Robin Mjelle

Multiple layers of gene regulation: the cell cycle, epigenetics and microRNAs

Thesis for the degree of Philosophiae Doctor

Trondheim, March 2015

Norwegian University of Science and Technology
Faculty of Medicine
Department of Cancer Research and Molecular Medicine

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Medicine
Department of Cancer Research and Molecular Medicine

ISBN 978-82-326-0804-1 (printed ver.)
ISBN 978-82-326-0805-8 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2015:73

Printed by NTNU Grafisk senter

Norges teknisk-naturvitenskapelige universitet

Det medisinske fakultet

Robin Mjelle

Multiple nivåer av genregulering: cellesylus, epigenetikk og mikroRNA

Hver celle i kroppen inneholder arvestoffet DNA. Hver gang en celle deler seg vil arvestoffet bli kopiert og videreført til neste celle. Slik vil alle celler til enhver tid inneholde den informasjonen som skal til for å vedlikeholde og opprettholde cellens funksjon. For å forsikre seg om at alt DNA blir kopiert uten feil har cellen bygd opp et reguleringssystem som sjekker for feil i DNA hver gang en celle deler seg. Prosessen for celledeling kalles cellesyklus og reguleringen av denne kalles cellesyklus-regulering. Reguleringen er basert på ulike kontrollpunkt i cellesyklus, kalt «checkpoints», som kontrolleres av bestemte proteiner som er ansvarlig for at DNA er riktig kopiert og at forholdene ligger til rette for celledeling. Arvematerialet er stadig utsatt for skade som repareres kontinuerlig av reparasjons-proteiner i cellen. Disse reparasjonsproteinene er spesielt aktive under celledeling for å hindre at feil i DNA blir nedarvet til datterceller. Skader på DNA kan forekomme under selve delingsprosessen eller i andre faser av cellesyklus spontant eller av ytre faktorer som kjemikalier og stråling. Hvis skaden skjer i gener som kontrollerer celledeling vil cellens naturlige kontrollpunkt kunne miste sin funksjon, som igjen kan føre til ukontrollert celledeling og kreft.

Vi har i denne avhandlingen undersøkt hvilke gener som er aktive under celledeling med særlig fokus på gener for proteiner som reparerer skadet DNA, såkalte DNA-reparasjonsgener. Vi har analysert ulike celletyper for å finne generelle reguleringsmønstre som er felles for alle celler. Ved å studere hvordan gener blir regulert under celledeling, vil man få et bedre grunnlag for å forstå eventuell feilregulering og kunne utvikle nye behandlingsmetoder mot kreft. Det har i senere tid vært et økende fokus på utnyttelse av cellenes reparasjonsmaskineri for målrettet terapi mot kreftceller. Kreftceller er avhengige av reparasjon av skadet DNA for å overleve og gjennom å eliminere sentrale reparasjonsprotein vil kreftcellene kunne miste evnen til å dele seg. Vi har i artikkel I vist at gener som vanligvis er avslått ved hjelp av epigenetiske markører, kan transkriberes samtidig som DNA replikeres i cellesyklus. Replikasjonsavhengig transkripsjon er ikke tidligere vist for denne typen gener og kan være en viktig mekanisme for å videreføre de epigenetiske markørene på disse genene. Vi har i artikkel II i avhandlingen vist at DNA reparasjonsgener ofte endrer uttrykk gjennom cellesyklus og at mange av disse genene er oppregulert i samme cellesyklus-fase som nytt DNA blir syntetisert, den såkalte S fasen. Dette tyder på at reparasjonsgenene har en sentral rolle i å regulere prosessene rundt DNA syntese og at gener som inngår i samme reparasjonsprosess ofte er uttrykt samtidig. Videre viser vi at gener som koder for proteiner som hjelper å åpne protein-dekket DNA, såkalte kromatin remodeleringsgener, er oppregulert i S fase av cellesyklus. Dette tyder på at disse genene er viktige for celledeling ved å tilrettelegge for proteiner som er avhengige av åpen DNA struktur.

I 1993 ble det oppdaget en klasse små RNA molekyl som senere skulle vise seg å være sentrale i reguleringen av aktiviteten og uttrykksnivået til gener. Disse små RNA molekylene er nå kjent som mikroRNA (miRNA) og er assosiert med en rekke sykdommer, deriblant kreft. Flere studier har vist at miRNA er feilregulert i kreftceller, noe som kan resultere i endret utrykk av gener som er regulert av

miRNA. Artikkel III og IV i avhandlingen er dedikert til proteinet Argonaute 2 (Ago2) hvis funksjon er å regulere utrykket til gener ved å binde til konserverte nukleotid-sekvenser i mål-mRNA og dermed bidra til degradering av mRNA eller hindre at mRNA blir translatert til protein. For å finne mål-mRNA i cellen binder Ago2 seg til miRNA som fungerer som guide. I artikkel IV undersøker vi om spesifikke miRNA har økt affinitet for Ago2 sammenliknet med de andre tre Ago-proteinene. Vi viser at struktur og lengde på miRNA kan avgjøre hvilke av de fire Ago-proteinene et gitt miRNA binder seg til. Dette er avgjørende for reguleringen siden ulike miRNA har forskjellige bindingssekvenser i mål-mRNA samtidig som Ago2 er funksjonelt forskjellig fra de andre Ago proteinene. I artikkel III viser vi at Ago2 er viktig i reguleringen av en gruppe relaterte gener kalt Hox-gener. Hox-genene er essensielle i tidlig embryonalutvikling for posisjonering av embryonale strukturelementer som armer, bein og øyne. Vi viser at embryonale museceller som ikke inneholder Ago2, har høyere uttrykk av flere av Hox genene, noe som tyder på at disse genene er regulert av miRNA, eller at Ago2 indirekte bidrar til reguleringen av disse genene.

Kandidat: Robin Mjelle

Institutt: Institutt for Kreftforskning og Molekylærmedisin

Hovedveileder: Hans E. Krokan

Biveileder: Pål Sætrom

Finansiering: Norges forskningsråd og Svanhild og Arne Must's Fond for Medisinsk Forskning

Ovennevnte avhandling er funnet verdig til å forsvares offentlig for graden

PhD i Molekylærmedisin.

Disputas finner sted i

Auditoriet MTA, Medisinsk-teknisk forskningssenter,

Tirsdag 14. April 2015, kl. 12:15.

1. Acknowledgements

First and foremost I would like to thank my two supervisors Hans E. Krokan and Pål Sætrom for their contribution to the work presented here, and to the projects that did not make it in the thesis. Hans has been a great leader and taught me much about research and being independent in my work. I appreciate all the stories you tell, about discoveries you made decades ago and persons you have met over the years, it reminds me of how talented researcher you are. I am grateful to Pål who took me in as a master student in 2009. You have been a great inspiration, and the main reason why I wanted to continue the academic path. Your passion and enthusiasm is a true motivation for people working with you. I hope we will make many interesting discoveries together in the time to come.

I would like to thank my brother Kenneth, who got me interested in biology and taught me how to sequence DNA way before the high-throughput techniques came around. You always walked the path before me, which made my choices so much easier. I often wish we lived in the same city.

I would like to thank my girlfriend Kristin for the wonderful times we have together. You make me forget the worries that I continuously make up, and remind me of the important things in life. I look forward to future trips and adventures together. We have a lot of fun left to do.

Finally I would like to thank my parents for supporting me and making me feel safe. You are important motivation for me to push forward with the things I like.

