex	2.60E-06
GO:0044444	cytoplasmic part	3.40E-06
GO:0031519	PcG protein complex	4.78E-06
GO:0000775	chromosome, centromeric region	7.66E-06
GO:0000791	euchromatin	1.37E-05
GO:0000793	condensed chromosome	8.77E-05
GO:0005720	nuclear heterochromatin	9.71E-05
GO:0016580	Sin3 complex	7.32E-04
GO:0035098	ESC/E(Z) complex	7.32E-04
GO:0070822	Sin3-type complex	7.32E-04
GO:0015630	microtubule cytoskeleton	7.36E-04
GO:0035189	Rb-E2F complex	0.001326
GO:0005635	nuclear envelope	0.003046
GO:0000794	condensed nuclear chromosome	0.003677
GO:0005819	spindle	0.005484
GO:0005856	cytoskeleton	0.006478
GO:0000780	condensed nuclear chromosome, centromeric region	0.010029
GO:0005815	microtubule organizing center	0.010464
GO:0000152	nuclear ubiquitin ligase complex	0.010688
GO:0072357	PTW/PP1 phosphatase complex	0.011088
GO:0070688	MLL5-L complex	0.017595
GO:0016581	NuRD complex	0.02002
GO:0031965	nuclear membrane	0.023471
GO:0005813	centrosome	0.028432
GO:0000151	ubiquitin ligase complex	0.030462
GO:0001940	male pronucleus	0.036949
GO:0001939	female pronucleus	0.066226
GO:0005719	nuclear euchromatin	0.066226
GO:0070557	PCNA-p21 complex	0.067852
GO:0005721	centromeric heterochromatin	0.085319
GO:0032993	protein-DNA complex	0.152255
GO:0044430	cytoskeletal part	0.174874
GO:0005680	anaphase-promoting complex	0.177245
GO:0005876	spindle microtubule	0.245745
GO:0031967	organelle envelope	0.364413
GO:0008024	positive transcription elongation factor complex b	0.385183
GO:0033553	rDNA heterochromatin	0.385183
GO:0000779	condensed chromosome, centromeric region	0.477379
GO:0031975	envelope	0.487438
GO:0071141	SMAD protein complex	0.630571
GO:0000781	chromosome, telomeric region	0.635734
GO:0000803	sex chromosome	0.69737
GO:0005669	transcription factor TFIID complex	0.69737
GO:0045120	pronucleus	0.77927
GO:0071339	MLL1 complex	0.77927

Molecular Function

GO id	GO name	adjusted -P
GO:0005515	protein binding	2.29E-84
GO:0008134	transcription factor binding	4.56E-57
GO:0044212	transcription regulatory region DNA binding	5.55E-40
GO:0000975	regulatory region DNA binding	1.59E-38
GO:0001067	regulatory region nucleic acid binding	1.59E-38
GO:0003677	DNA binding	2.31E-37
GO:0005488	binding	1.88E-35
GO:0019899	enzyme binding	3.00E-35
GO:0000988	protein binding transcription factor activity	1.76E-31
GO:0003712	transcription cofactor activity	6.78E-30
GO:0000989	transcription factor binding transcription factor activity	1.63E-29
GO:0003700	sequence-specific DNA binding transcription factor activity	1.22E-28
GO:0001071	nucleic acid binding transcription factor activity	1.40E-28
GO:0003682	chromatin binding	5.09E-23
GO:0003676	nucleic acid binding	1.52E-21
GO:0003713	transcription coactivator activity	9.60E-20
GO:0019901	protein kinase binding	2.18E-18
GO:0043565	sequence-specific DNA binding	1.53E-17
GO:0000981	sequence-specific DNA binding RNA polymerase II transcription factor activity	2.72E-17
GO:0019900	kinase binding	1.24E-16

GO:0001085	RNA polymerase II transcription factor binding	1.41E-16
GO:0070491	repressing transcription factor binding	3.83E-16
GO:0003674	molecular_function	3.42E-13
GO:0003714	transcription corepressor activity	5.14E-13
GO:0042802	identical protein binding	9.08E-13
GO:0042393	histone binding	2.97E-12
GO:0002039	p53 binding	5.82E-12
GO:0035258	steroid hormone receptor binding	8.44E-12
GO:0035257	nuclear hormone receptor binding	1.80E-11
GO:0051427	hormone receptor binding	2.35E-10
GO:0046983	protein dimerization activity	2.74E-10
GO:0035326	enhancer binding	1.09E-08
GO:0003705	sequence-specific enhancer binding RNA polymerase II transcription factor activity	1.28E-08
GO:0001047	core promoter binding	1.83E-08
GO:0042826	histone deacetylase binding	2.35E-08
GO:0019904	protein domain specific binding	2.44E-08
GO:0031625	ubiquitin protein ligase binding	3.05E-08
GO:0050681	androgen receptor binding	1.01E-07
GO:0004693	cyclin-dependent protein kinase activity	1.22E-07
GO:0001158	enhancer sequence-specific DNA binding	2.10E-07
GO:0000976	transcription regulatory region sequence-specific DNA binding	2.99E-07
GO:0001012	RNA polymerase II regulatory region DNA binding	4.12E-07
GO:0043566	structure-specific DNA binding	5.03E-07
GO:0035064	methylated histone residue binding	2.38E-06
GO:0046332	SMAD binding	4.13E-06
GO:0042803	protein homodimerization activity	4.61E-06
GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding	5.27E-06
GO:0000982	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity	6.44E-06
GO:0035035	histone acetyltransferase binding	8.73E-06
GO:0001103	RNA polymerase II repressing transcription factor binding	1.10E-05
GO:0035497	cAMP response element binding	2.32E-05
GO:0046982	protein heterodimerization activity	5.53E-05
GO:0001077	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription	8.16E-05
GO:0043425	bHLH transcription factor binding	9.06E-05
GO:0000980	RNA polymerase II enhancer sequence-specific DNA binding	1.02E-04
GO:0004674	protein serine/threonine kinase activity	1.51E-04
GO:0030332	cyclin binding	1.72E-04
GO:0016538	cyclin-dependent protein kinase regulator activity	4.21E-04
GO:0008757	S-adenosylmethionine-dependent methyltransferase activity	5.67E-04
GO:0004672	protein kinase activity	6.06E-04
GO:0033613	activating transcription factor binding	6.67E-04
GO:0016773	phosphotransferase activity, alcohol group as acceptor	6.93E-04
GO:0005102	receptor binding	9.68E-04
GO:0042054	histone methyltransferase activity	0.001009
GO:0032129	histone deacetylase activity (H3-K9 specific)	0.001457
GO:0046969	NAD-dependent histone deacetylase activity (H3-K9 specific)	0.001457
GO:0008022	protein C-terminus binding	0.001498
GO:0004842	ubiquitin-protein ligase activity	0.001813
GO:0016301	kinase activity	0.002189
GO:0019207	kinase regulator activity	0.00224
GO:0017136	NAD-dependent histone deacetylase activity	0.002608
GO:0034979	NAD-dependent protein deacetylase activity	0.002608
GO:0001078	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in negative regulation of transcription	0.002616
GO:0019787	small conjugating protein ligase activity	0.0035
GO:0003690	double-stranded DNA binding	0.003949
GO:0000987	core promoter proximal region sequence-specific DNA binding	0.004307
GO:0008170	N-methyltransferase activity	0.00462
GO:0001159	core promoter proximal region DNA binding	0.004655
GO:0018024	histone-lysine N-methyltransferase activity	0.005857
GO:0030957	Tat protein binding	0.006419
GO:0035173	histone kinase activity	0.006735
GO:0016278	lysine N-methyltransferase activity	0.008395
GO:0016279	protein-lysine N-methyltransferase activity	0.008395
GO:0008276	protein methyltransferase activity	0.010676
GO:0001106	RNA polymerase II transcription corepressor activity	0.02002
GO:0070888	E-box binding	0.020693
GO:0016772	transferase activity, transferring phosphorus-containing groups	0.025626
GO:0008301	DNA bending activity	0.03143
GO:0042974	retinoic acid receptor binding	0.046015
GO:0008094	DNA-dependent ATPase activity	0.051185