Contents

2. List of papers

Paper I:

Transcription profiling during the cell cycle shows that a subset of Polycomb-targeted genes is upregulated during DNA replication

Javier Peña-Diaz, Siv A. Hegre, Endre Anderssen, Per A. Aas, Robin Mjelle, Gregor D. Gilfillan, Robert Lyle, Finn Drabløs, Hans E. Krokan and Pål Sætrom

Nucleic Acids Res, 2013. 41(5): p. 2846-56.

Paper II:

Cell cycle regulation of human DNA repair and chromatin remodeling genes

Robin Mjelle, Siv Anita Hegre, Per Arne Aas, Geir Slupphaug, Finn Drabløs, Pål Sætrom and Hans E. Krokan

Submitted 2014

Paper III:

Argonaute2 affects Hox gene expression and the late growth response in mouse embryonic fibroblasts

Marie Lundbæk, Robin Mjelle and Pål Sætrom

Submitted 2014

Paper IV:

Deep sequencing of Ago2 knock-out cells reveals distinct sequence and structure features associated with miRNA sorting.

Robin Mjelle, Marie Lundbæk, Jan-Preben Mossin and Pål Sætrom

Manuscript

3. Abbreviations

Ago	Argonaute
APC	Anaphase promoting complex
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3-related protein
BER	Base excision repair
BIR	Break-induced replication
CDK	Cyclin-dependent kinase
CKI	Cyclin-dependent kinase inhibitor
CLL	Chronic lymphocytic leukemia
CPD	Cyclobutane Pyrimidine Dimer
CPG	Cancer predisposition gene
DDR	DNA damage response
dHJ	Double Holliday junction
DSB	Double strand break
ES cells	Embryonic stem cells
FA	Fanconi Anemia
Fox protein	Forkhead box protein
G_1 phase	Growth phase 1 or Gap phase 1
G_2 phase	Growth phase 2 or Gap phase 2
GGR	Global genomic repair
H3	Histone 3
H3K27me3	Histone 3 lysine 27 tri-methylation
H3K4me3	Histone 3 lysine 4 tri-methylation
Hox genes	Homeotic genes
HR	Homologous recombination
KAP-1	KRAB-associated protein-1
lincRNAs	Long intergenic non-coding RNAs
lncRNAs	Long non-coding RNAs
MEF	Mouse embryonic fibroblast
miRNA	MicroRNA
MMEJ	Microhomology-mediated end joining
MMR	Mismatch repair
MMS	Methyl methanesulfonate
M phase	Mitosis or mitotic phase
mRNA	Messenger RNA
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
nt	Nucleotide
PAZ	Piwi Argonaute-zwille

PcG	Polycomb group
PCNA	Proliferating cell nuclear antigen
PCR	Transcriptional-repressive complexes
pri-miRNA	Primary miRNA
qRT-PCR	Quantitative real-time PCR
Rb	Retinoblastoma
RISC	RNA-induced silencing complex
S.cerevisiae	*Saccharomyces cerevisiae* (yeast)
SDSA	Synthesis-dependent strand annealing
SILAC	Stable isotope labeling by amino acids in cell culture
siRNA	Small interfering RNA
S phase	DNA synthesis phase or DNA synthetic phase
ssDNA	Single-stranded DNA
ssRNA	Single stranded RNA
SWI/SNF	SWItch/Sucrose NonFermentable
TCR	Transcription coupled repair
UTR	Untranslated region
XP	*Xeroderma pigmentosum*
YY1	Yin Yang 1

Preface

Eukaryotes have evolved a complex system of gene regulation compared to prokaryotes. In eukaryotes, protein expression is regulated at multiple levels in the cells, including regulation by modifications of histones and DNA, regulation of mRNA levels and post transcriptional modifications. This complexity is believed to be important in the generation and maintenance of the different cell types found in multicellular organisms (reviewed in (1)).

In prokaryotes, there is a strong correlation between genome size and gene number (2). In Eukaryotes, however, a large proportion of nuclear DNA is non-coding. The non-coding parts of the genome have been subject to increased interest in the last two decades and many important discoveries have been made with regards to the role of non-coding RNAs in gene regulation. One type of non-coding RNA that is shown to play an essential role in gene regulation is a class of ~22 nucleotide long small RNAs called microRNAs (miRNAs). The first miRNA was discovered by Victor Ambros' lab in 1993 in the nematode *C. elegans* (3). At that time little was known about the importance of these small RNAs apart from their role in regulating the timing of development in the worm. In 2001, a much larger set of small RNAs was discovered by three individual laboratories (4-6), and the term miRNAs was established. In the years to follow, many laboratories started working on these small RNAs which have later been implicated in many biological processes and diseases. Today, small RNAs are used both as a tool for silencing genes and as potential biomarkers for diseases, and they are often included in studies of gene expression. Even though small RNA biology has been intensively studied for more than ten years, several central questions remain to be answered. In the current thesis we investigated the role of the miRNA binding protein Argonaute 2 (Ago2) by sequencing small RNAs and mRNA in Ago2 knock-out cells. We showed that Ago2 knock-out strongly affects *Hox* gene-expression as well as late growth responses in serum starved MEF cells. Moreover, by analyzing miRNAs from the same Ago2 knock-out cells, we characterized sequence and structure features in miRNAs differential expressed between knock-out and wild-type, indicating novel sorting characteristics for Ago2. We also included a meta-analysis of published data from deep-sequenced small RNAs extracted from protein immunoprecipitation (IP) of the different Ago proteins. This analysis confirmed the sorting features found in the Ago2 knock-out cells.

The field of chromatin remodeling has gained more interest in recent years although the first ATP-dependent chromatin remodeling complex, the yeast SWI/SNF, was discovered already

in in 1992 (7). The increased interest is partly due to advances in technology resulting in high resolution structure of the nucleosome as well as new sequencing methods. Several research groups have started to work on how remodeling enzymes corporate with DNA repair proteins to facilitate repair of damaged DNA. Adding to this complexity comes the discovery that small RNAs may have a role in mediating DNA damage response; however, the exact mechanism for this interaction is not known (8,9). In the current thesis we try to integrate DNA repair genes and chromatin remodeling genes, focusing on their role in the cell cycle. We analyzed cell cycle expression profiles from multiple cell lines and found that several DNA repair pathways and chromatin remodeling genes are enriched in specific cell cycle phases, indicating that the cell prefers certain repair pathways in specific cell cycle phases. Moreover, we provide an extensive update on DNA repair and chromatin remodeling genes. We also found that a subset of Polycomb-target genes display replication dependent transcription. This may be a mechanism for expressing Polycomb-associated RNAs, which perhaps contribute in maintaining the repressive state of these genes during cell division.

4. Introduction

4.1. Mechanisms and pathways of DNA repair

The maintenance of genome integrity is crucial for cell survival and to prevent mutations to be inherited. Cells have evolved several mechanisms to counteract the effects from agents that could inflict damage to the DNA. The DNA damage response (DDR) comprises a series of pathways that are distinct in terms of which lesion they process, but still share many of the same proteins (10). DNA repair pathway choice is dependent on damage type, but also to a large degree on cell cycle phase (11-13) and protein modifications (14,15). The following sections describe the major repair pathways, which proteins are involved and the type of lesion they target.

4.1.1. The base excision repair pathway

Base excision repair (BER) is a repair pathway that corrects small base lesions caused by oxidation, deamination and alkylation. The pathway is initiated by one of the many mammalian DNA glycosylases that recognize and remove the damaged base (reviewed in (16)). The pathway is further completed by a short patch or long patch repair route that uses a distinct set proteins to fill and ligate the gap generated by the DNA glycosylase. Short patch BER does not depend on replication and is equally efficient in proliferating and nonproliferating cells. The major proteins needed for short patch BER after the initiate repair by a DNA glycosylase are AP-endonuclease APE1, DNA polymerase β (polβ), DNA ligase 1 or 3 (LIG1/3), Polymerase 1 (PARP) and XRCC1. Long patch BER on the other hand depends on replication proteins and is mainly acting in proliferating cells. After the damaged strand is cleaved by APE1, long patch BER uses DNA polymerase δ and ε, PCNA, FEN1 and LIG1 to incorporate new nucleotides and repair the damage. The BER pathway is known to be up-regulated during S phase (17).

4.1.2. The nucleotide excision repair pathway

Nucleotide excision repair (NER) is the principal repair mechanism for helical disorders, replication-blocking DNA adducts such as those induced by UV light, and other chemicals toxic to the genome (17). In humans, defects in the NER-genes leads to increased sensitivity to UV radiation and the clinical disorder *Xeroderma pigmentosum* (XP) and increased risk of skin cancer (18). NER has two modes of repair: global genome repair (GGR) and

transcription coupled repair (TCR). The *Xeroderma pigmentosum* group A protein (XPA) is indispensable in both pathways and acts as an important regulator in verifying the damage and recruiting other factors to the damage site (19,20). TCR is a specialized pathway that removes lesions on the transcribed strand of DNA that block RNA polymerase (21). GGR and TCR differ mostly in the initial steps of repair. XPC-RAD23B recognizes structural changes in the DNA and initiates GGR by binding the strand opposite of the lesion. TCR is initiated by stalled RNA polymerase on the transcribed strand (22,23). Stalled polymerase represents a signal that recruits the two proteins CSA and CSB, which facilitate removal of the damage and restart of transcription (24,25). After the initial damage recognition by either GGR or TCR the transcription factor II H (TFIIH) complex is recruited. TFIIH is required for the unwinding of the DNA around the lesion facilitating the recruitment of XPA and replication protein A (RPA) and proteins from the 10/11-subunit complex. RPA activates XPG and ERCC1, which cleave the damaged DNA fragment. DNA polymerases δ, ε, and κ fill the single strand gap, a process which is coordinated by PCNA (26). DNA Ligase3-XRCC1 and DNA Ligase I finally seal the DNA (26). In non-dividing cells this is performed by XRCC1–LIG3α (27). The NER pathway is known to be important in G_1 phase, but it also function at a global level irrespective of cell cycle phase (17,28).

4.1.3. The mismatch repair pathway

Mismatch repair (MMR) is a repair pathway specific for base-base mismatches and insertions/deletions loops generated during DNA replication and recombination. Some studies have reported constant levels of MMR-genes throughout the cell cycle (29), whereas other have reported that the pathway mainly functions during S phase (30). Some of the genes that are associated with MMR are: RFC, RPA, ExoI, MSH2, MSH3, MSH6, MutSα/β and MutLα. MutSα recognizes short base-base mismatches of 1-2 nucleotides, while MutSβ recognizes larger mismatches. It has been suggested that PCNA is helping MutSα and MutSβ to recognize mispairs in newly replicated DNA, in addition to its role in DNA synthesis (31-33). RFC is required for 3' directed mismatch excision during repair and facilitates loading of PCNA (33). RPA is a DNA-binding protein that seems to be involved in all stages of MMR (34). ExoI is an exonuclease that interacts with MSH2 and MLH1 to perform 5'→3' directed repair (35,36).

4.1.4. The fanconi Anemia pathway

Fanconi Anemia (FA) is a genetic disorder caused by mutations in genes responsible in the repair of interstrand crosslinks. The Fanconi anemia pathway involves more than 15 genes and its main function is to repair interstrand crosslinks (37). The fanconi proteins coordinate multiple repair processes such as translesion synthesis, nuclease activity, homologous recombination, non-homologous end joining and mismatch repair. Crosslinks in DNA can occur both during transcription and replication and repair is therefore needed in all stages of the cell cycle. When no homologous template is present the damage can be repaired using nucleotide excision repair (38). If a template is present homologous recombination is used to remove the crosslink (39). The fanconi pathway is activated through phosphorylation of ATR which together with its partner ATRIP (ATR interacting protein) coordinate the downstream activation of the fanconi core complex (40). The core complex activates the FANCI/FANCD2 (I-D2) complex which in turn recruits FAN1 (41). FANCD2 is found to be phosphorylated by ATM, an essential step in establishing the intra-S checkpoint (40). Together with the I-D2 complex, FAN1 localizes at sites of DNA damage. The I-D2 complex also activates BRCA2 and FANCN which loads the homologous recombination protein RAD51 onto the DNA (37).

4.1.5. DNA double-strand repair

DNA double-strand break (DSB) is a severe type of lesion that can lead to oncogenic translocations. Three mechanisms exist to repair DSBs: non-homologous end joining (NHEJ) (42); microhomology-mediated end joining (MMEJ) (43); and homologous recombination (HR) (44). The MRN complex consisting of Mre11, Rad50 and Nbs1 plays an important role in the initial processing of DSBs prior to repair by the more specific pathways. The MRN complex activates the checkpoint kinase ATM in response to DSBs. The decision on which pathway to choose depends on which cell cycle phase the damage occur and on which factors that are able to access the free DNA ends (45). The process of creating 3′ single-stranded DNA (ssDNA) tails, called resection, is a critical determinant for repair pathway choice (45,46). Once resection is initiated, the NHEJ-proteins XRCC5 and XRCC6 will not bind and the cell is committed to HR. NHEJ is shown to be active throughout the cell cycle but the activity increases from late G_1 to G_2/M phase (47). DSBs that occur during chromosome segregation in M phase are likely to be repaired by NHEJ (48). Because homologous recombination prefers sister chromatids as template for repair it is suggested that

HR is restricted to the S and G_2 phases of the cell cycle when sister chromatids are present (44,49,50). During G_2/M phase the chromatin is highly condensed which makes homology search by HR difficult. However, HR can function during G_2 if the sister chromatids are in proximity to one another. Homologous recombination is divided into three stages: presynapsis, synapsis and postsynapsis. In brief, during presynapsis the DNA is prepared for repair by forming a region of single-stranded DNA (ssDNA) which is bound by the ssDNA-binding protein RPA (replication protein A). During synapsis, Rad51 performs homology search on the RPA-covered ssDNA (44). Finally, in postsynapsis the HR pathway is separated into three sub-pathways (BIR, SDSA and dHj) with separate acting proteins. BRCA2, BLM, GEN1 and EME1 are important proteins in this final step of the HR pathway (51,52).

4.1.6. Direct DNA damage reversal

In addition to the abovementioned repair systems, the cell can repair lesions without the incision of the DNA sugar-phosphate backbone or removal of the damaged base. This mechanism is called direct damage reversal and is essentially error-free, however, it has a narrower substrate range than other repair pathways (53,54). Three types of direct DNA repair have been described: i) removal of UV light-induced photo lesions by photolyases; ii) removal of O^6- methyl guanine (6-meG) by O^6- alkylguanin-DNA alkyltransferases (AGTs); iii) removal of *N*-alkylated base adducts by AlkB family dioxygenases.