GO:0001102	RNA polymerase II activating transcription factor binding	0.05239
GO:0019887	protein kinase regulator activity	0.055265
GO:0016881	acid-amino acid ligase activity	0.060825
GO:0031078	histone deacetylase activity (H3-K14 specific)	0.066226
GO:0032041	NAD-dependent histone deacetylase activity (H3-K14 specific)	0.066226
GO:0046970	NAD-dependent histone deacetylase activity (H4-K16 specific)	0.066226
GO:0031493	nucleosomal histone binding	0.067852
GO:0004197	cysteine-type endopeptidase activity	0.07861
GO:0046872	metal ion binding	0.081449
GO:0016740	transferase activity	0.107061
GO:0043169	cation binding	0.138419
GO:0043167	ion binding	0.145099
GO:0004861	cyclin-dependent protein kinase inhibitor activity	0.162818
GO:0008353	RNA polymerase II carboxy-terminal domain kinase activity	0.162818
GO:0003824	catalytic activity	0.180336
GO:0032452	histone demethylase activity	0.196156
GO:0008270	zinc ion binding	0.197044
GO:0070644	vitamin D response element binding	0.197492
GO:0001076	RNA polymerase II transcription factor binding transcription factor activity	0.21348
GO:0000979	RNA polymerase II core promoter sequence-specific DNA binding	0.245745
GO:0016879	ligase activity, forming carbon-nitrogen bonds	0.246898
GO:0004407	histone deacetylase activity	0.286002
GO:0030374	ligand-dependent nuclear receptor transcription coactivator activity	0.286002
GO:0030971	receptor tyrosine kinase binding	0.381996
GO:0033558	protein deacetylase activity	0.381996
GO:0034648	histone demethylase activity (H3-dimethyl-K4 specific)	0.385183
GO:0051525	NFAT protein binding	0.385183
GO:0004860	protein kinase inhibitor activity	0.407892
GO:0047485	protein N-terminus binding	0.438363
GO:0019210	kinase inhibitor activity	0.465332
GO:0008168	methyltransferase activity	0.615456
GO:0003886	DNA (cytosine-5-)-methyltransferase activity	0.630571
GO:0043125	ErbB-3 class receptor binding	0.630571
GO:0001046	core promoter sequence-specific DNA binding	0.635734
GO:0070412	R-SMAD binding	0.69737
GO:0032451	demethylase activity	0.77927
GO:0001104	RNA polymerase II transcription cofactor activity	0.79583
GO:0008234	cysteine-type peptidase activity	0.804387
GO:0009008	DNA-methyltransferase activity	0.930108
GO:0019213	deacetylase activity	0.991478