UV lesion are generally of two types: cyclobutane pyrimidine dimers (CPDs) and pyrimidine pyrimidones (6-4) photoproducts (6-4 PPs) (53). The photolyase enzymes are specifically targeting either of the two lesions and they use light to catalyze the repair process (53,55,56). Alkylation damage could occur as a result of exposure to a variety of environmental compounds including methyl-N'-nitro-N-nitroso-guanidine (MNNG), N-methyl-N-nitrosurea (MNU) and methyl methane-sulfonate (MMS). They could also occur from endogenous compounds like S- adenosylmethionine (SAM) (54). These compounds can create O-alkylated and N-alkylated adducts which is toxic to the cell (54). Although lesions of these types could be repaired by the BER pathway, two direct repair routes exist: AGTs reverse O^6- alkylated guanins and AlkB reverse *N*-alkylated lesions (53).

4.2. Mechanisms and functions of chromatin remodeling

The size of the eukaryotic genome requires tight packing and organizing of the DNA (57), while at the same time ensuring proper accessibility of genes and regulatory elements. The

DNA is wrapped around histone proteins that together make up the nucleosome which is the basic unit of chromatin (57,58). Nucleosomes are positioned on the DNA in an ordered way and they are typically depleted at many enhancers, promoters and terminator regions (59,60). The positioning bias varies between organisms and is found to be more pronounced in yeast than in humans (61). The accessibility of chromatin also varies across the genome (62) and is dependent on the histone proteins and how they are modified (63-65).

Chromatin remodeling proteins are often defined as proteins that can allow access to DNA bound to nucleosomes by altering the nucleosome structure (66). Remodeling proteins gained increased interest when it was shown that these proteins together with other modifying complexes could add and remove modifications on the histone proteins and thereby affect transcription and subsequently the expression of genes (67,68). One common feature for the remodeling proteins is the ATPase domain that enables hydrolysis of ATP which generates the energy necessary to alter histones (69). The chromatin remodeling proteins interact with nucleosomes in at least two ways, either by sliding nucleosomes along the chromatin or by exchanging histones within nucleosomes (70). In both cases the result is recruitment of DNA binding factors (70).

Several families of chromatin remodeling proteins are described, with the main eukaryotic being: SWI/SNF, ISWI, NURD/Mi-2/CHD, INO80 and SWR1 (71). The most studied of these are the SWI/SNF, the INO80 and the ISWI remodelers. They share a very similar ATPase domain that probably have the same biological function (71), and a catalytic domain close to the C termini that varies between the two families (Fig. 1). The catalytic domain of the SWI/SNF remodelers is the bromodomain that binds acetylated histone tails, whereas the ISWI remodelers have two different catalytic domains, known as SANT and SLIDE (72).

The ISWI remodelers have a special role in DNA replication which involves ordering of nucleosomes after the chromatin has been opened (73). The main function of the ISWI remodelers seems to be in ordering nucleosomes to promote transcriptional repression (74,75). This is in contrast to the SWI/SNF remodelers that promote transcription by disordering nucleosomes to facilitate transcription factor binding (76). Although several studies have shown that ISWI remodelers are responsible for nucleosome ordering (77) and SWI/SNF remodelers for nucleosome disordering (78), their role can vary depending on chromatin context (71).

INO80 is another important family of chromatin remodeling proteins. INO80 is different from SWI/SNF and ISWI in several ways. First, the ATPase subunit of INO80 has a spacer region

that split the ATPase domain, a feature shown to be important for protein-protein interactions (79). Second, in contrast to SWI/SNF and ISWI that do not exhibit helicase activity in spite of have a helicase domain, the *S. cerevisiae* INO80 has actually been shown to exhibit ATP-dependent DNA helicase activity (80). The INO80 proteins are responsible for the organization of nucleosomes around the promoter of actively transcribed genes (81) and are shown to be important transcriptional regulators. Indeed, from studies in yeast and plants, it has been shown that INO80 is responsible for regulating as much as 20% of the genes (81-83). The transcription factor Yin Yang 1 (YY1) is one of the subunits of INO80 and is believed to be responsible for at least some of the recruitment of INO80 to promoter regions (84). Recruitment of INO80 to promoter regions is dependent on transcription factors, and the transcription factor Pho4p, which binds the PHO5 promoter, is the central factor in the recruitment (85). The regulation of the INO80 complex itself is important in the adaption and response to external stimuli and signaling. Like most proteins, INO80 can be regulated through post-transcriptional modifications of its subunits (86). As an example, when the cell is responding to damage, INO80 is phosphorylated at the Ies4 subunit by the kinase ATM (86), causing the protein to become activated. Another way of regulating the activity of INO80 is through inositol lipids which also regulate the activity of for ISWI and SWI/SNF (87). When the inositol polyphosphate kinase Arg82 exerts its function on INO80, INO80 dissociates from the promoters it binds to and ultimately reduces the expression of associated genes (87).

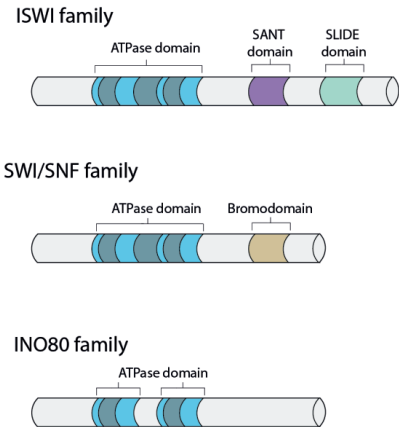


Figure 1: ISWI and SWI/SNF remodelers share a similar ATPase domain which is also found in INO80. However, here it is separated by a spacer region. The ISWI remodelers contains a SANT and SLIDE domain at the C-terminal that probably function in binding histone tails and DNA linkers. The SWI/SNF remodelers contain a bromodomain in its ATPase subunit that binds acetylated histone tails.

Several lines of evidence have shown that INO80 is important for transcription regulation (81). Moreover, recent reports show that INO80 also contributes to DNA replication, and increased expression of INO80 in S phase has been observed (88-90). Experiments have shown that INO80 mutants are highly sensitive to replication-blocking agents, suggesting that INO80 is directly involved in chromatin organization at the replication fork (66,91). Moreover, INO80 mutants show severe problems in resuming DNA replications after treatment with methyl methanesulfonate (MMS), a well-established DNA replication blocking agents. This observation has been explained by decreased levels of the ubiquitinated forms of PCNA, and that INO80 is required for PCNA ubiquitination through its ATPase and remodeling properties (92). However, it is not known if this is a direct effect of INO80 or if the effect is indirect through the recruitment of other proteins by INO80 (66).

INO80 and many of the other remodeling proteins have received increased interest in the field of DNA repair. The following section describes some of the main mechanisms of how chromatin is reshaped and altered by remodeling proteins in response to DNA damage.

4.3. Chromatin remodeling and DNA repair

There has in recent years been a strong focus in the field of DNA repair to understand how DNA lesions in closed chromatin are made accessible for the repair machinery (93). Damages that are formed in heterochromatin are thought to be processed slower than those formed in euchromatin (94,95), and the damage response itself is dependent on the chromatin landscape within which the damage occurs (96-98). ATP-dependent chromatin remodeling enzymes are thought to play a key role in organizing the chromatin before and after DNA repair (99). These enzymes are able to use energy derived from ATP hydrolysis to change chromatin structure. The enzymes can shorten or extend the linker DNA between nucleosomes, load or remove nucleosome from the DNA or replace histone variants on the nucleosomes (99).

In response to DNA damage, chromatin opens up to increase the access for the repair machinery (100). In the case of double strand breaks, several proteins from the phosphatidylinositol 3-kinase-like family of kinases including ATM, ATR, and DNA–PK are important (101). Ataxia telangiectasia mutated (ATM) is a serine/threonine protein kinase protein responsible for the relaxation of chromatin (94) and is induced in response to DNA damage (102). One of the substrates for ATM is KAP-1 (94,103). ATM is thought to work

through the phosphorylation of KAP-1, which perturbs heterochromatin when the interaction between KAP-1 and heterochromatin is diminished (94,104). Although ATM has been shown to play a central role in the repair of at least a subset of DSBs (105), some controversies exist on its importance (106). The requirement for ATM seem to correlate with chromatin complexity and in heterochromatin regions the cell is unable to access and repair damaged DNA unless ATM is active and phosphorylates KAP-1 (94).

The INO80 complex is recruited to DSB, suggesting a role in direct repair. Specifically, in budding yeast the INO80 complex is recruited to a 1.6 kb region around the DSB and this recruitment is shown to be guided by the phosphorylated form of H2AX (γ-H2AX) (107). The function of INO80 in DSB include both non-homologous end-joining and homologous recombination (101), which are the two mechanisms the cell uses for DSB repair. A more detailed mechanism was described by Tsukuda et al. when they showed that absence of histone removal at DSB regions resulted in delayed recruitment of the repair protein Rad51. They hypothesized that INO80 was involved in this process through its regulation of H2A (108).

The role of chromatin remodeling proteins in DNA repair goes beyond the repair of DSB. Chromatin remodeling genes are activated in response to UV irradiation (109-112). UV damage is repaired primarily by nucleotide excision repair (NER) (113). The NER pathway is divided into two pathways: global genomic repair (GGR), which works genome wide and transcription-coupled repair (TCR), which repairs lesions in regions actively transcribed by RNA pol II (114). The cell faces the same challenge for UV induced damage as for DSBs in recognizing and accessing the damaged site. In response of UV damage the DNA is unwrapped from the histones allowing repair proteins to access (115). As expected, TCR is a more rapid process than GGR, possibly impart of the open chromatin state around the transcription site which leads to better recognition and access to the damage (116). It has also been shown that Histone 2A, 3 and 4 are ubiquitinylated at the sites of damage, causing a more relaxed DNA-histone interaction (117,118). In eukaryotic cells these modifications are achieved by the DDB1–DDB2 complex, which also recognizes the damage (117). In addition to the DDB1–DDB2 complex, both the SWI/SNF and INO80 chromatin-remodeling complex have been reported to promote removal of UV-generated lesions in the NER pathway (119,120). The SWI/SNF complex is shown to contribute to removal of pyrimidine(6-4)pyrimidone photoproducts (6–4PPs) (119,121), whereas INO80 is involved in the removal of both Cyclobutane Pyrimidine Dimers (CPDs) and 6–4PPs (120). The INO80 complex has

also been shown to contribute to removal of UV lesions in non-transcribed regions through GGR (120).

Histone acetylation is another mechanism that is regulated in response to UV damage (122,123). In humans and yeast, histone 3 (H3) acetylation is increased after UV damage and correlates with NER activity (122,123). Since H3 acetylation in general is associated with open chromatin this modification is thought to facilitate the access of NER proteins and increase repair (109).

In summary, the interplay between DNA repair genes and chromatin remodeling proteins is highly complex and an essential factor for maintaining genome integrity. The remodeling proteins are highly conserved during evolution and are critical for repair of various lesions and for repair pathways function, including MMR, BER and especially DSB repair.

4.4. Gene regulation by histone modification and Polycomb proteins

An important group of chromatin remodeling proteins is the Polycomb group (PcG) proteins. The PcG proteins were discovered in *Drosophila melanogaster* in 1978 in a famous Nature paper by Edward B. Lewis, which later resulted in the Nobel prize (108). Lewis et al. showed that PcG regulates a set of genes, which was later known as the Hox genes, during embryonic development in the fly. The PcG proteins are responsible for forming the H3K27-me2/me3 marks on the histone proteins (124). The PcG proteins are part of a large Polycomb repressive complex (PRC), which are separated into two sub-complexes PCR1 and PCR2, containing several proteins with distinct functions (124). PCR1 has four main subunits, Cbx, Ring1, Phc, and Bmi1/Mel18 (reviewed in (125)). One of the important features of the Cbx subunit is the ability to bind di- or trimethylated K9 and K27 lysine residues on the H3 peptides (126). PCR2 has three main subunits, Ezh2/Ezh1, Eed and Suz12 and various combinations of these exist (126). One very important feature of PCR2 is its ability to methylate H3 lysine 27, and PCR2 may be the only protein complex that are able to produce di- and tri-methylated H3 lysine 27 (127). Since the Cbx subunit of PCR1 is able to bind di- and tri-methylated H3 lysine 27, a model has been proposed where PCR2 methylation is required for PRC1 recruitment to target genes (127). Although PCR1 and PCR2 have distinct subunits, mutations in either of the protein complexes affect the other protein complex, indicating that PCR1 and PC2 are functionally inter-dependent under certain conditions (128,129).

The PcG proteins were first identified as proteins that regulate cell fate decisions at different stages of development by interfering with the Hox genes. Later it was shown that PcG proteins are recruited to genes involved in other processes, for instance genes associated with cancer (130). Several groups have shown that proliferation is altered in stem cells and normal cells as a result of the interaction between PcG and the cyclin-dependent kinases CDKN2B and CDKN2A, encoding the tumor suppressors INK4B, INK4A and ARF (131,132). These genes are associated with the Rb and p53 pathways, two critical G_1/S phase regulators (reviewed in (132,133)). Genes targeted by PcG have a much higher probability to be silenced by DNA methylation compared to other genes (134-136), and in fact, p53 and Rb are often epigenetically silenced by DNA methylation in cancer (137). At least two models exist that explain the increase in DNA methylation of PcG target genes. One scenario is that PcG proteins get up-regulated in cancer, leading to recruitment of DNA methyltransferases and increased methylation. This is supported by the observed up-regulation of PcG in many cancers (138). However, this model is challenges by the fact that cancer cells in general show increased proliferation and the expression of many of the PcG proteins correlates with proliferation (130). Another model has therefore been proposed which tries to explain the increased silencing of PcG targets by de-regulated PcG co-proteins (130). In this model transcription factors are proposed to play a central role and especially the CFTF family of transcription factors, but also other families like Hox, Sox, Runx, Fox, Pax and Gata (130). Much of the complexity in eukaryotes is attributed to gene regulation by transcription factors (139); however, a detaled description of their role is beyond the scope of the present thesis.

In 2006, a paper was released describing a 2.2kb long non-coding RNA (lincRNA) termed HOTAIR, that interacts with PCR2 and presumably works as a scaffold for histone modification enzymes to bind DNA or as a guide RNA to recruit PcG to their target genes (140-142). This discovery, and the discovery of lincRNAs in general, has been very important in the understanding of how proteins are guided and bind to specific regions in the genome (reviewed in (143)).

4.5. MicroRNA biology and function

MicroRNAs are small ~22 nt RNA molecules that function in a single-stranded form to repress mRNAs (144). In humans, almost 2000 unique miRNAs have been discovered and deposited in public databases (145), and it has been predicted that more than 60% of the human genes are conserved miRNA targets (146). MicroRNAs were first discovered in the model organism of nematodes, *Caenorhabditis elegans*, as a regulator of developmental

timing (3,147). Later, a large number of studies have demonstrated the importance of miRNAs in a wide range of biological processes (148,149).