Table 14: Common GO term enrichment of Whi5 and Rb interactors

Biological Process

GO id	GO name	pRb	Whi5
GO:0060255	regulation of macromolecule metabolic process	6.99E-96	1.65E-07
GO:0019222	regulation of metabolic process	4.86E-93	3.11E-06
GO:0080090	regulation of primary metabolic process	1.27E-89	2.12E-05
GO:0006351	transcription, DNA-dependent	6.96E-89	0.003849
GO:0031323	regulation of cellular metabolic process	8.67E-89	2.59E-05
GO:0032774	RNA biosynthetic process	7.72E-88	0.003973
GO:2000112	regulation of cellular macromolecule biosynthetic process	3.47E-87	2.74E-04
GO:0010468	regulation of gene expression	7.26E-87	0.001819
GO:0006355	regulation of transcription, DNA-dependent	2.53E-86	4.56E-04
GO:0010556	regulation of macromolecule biosynthetic process	3.68E-86	2.94E-04
GO:0051252	regulation of RNA metabolic process	8.95E-85	8.74E-04
GO:0044260	cellular macromolecule metabolic process	1.09E-84	0.028798
GO:0031326	regulation of cellular biosynthetic process	1.60E-84	6.10E-04
GO:0009889	regulation of biosynthetic process	4.75E-84	6.10E-04
GO:0043170	macromolecule metabolic process	1.80E-82	0.024393
GO:0051171	regulation of nitrogen compound metabolic process	2.46E-82	3.42E-04
GO:0019219	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	7.55E-81	3.08E-04
GO:0090304	nucleic acid metabolic process	1.32E-78	0.290501
GO:0006366	transcription from RNA polymerase II promoter	6.17E-78	5.64E-04
GO:0050794	regulation of cellular process	2.19E-72	3.19E-06
GO:0048523	negative regulation of cellular process	8.10E-71	1.17E-06
GO:0050789	regulation of biological process	8.48E-69	1.48E-06
GO:0006357	regulation of transcription from RNA polymerase II promoter	2.34E-68	0.002118
GO:0007049	cell cycle	2.56E-67	5.33E-08
GO:0048522	positive regulation of cellular process	2.84E-66	0.009554
GO:0048519	negative regulation of biological process	1.07E-65	1.04E-07
GO:0010604	positive regulation of macromolecule metabolic process	1.05E-64	0.02843
GO:0065007	biological regulation	2.13E-64	4.03E-06
GO:0031325	positive regulation of cellular metabolic process	2.45E-63	0.003248
GO:0009893	positive regulation of metabolic process	4.28E-63	0.001048
GO:0010605	negative regulation of macromolecule metabolic process	7.54E-63	0.01749
GO:0009892	negative regulation of metabolic process	8.31E-61	0.061735
GO:0048518	positive regulation of biological process	1.09E-60	0.003831
GO:0031324	negative regulation of cellular metabolic process	4.04E-60	0.151956
GO:0051726	regulation of cell cycle	2.68E-59	5.42E-11
GO:0010558	negative regulation of macromolecule biosynthetic process	2.78E-57	0.017832
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	6.01E-57	0.017832
GO:0009890	negative regulation of biosynthetic process	6.51E-57	0.041866
GO:0010629	negative regulation of gene expression	1.09E-56	0.053839
GO:0045892	negative regulation of transcription, DNA-dependent	3.22E-56	0.142866
GO:0031327	negative regulation of cellular biosynthetic process	4.15E-56	0.041866
GO:0051253	negative regulation of RNA metabolic process	5.70E-55	0.149992
GO:0045934	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	4.55E-54	0.01043
GO:0051172	negative regulation of nitrogen compound metabolic process	8.89E-54	0.01043
GO:0051254	positive regulation of RNA metabolic process	5.60E-52	0.916311
GO:0045935	positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	6.39E-52	0.377309
GO:0000278	mitotic cell cycle	1.22E-51	1.91E-04
GO:0051173	positive regulation of nitrogen compound metabolic process	2.38E-51	0.394074
GO:0031328	positive regulation of cellular biosynthetic process	9.40E-50	0.459444
GO:0009891	positive regulation of biosynthetic process	2.50E-49	0.459444
GO:0051276	chromosome organization	4.93E-48	4.90E-04
GO:0016568	chromatin modification	1.46E-47	0.110313
GO:0022402	cell cycle process	3.58E-46	1.13E-07
GO:0006325	chromatin organization	3.28E-45	0.001177
GO:0043412	macromolecule modification	2.02E-43	9.65E-06
GO:0051329	interphase of mitotic cell cycle	2.13E-43	3.59E-07
GO:0006464	protein modification process	3.03E-43	4.93E-08
GO:0051325	interphase	4.68E-43	4.71E-08
GO:0009987	cellular process	5.90E-40	0.836662
GO:0006996	organelle organization	5.93E-39	0.062529
GO:0022403	cell cycle phase	8.75E-39	1.32E-06
GO:0033554	cellular response to stress	8.93E-38	0.006262

GO:0045786	negative regulation of cell cycle	9.74E-38	1.96E-05
GO:0016043	cellular component organization	9.38E-37	0.482921
GO:0006950	response to stress	1.82E-36	5.99E-04
GO:0051716	cellular response to stimulus	9.30E-36	0.08005
GO:0044267	cellular protein metabolic process	1.42E-35	0.103682
GO:0019538	protein metabolic process	6.31E-34	0.358501
GO:0016569	covalent chromatin modification	2.03E-33	0.213116
GO:0007050	cell cycle arrest	2.33E-33	1.47E-04
GO:0065009	regulation of molecular function	6.39E-33	0.908991
GO:0006974	response to DNA damage stimulus	1.06E-32	0.047288
GO:0016570	histone modification	2.93E-32	0.213116
GO:0010564	regulation of cell cycle process	7.08E-32	1.34E-06
GO:0050896	response to stimulus	5.10E-31	0.008357
GO:0071842	cellular component organization at cellular level	1.05E-30	0.865325
GO:0071156	regulation of cell cycle arrest	5.42E-30	6.12E-04
GO:0000082	G1/S transition of mitotic cell cycle	2.80E-27	6.69E-07
GO:0031399	regulation of protein modification process	3.14E-26	0.001156
GO:0007346	regulation of mitotic cell cycle	1.31E-25	0.004494
GO:0006259	DNA metabolic process	8.67E-25	0.008619
GO:0000075	cell cycle checkpoint	3.71E-23	6.12E-04
GO:0006793	phosphorus metabolic process	5.88E-21	2.95E-07
GO:0006796	phosphate metabolic process	5.88E-21	2.95E-07
GO:0016310	phosphorylation	1.17E-20	2.36E-06
GO:0006468	protein phosphorylation	1.10E-19	1.46E-09
GO:2000602	regulation of interphase of mitotic cell cycle	1.87E-18	0.349616
GO:0007093	mitotic cell cycle checkpoint	1.13E-17	0.151097
GO:0006260	DNA replication	9.94E-15	0.003514
GO:0042325	regulation of phosphorylation	1.56E-14	0.003133
GO:0051338	regulation of transferase activity	1.68E-14	0.011437
GO:0019220	regulation of phosphate metabolic process	2.54E-13	0.002528
GO:0051174	regulation of phosphorus metabolic process	2.54E-13	0.002528
GO:0043549	regulation of kinase activity	6.00E-13	0.008581
GO:0006281	DNA repair	1.16E-12	0.733641
GO:0001932	regulation of protein phosphorylation	9.39E-12	0.001728
GO:0090068	positive regulation of cell cycle process	6.15E-11	0.808535
GO:0045859	regulation of protein kinase activity	1.24E-10	0.005434
GO:0000079	regulation of cyclin-dependent protein kinase activity	1.46E-10	0.505837
GO:0000083	regulation of transcription involved in G1/S phase of mitotic cell cycle	3.10E-10	2.67E-07
GO:0051052	regulation of DNA metabolic process	1.93E-09	0.745754
GO:0071900	regulation of protein serine/threonine kinase activity	8.83E-08	0.001378
GO:0000279	M phase	3.04E-07	0.125522
GO:0010948	negative regulation of cell cycle process	3.74E-06	0.557776
GO:0040029	regulation of gene expression, epigenetic	4.42E-06	0.086266
GO:0031331	positive regulation of cellular catabolic process	4.09E-05	0.364043
GO:0006261	DNA-dependent DNA replication	6.84E-05	0.001291
GO:0045814	negative regulation of gene expression, epigenetic	1.06E-04	0.086266
GO:0006275	regulation of DNA replication	0.001218	0.114594
GO:0016458	gene silencing	0.003469	0.114555
GO:0071901	negative regulation of protein serine/threonine kinase activity	0.064917	0.993544
GO:0071824	protein-DNA complex subunit organization	0.065016	0.415412
GO:0006342	chromatin silencing	0.07391	0.086266