MicroRNAs are transcribed in the nucleus by RNA polymerase II into long primary transcripts (pri-miRNA) that contain one or more miRNA hairpin structure (144) (Fig. 3). The hairpins have a characteristic structure with a double stranded stem and a terminal loop, both of which serve as recognition features for proteins in the maturation process (150,151). Mechanistically, a protein complex called the Microprocessor, consisting of the RNase III family nuclease Drosha, and its co-partner DGCR8, uses the flanking ssRNA segments in the pri-miRNA to bind and cleave the pri-miRNA approximately 11 nts the stem-ssRNA junction (152) (Fig. 3). DGCR8 is proposed to function as an anchor that measures the correct distance from the stem-ssRNA junction (152). The Microprocessor requires at least 40 nts on each side of the hairpin duplex for proper processing of the pri-miRNA (153). A long-standing question in miRNA biogenesis has been how the cell is able to distinguish an authentic pri-miRNA from other types of hairpins transcribed in the nucleus. Bartel et al. reported three possible primary-sequence determinants in non-nematode pri-miRNAs, including a SRp20-binding motif (CNNC) down-stream of the pri-miRNA hairpin, a UGU or GUG motif in the terminal loop and a basal UG motif (151). They found that 79% of the conserved human miRNAs contained one of the three pri-miRNA determinants, indicating that they have a function in pri-miRNA recognition.

In 2007, Bartel et al. described an alternative to the canonical pri-miRNA processing in which intronic regions get transcribed into pri-miRNA like molecules that enter the miRNA-processing pathway without cleavage by Drosha. They termed these molecules “mirtrons” (154). Instead of using Drosha, mirtrons get processed by the endogenous splicing machinery.

After the pri-miRNA has been processed by the Microprocessor complex (or the splicing machinery for mirtrons), it leaves a precursor miRNA of about 70 nts in length that are exported from the nucleus to the cytoplasm (155,156). The export is mediated by Exportin-5 and its co-factor Ran-GTP (157). Once in the cytoplasm, the RNase III protein Dicer recognizes the terminal loop in the pre-miRNA and removes it to leave a double-stranded miRNA duplex. The precise mechanism for Dicer cleavage is not fully known. However, at least three models have been established trying to explain this process. Two models are that Dicer measures a fixed distance (usually 22 nts) from either the 3' end overhang (3' counting rule) (158,159) or from the 5' phosphate group (5' counting rule) (160) and thereby determine the cleavage site. A third model is that Dicer recognizes the terminal loop in addition to the

3'/5' ends and uses both features to determine the cleavage site (161). One strand of the miRNA duplex is subsequently incorporated into the RNA-induced silencing complex (RISC, or miRISC for miRNA-containing RISC), where it binds to a protein in the AGO protein

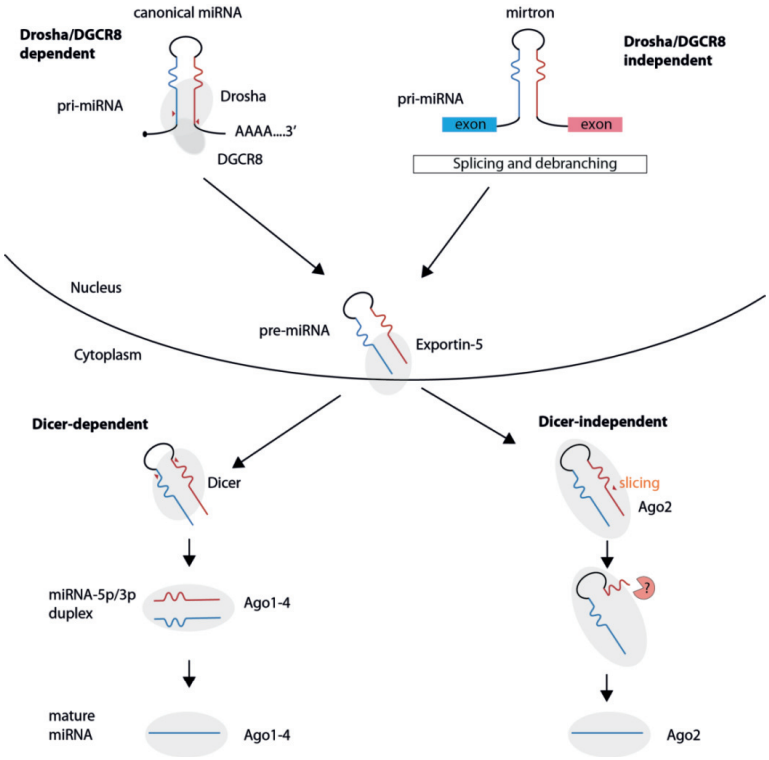


Figure 3: MiRNA biogenesis pathway. A detailed description of the different steps in the pathway is found in the text.

family. The strand that is not incorporated into RISC is usually degraded by nucleases in the cytoplasm, but can sometimes act as a functional miRNA (162). Ago2 has the ability to cleave miRNA precursors to create mature miRNA in the absence of Dicer, however this is regarded as a rare exception (163,164). Once the mature miRNA is loaded into RISC is has the ability to target and repress complement mRNA sequences. The targeting itself is a complex mechanism that involves base-pairing between a specific part of the miRNA and usually the 3' untranslated region (3' UTR) of the target mRNA. Illustration of miRNA directed repression is shown in figure 4.

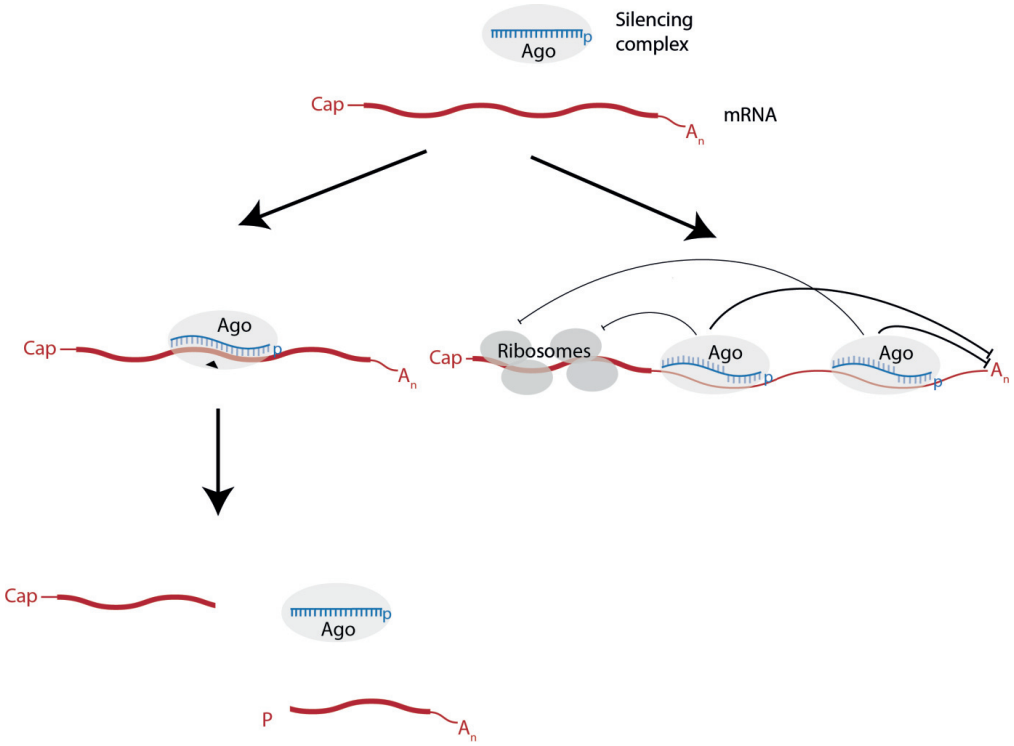


Figure 4: MicroRNA mediated gene repression. A miRNA (blue) incorporated in the silencing complex pairs to a target mRNA (red) together with an Argonaute protein (Ago) to direct mRNA repression. If the base-pairing between the miRNA and the targeted mRNA is perfect or near perfect the repression can occur through mRNA cleavage (left side in figure). If the miRNA binds the targeted mRNA with only partial complementarity, the repression will not involve Ago-mediated cleavage. In this second process, the miRNA binds the mRNA using a short sequence at the 5' end of the miRNA called the miRNA seed. This type of repression involved recruitment of factors that causes translational inhibition (thin repressive lines) or, more common, destabilization of the mRNA followed by shortening of the poly(A) tail and degradation (thick repressive lines).