Cellular Component

GO id	GO name	pRb	Whi5
GO:0005634	nucleus	1.25E-102	0.012589
GO:0043231	intracellular membrane-bounded organelle	4.22E-72	0.253857
GO:0043227	membrane-bounded organelle	7.95E-72	0.253857
GO:0043229	intracellular organelle	6.35E-62	0.138068
GO:0043226	organelle	1.14E-61	0.138068
GO:0044424	intracellular part	7.73E-59	0.458662
GO:0005622	intracellular	1.01E-49	0.626387
GO:0044427	chromosomal part	1.24E-42	0.179654
GO:0005694	chromosome	3.40E-42	0.094417
GO:0043234	protein complex	2.10E-36	0.148843
GO:0000228	nuclear chromosome	7.35E-34	0.478186
GO:0044454	nuclear chromosome part	1.37E-31	0.411481
GO:0000307	cyclin-dependent protein kinase holoenzyme complex	6.64E-14	0.04808

Molecular Function

GO id	GO name	pRb	Whi5
GO:0004674	protein serine/threonine kinase activity	1.51E-04	2.74E-04
GO:0016538	cyclin-dependent protein kinase regulator activity	4.21E-04	0.422036
GO:0004672	protein kinase activity	6.06E-04	7.79E-05
GO:0016773	phosphotransferase activity, alcohol group as acceptor	6.93E-04	0.011603
GO:0032129	histone deacetylase activity (H3-K9 specific)	0.001457	0.37453
GO:0046969	NAD-dependent histone deacetylase activity (H3-K9 specific)	0.001457	0.37453
GO:0016301	kinase activity	0.002189	0.01315
GO:0019207	kinase regulator activity	0.00224	0.151097
GO:0017136	NAD-dependent histone deacetylase activity	0.002608	0.993544
GO:0034979	NAD-dependent protein deacetylase activity	0.002608	0.993544
GO:0016772	transferase activity, transferring phosphorus-containing groups	0.025626	0.339772
GO:0019887	protein kinase regulator activity	0.055265	0.114594
GO:0031078	histone deacetylase activity (H3-K14 specific)	0.066226	0.37453
GO:0032041	NAD-dependent histone deacetylase activity (H3-K14 specific)	0.066226	0.37453
GO:0046970	NAD-dependent histone deacetylase activity (H4-K16 specific)	0.066226	0.37453
GO:0004407	histone deacetylase activity	0.286002	0.482921
GO:0033558	protein deacetylase activity	0.381996	0.482921

APPENDIX C

Fig.1. A Phylogenetic tree for Whi5 homologues

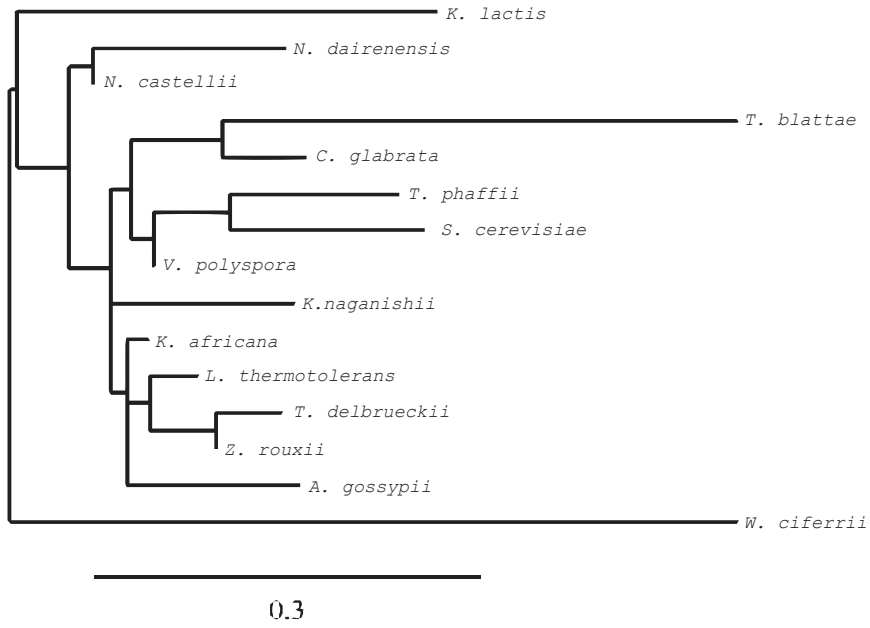


Fig.1. B Phylogenetic tree for Rb homologues

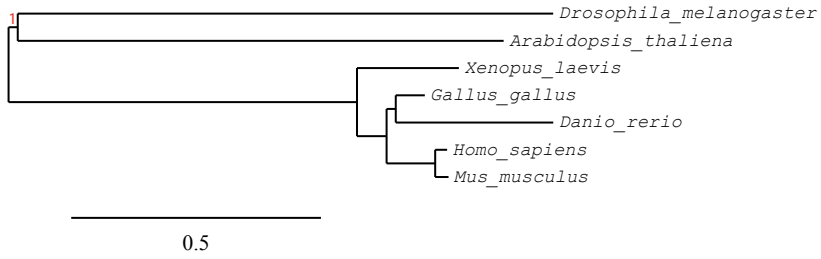
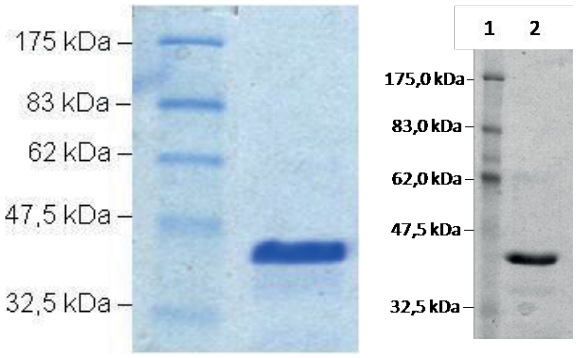
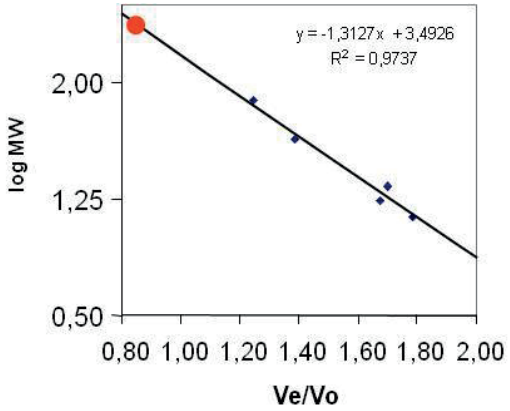


Fig.2. Experimental Data of Whi5 protein

A



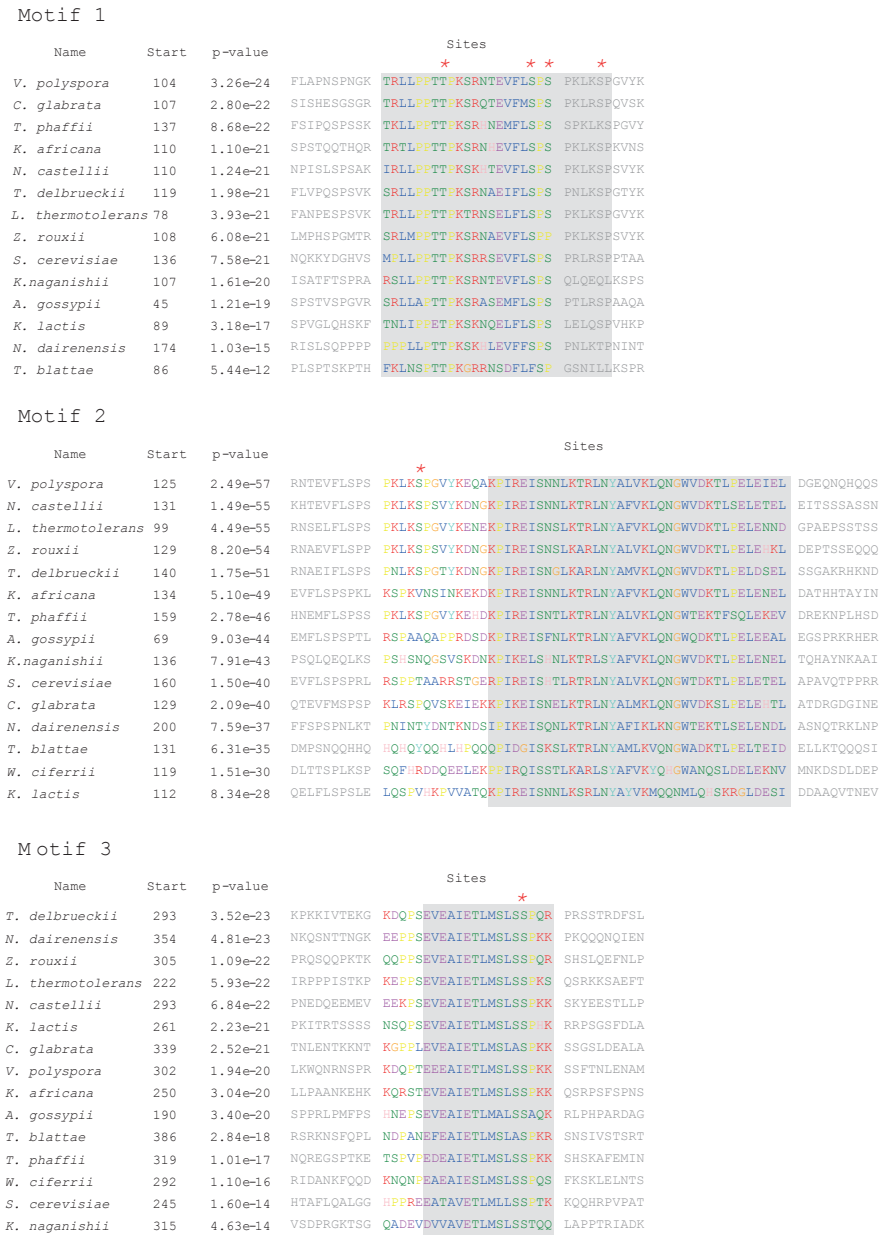
B



Analytical size-exclusion chromatogram of pure, recombinant Whi5. Calibration curve was obtained with the following globular proteins: transferrin (80kDa), BSA (66kDa), Ovalbumin (43 kDa), Chimotrypsin (23 kDa), Myoglobin (17 kDa) e Cytochrome C (13,6 kDa). SDS-PAGE analysis of a recombinant, IMAC-purified Whi5Sc produced in E. coli cells

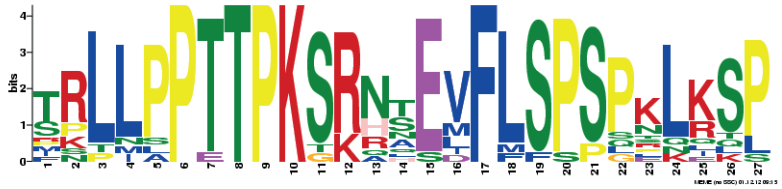
Fig.3. A. Motifs for 15 Whi5 related sequences: Shaded sequences are refined motifs and colored sequences are Meme motifs for 15 sequences. B. Motif logos for refined motifs