4.5.1. MicroRNA targeting

The average miRNA is 22 nts in length (165) and contains a region of 6 nts at the 5' end called the seed region that is critical for pairing to target mRNAs (166). David Bartel et al. showed that >60% of human protein-coding genes have conserved binding sites for miRNAs in their corresponding mRNAs (146). The pairing between miRNAs and mRNAs is generally divided into four main binding sites; a 6mer, two 7mers, and one 8mer (Fig. 5) (166). The 6mer is a perfect 6nt match between nucleotides 2-7 of the miRNA and the mRNA. The two types of 7mer sites are 1) a seed match with additional paring at miRNA nucleotide 8 in the miRNA 2) a seed match followed by an A at target position 1. The strongest binding sites are

8mers which comprises the seed match and both additional paring at miRNA nucleotide 8 in the miRNA and an A at target position 1.

Since a 6mer match is expected to occur once every 4,096 nucleotides in the genome, clearly other factors must be involved in determining a functional miRNA site. Several additional structure and sequence features are shown to influence the repression by miRNAs. For instance, closely spaced sites result in more down-regulation of mRNAs compared to if the sites are further apart (166,167), and similar if there is additional pairing at nucleotides closer to the 3' end of the miRNA, called 3′-compensatory pairing (166). 3′-compensatory pairing is thought to compensate for mismatches in the seed region and enhance recognition of mRNA targets (166). The miRNA seed region has strong affinity of target mRNAs, mainly as a result of how the Ago proteins are pre-organizing the 5′ region of the miRNA to favor Watson–Crick pairing in this region (162).

David Bartel and his group described another type of pairing between miRNAs and mRNA involving the central region of the miRNAs (168). They showed that in the absence of perfect seed pairing and 3'-compensatory pairing the miRNA could bind to an mRNAs using 11-12 nucleotides in the central region of the miRNA. Despite being less conserved than the 5' and 3' region of the miRNA, Bartel et al. showed that "central pairing" is indeed functional in reducing mRNA and protein abundance.

General sequence features have been shown to increase target site efficiency. As an example, target sites in regions with high AU-content are more efficient than those residing in other regions (166). Furthermore, target sites positioned away from the central region of the 3' UTR show more efficient repression, especially if the 3'UTR is long (166).

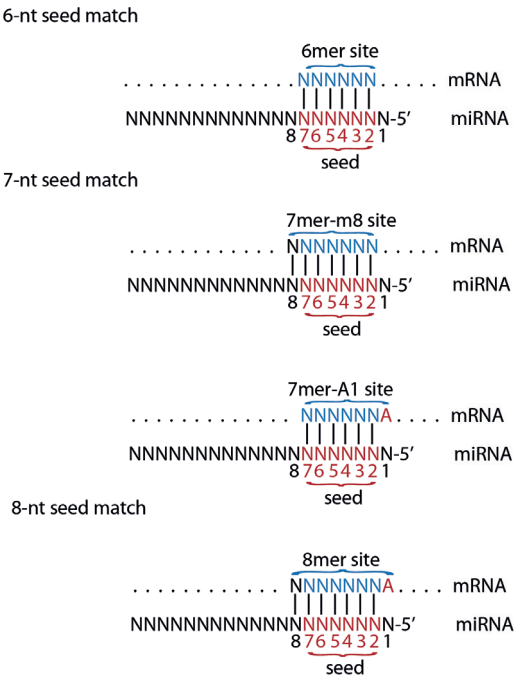


Figure 5: Four types of miRNA sites. A detailed description of the sites is found in the text.

A recent study by MacRae et al. uses crystal structures of human Ago2 to explain the structural basis for miRNA targeting (169). By crystalizing Ago2 bound to a miRNA with and without target mRNA, they showed that miRNA targeting takes place in a stepwise mechanism. First, the guide nucleotides 2-5 are exposed to the target mRNA. This initial targeting creates a conformational change in the miRNA that exposes nucleotide 2-8 and 13-16 for extended target recognition. Specifically, they show that seed pairing leads to a rearrangement of the 3' end of the miRNA, increasing the interaction between the miRNA and the mRNA target. These results explain the conservation pattern previously observed for miRNA targets (170). It also explains why miRNAs sometimes have pairing at the 3' end, as identified by Bartel et al. (166).

4.6. The Argonaute proteins

Argonaute proteins (Ago proteins) are the backbone proteins in all small RNA-guided silencing processes. They are conserved in all examined eukaryotes, with the exception of *Saccharomyces cerevisiae*, which do not possess small RNA regulation (171). The Ago

proteins are the direct binding proteins for small RNAs and have their names after the squid like phenotype of *Arabidopsis thaliana* Ago1 mutants (172). The mammalian genome contains four homologous Ago genes: Ago1-4 (173). The four Ago genes have diverged from a common ancestral Ago gene and three of the genes (Ago1 and Ago3-4) are found as tandem repeats in the same human chromosome, while Ago2 is located on a different chromosome (173).

The Ago proteins are characterized by four protein domains; N (amino-terminal); PAZ (PIWI–ARGONAUTE–ZWILLE); MID (middle); and PIWI (Fig 6). The PAZ domain contains a small RNA binding pocket that binds the 3’ end whereas the MID domain binds the 5′ end of the small RNAs by interaction of the 5' terminal base of the RNA with a conserved

Figure 6: Position of domains in the Argonaute protein (human Ago2 is shown here). All Ago proteins share two domains: The PAZ domain (yellow) and the PIWI domain (orange). The PAZ domain contains a specific binding pocket that fixes the two-nucleotide 3’ overhang. The show high degree of homology to RNase H and this domain is referred to as the slicer domain in human Ago2. The MID domain lies between the PAZ and the PIWI domain and binds the 5' phosphate group of small RNAs. The MID domain is also shown to be important for the interaction between Ago and other proteins.

tyrosine residue in the protein (174). The N domain is required for correct loading of small RNA and plays a role in unwinding the small RNA duplex (175). High resolution crystal structure of human Ago2 and miR-20a showed that the miRNA is bound at both ends by the MID and PAZ domains and that the binding of a miRNA increases the stability of the Ago protein by locking it in a stable conformation (176). The crystal structure revealed that all domains in the Ago2 protein are interacting with the miRNA, which is probably the reason why the stability of the protein is increased in the presence of a miRNA (176).

One of the great mysteries in small RNA biology has been how the Ago proteins choose the correct strand in the RNA duplex and how the strand is loaded onto the Ago proteins. One approach has been to examine the stability of the duplex, and the strand with less stable 5′ end is shown to be preferentially loaded (177), a phenomenon known as “the asymmetry rule” (178). Furthermore, it has been reported that Ago1 and Ago2 both possess ATP-independent strand-dissociating activity of miRNA duplexes, a function reminiscent of RNA chaperone

activity (179). Ago2, which is the only member of the Ago protein family that possesses catalytic activity (180), is able to cleave perfectly paired siRNA duplexes and more rarely also miRNA duplexes (181). The catalytic property of Ago2 enables rapid dissociation of the passenger strand which could have important implications for small RNA activity and target repression by rapidly activating RISC (182-184).