A

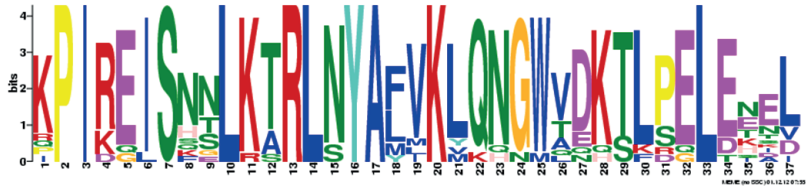


B

Motif 1

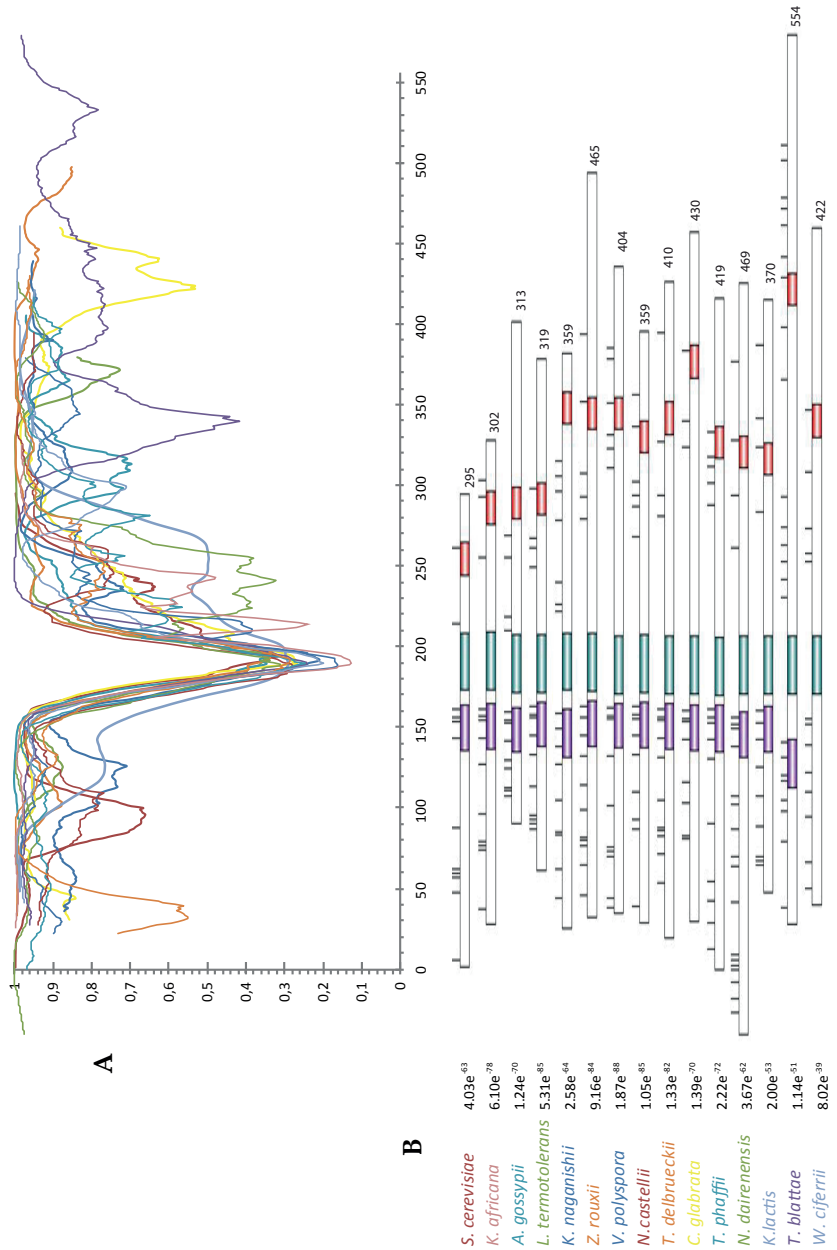


Motif 2



Motif 3

Fig.4 Conservation of structural disorder among Whi5 homologues in Fungi. (refined motifs)



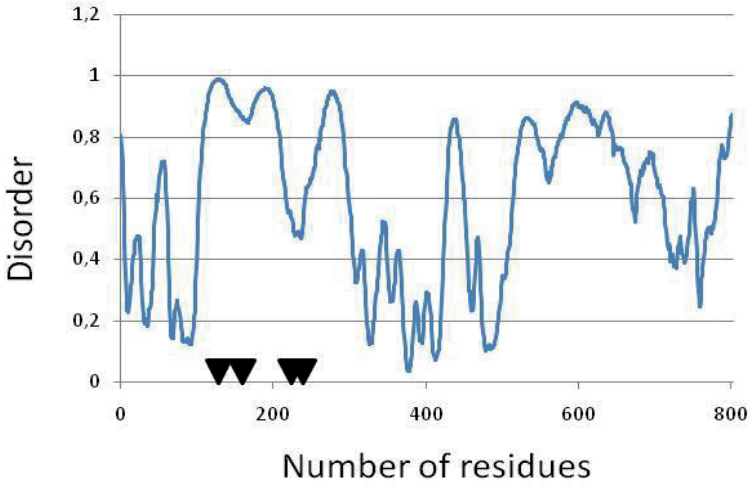


Fig.5 Swi6 disorder plot.

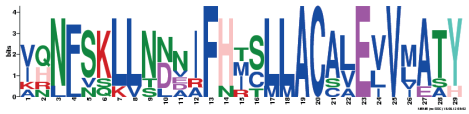
Fig.6 A. Motifs for Rb related sequences B. Motif logos for motifs

B

Motif 1

Motif 2

Motif 3

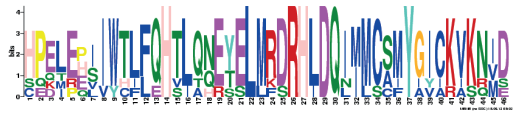


Motif 4

Motif 5

Motif 6

Motif 7



Motif 8

Motif 9

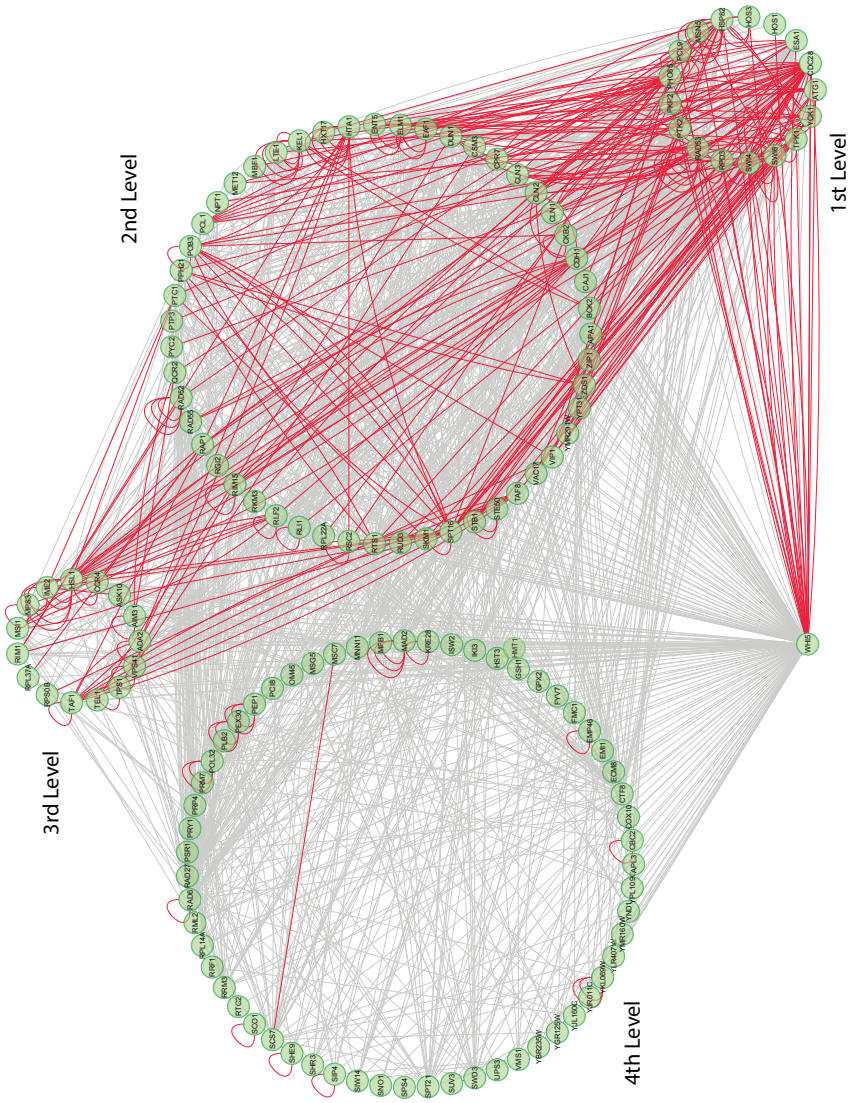


Fig 7. Interaction network of Whi5.
The redline and the grey line represent physical and genetic interactions respectively. 1st level proteins directly interact with Whi5, 2nd level proteins interact with 1st level proteins and 3rd level proteins interact with 2nd level proteins. 4th level proteins are genetically interact with all other proteins. Total Interactors (Nodes) are 143 and Total Edges are 1547.

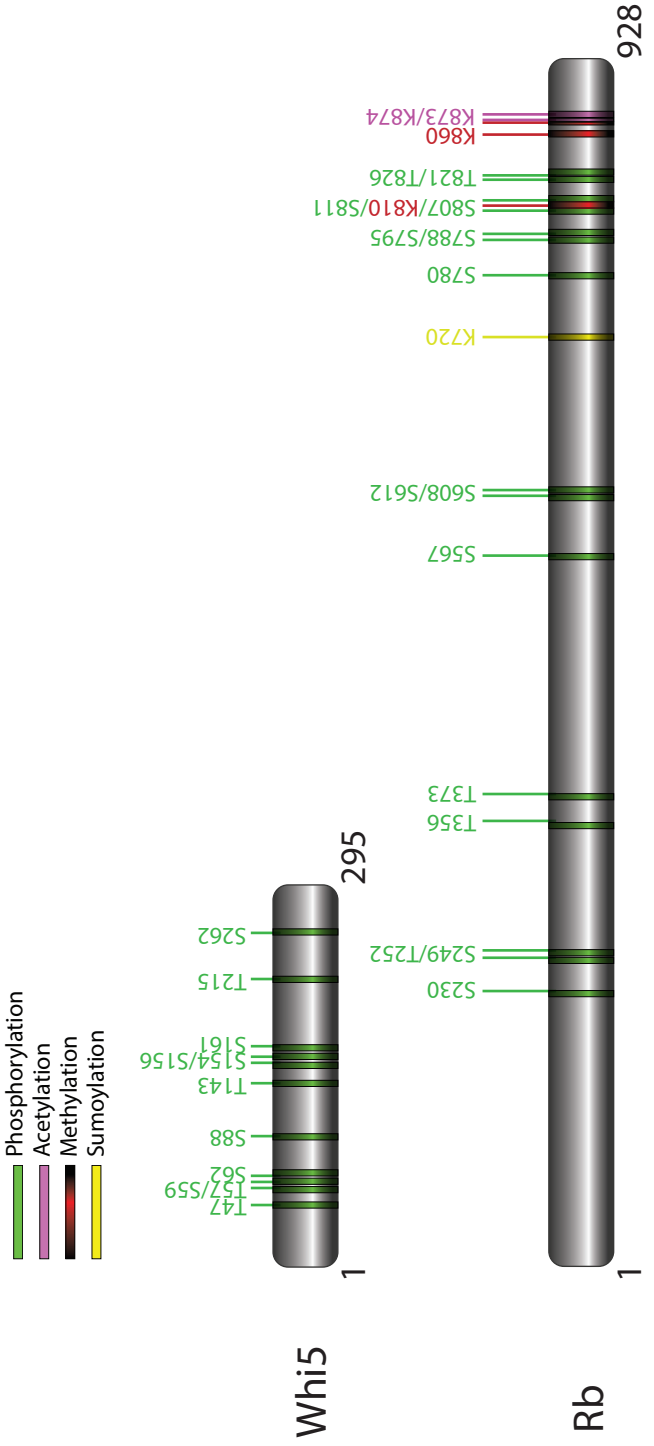


Fig 8. Protein modification of Whi5 and Rb.

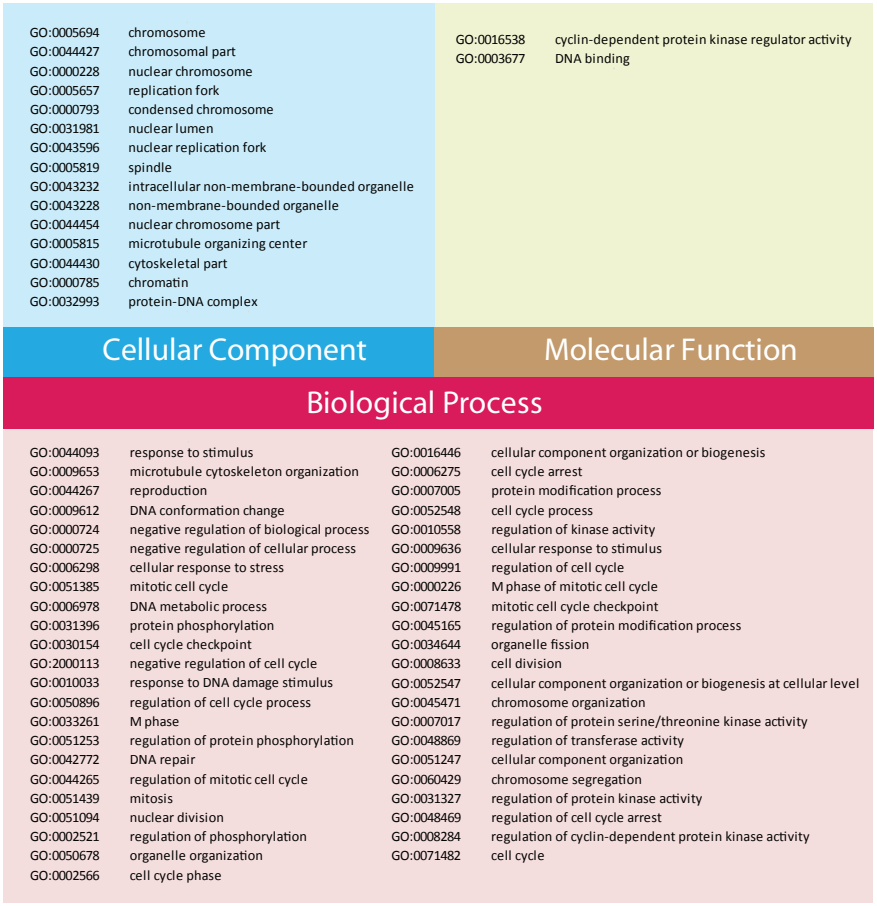


Fig 9. Common GO terms enrichment of E2F and SBFTarget genes.

APPENDIX D

List kinases of DNA Replication Initiation

Budding yeast

Proteins	Components	Target Known Proteins in DNA Replication Initiation	Known role of Phosphorylation
G1-CDK	Cln1,2 and Cdc28	Sic1	Multisite phosphorylation inactivate the protein [1-5]
S-CDK	Clb5,6 and Cdc28	ORC	Prevent relicensing[6, 7]
		CDC6	Promote degradation[7, 8]
		MCM2-7	Activate the complex[7]
		Sld2	Multisite phosphorylation helps to bind to Dpb11 and form Pre-LC[7, 9, 10]
		Sld3	Helps to bind to CDC45 and Dpb11[7, 9, 10]
DDK	CDC7-DBF4	MCM2-7	Activate the complex[7, 11-14]

Mammalian system

Proteins	Components	Target Known Proteins	Known role of Phosphorylation
CycD-CDK4	Cyclin D (D1, D2, and D3) and CDK4	pRb P107 P130	Modify the protein and binding pattern with E2F proteins [15]
CycD-CDK6	Cyclin D (D1, D2, and D3) and CDK6	pRb, P107, P103	Similar function like CycD-CDK4[15]
CycE-CDK2	Cyclin E (E1 and E2) and CDK2	RB1, P107 P103	Modify the protein and binding pattern with E2F proteins[15]
		CDC6	Prevent from degradation[16]
		MCM2-7	Activate the complex[7]
		RecQl4 (CycA/Cyc E?)	Unknown function [17]
		TopBP1 (CycA/Cyc E?)	Unknown function [18]
		Treslin	Promote formation of a Treslin-TopBP1 Complex [19]
		GEMC1	Promote affinity for TopBP1 and stimulates Cdc45 binding. [20, 21]
		DUE-B	Promote assembly of a TopBP1–DUE-B–Cdc45 complex. [22]
		CDT1 (CycA/Cyc E?)	Degradation of Cdt1 in S phase[23]
CycA-CDK2	Cyclin A (A1 and A2) and CDK2	E2F	Disrupts the formation of E2F1–DP1 heterodimers and reduces their DNA-binding properties [29]
		ORC1	Promote Degradation [24]
		CDC6	Exclude from nucleus[16]
DDK	CDC7-DBF4	MCM2-7	Activate the complex[7]

APPENDIX E

Tools and Web service

Name	URL
BioGRID	www.thebiogrid.org, version 3.2
BioModels	http://www.ebi.ac.uk/biomodels-main/
Cell Illustrator	http://www.cellillustrator.com/
CellDesigner	http://www.celldesigner.org/
ClustalW2	http://www.ebi.ac.uk/Tools/msa/clustalw2/
Composition Profiler	(http://www.cprofiler.org/cgi-bin/profiler.cgi/)
Cytoscape 2.8	(http://www.cytoscape.org/)
Disprot	
EndNote Web 3.5	https://www.myendnoteweb.com/EndNoteWeb.html
FoldIndex	(http://bip.weizmann.ac.il/fldbin.findex)
GPS2.1 (Group-based Prediction System, ver 2.1)	http://gps.biocuckoo.org/
iRefWeb	(http://wodaklab.org/iRefWeb/)
NetPhosYeast	http://www.cbs.dtu.dk/services/NetPhosYeast/
Pfam	(http://pfam.sanger.ac.uk/)
PONDR	(www.PONDR.com, Molecular Kinetics)
PPSP	(http://ppsp.biocuckoo.org/index.php
PRALINE	http://www.ibi.vu.nl/programs/pralinewww/
Proteus	(http://wks16338.biology.ualberta.ca/proteus),
ProtParam	(http://web.expasy.org/protparam/)
Pubmed	http://www.ncbi.nlm.nih.gov/pubmed
Revigo	http://revigo.irb.hr/
rxncon	http://rxncon.org/
UniProt	http://www.uniprot.org/
VSL2	(http://www.ist.temple.edu/disprot/Predictors)