Controversies still exist whether subpopulations of miRNAs preferentially bind to different Ago proteins. Some reports show that this is indeed the case (185), and others show that miRNAs bind randomly to Ago proteins (186,187). One report by Meister et al. indicates that the four Ago proteins prefer miRNAs with specific lengths (186). Specifically, they show that Ago2 prefers miRNAs of length 22, whereas Ago1 and Ago3 have a tendency to bind miRNAs of length 23 and 24 nts. Such length preferences were studied in more detail by Makeyev et al. who showed that Ago2-loaded miRNAs undergo efficient 3′-terminal trimming to yield miRNAs 1-2 nts shorter than those loaded onto the other Ago proteins (188). They further showed that a specific residue in the PAZ domain of Ago2 provides the 3' trimming property, possibly by creating a looser interaction between the miRNA 3' end and the protein, making the miRNA 3' end more accessible to a trimming exonuclease. Although miRNAs are diverse with regard to sequence, most miRNAs tend to start with A or U at the 5’ end (176,189,190). Crystal structures of the Ago2 MID domain in complex with nucleoside UMP or AMP (nucleotide U or A), revealed specific interactions between these two nucleosides and the MID domain, explaining the previously observed preference for A and U at the 5’ end of the miRNA (191). In detail, they show that a conserved region in the MID domain, called the nucleotide specificity loop, is responsible for the interaction with the nucleotide base, and that this region is important for the base preference and sorting of small RNAs to the Ago proteins. Several studies have demonstrated the importance of the 5’ nucleotide in small RNAs. In *T. thermophiles,* the miRNA 3’ end is released from the Ago PAZ domain when the miRNA binds a target mRNA, whereas the miRNA 5’ end preserves its interaction with the MID domain (192). In *Arabidopsis thaliana,* the 5′ nucleotide is functioning as a sorting signal for small RNAs to the different plant Ago proteins (193-195). Similar observations have been reported for flies for miRNAs associated with Ago1 and Ago2 (189,196).

The expression of the four Ago proteins varies between tissues and Ago4 in particular seems to be expressed in a limited number of tissues compared to the other Ago proteins (197-199). Moreover, the expression of Ago3 and Ago4 is generally lower than that of Ago1 and Ago2

(197). It is speculated that differences in codon distribution in the N-terminal region are responsible for the differences in expression, both at the protein and mRNA level (197). In the past few years, more studies on the function of the different Ago proteins have been conducted. We know that Ago proteins are involved in a wide range of cellular processes including transcription, alternative splicing and possibly also DNA repair (reviewed in (200)). Ago2 in particular appears to be indispensable in some processes. Studies in mice have shown that Ago2 is essential during early development (201,202) and necessary for the processing of miR-451 which plays an essential role in hematopoiesis (163,203). Moreover, changes in Ago2 expression during brain development is shown to influences miRNA activity (188). Taken together, much is known about the structure and function of the Argonaute proteins both in relation to small RNA regulation and other processes; however, there are still mechanisms far from fully understood. For example, how is the loading of small RNAs into the Argonaute proteins regulated, and how are the small RNAs dissociated from the protein to ensure rapid turnover in response environmental changes or during developmental transitions.

4.6.1. Nuclear function of Argonaute proteins

Argonaute proteins are present both in the cytoplasm and in the nucleus and that they have conserved functions in both compartments of the cell (204-207). Small RNAs in complex with Ago1 and Ago2 are shown to bind complementary regions in the DNA of promoters regions and transcription start sites (208). Sequencing of small RNAs bound to Ago1, Ago2 and Ago3 shows that ~0.02% of the sequences map to promoter regions, most of which are associated with Ago1 (205). Ago1 is shown to interact directly with RNA polymerase II (205,209,210), particularly at promoters of actively transcribed genes marked with histone 3 lysine 4 tri-methylation (H3K4me3) (205). In *Drosophila,* Ago2 localizes to euchromatin but not heterochromatin (211). While both Ago1 and Ago2 are actively transported into the nucleus, the nuclear role and localization are different between the two proteins. Ago1 mainly localizes to the nuclear interior, whereas Ago2 is found in the inner nuclear envelope (205,212).

Several examples exist showing that particular miRNAs bind to promoter regions and act in transcriptional gene silencing (TGS). Ago1 was found to be enriched at the CCNB1 gene promoter and Ago1 together with the three miRNAs, miR-744, miR-1186 and miR-466d-3p induces the expression of CCNB1 (210). Several similar interactions are reported, for instance *POLR3D* silencing by miR-320 (213), *PGR* and *IGSF1* silencing by miR-423-5p (214) and

CDKN1 activation by miR-373 (215).

The miRNA miR-29b is shown to be primarily localized in the nucleus of cyclic cells (216). It was shown that a hexanucleotide 3′ terminal motif (AGUGUU) of miR-29b specifies cellular localization of this miRNA and that Ago2 and Importin8 are responsible for the import (216). The exact role of miR-29b in the nucleus is not known; however, Hwang et al. speculated that it might be involved in regulating transcription or splicing, in addition to or instead of canonical miRNA gene silencing (216). Ago2 associated small RNAs are previously shown to be involved in alternative splicing. Corey et al. showed that small RNAs recruit Ago2 to pre-mRNA transcripts and thereby promote intron/exon exclusion (217). Moreover, the endogenous non-coding RNA MALAT was shown to interact with Ago1 and Ago2 (218) and to regulate alternative splicing (219).

Another surprising function of Ago proteins was revealed by Qi et al. who showed that ~21-nucleotide small RNAs (later termed diRNAs) are transcribed around sites of DSB, both in plants and animals (9) and are necessary for proper DSB repair together with Ago2. They further showed that knock-down of Ago2 reduced DSB repair, indicating that Ago2 is important for diRNA function. Subsequently, it was shown that diRNAs are restricted to homologous recombination (HR) by recruiting the repair protein Rad51, and that the catalytic activity of Ago2 is required for efficient DSB repair (220). The exact mechanism of diRNAs is not fully known. However, these findings reveal a novel function for Ago2 and small RNAs in the nucleus.

4.7. The eukaryotic cell cycle

The cell cycle, or cell division cycle, is separated into distinct phases that each are characterized by a specific molecular process (reviewed in (221)). These phases are Growth phase 1 or Gap phase 1 (G_1); DNA synthesis phase (S); Growth phase 2 or Gap phase 2 (G_2) and Mitosis (M). Cells that are not dividing are often said to be in G_0, a state of quiescence. Based on the physical appearance in the microscope the cell cycle is divided into two broader phases: interphase and mitosis. Mitosis only constitutes about 5% of the cell cycle time, which implies that most of the time the cell is in interphase (222). During mitosis the cell separates the chromosomes and eventually divides into two daughter cells (cytokinesis). The length of the cell cycle varies largely between cell types in culture, but the average cell length tend to follow the circadian clock with a duration of about 24 hours (223,224). The longest cell phase is usually G_1 phase, during which the cell prepares for DNA replication by

producing the required proteins and enzymes. During S phase the cell duplicates its DNA and as it enters G_2 phase the cell contains twice as much DNA as in G_1 phase. Usually the cell proceeds rapidly through G_2 phase and enters mitosis 2-4 hours after leaving S phase. The resulting two daughter cells will then enter G_1 phase and start a new cell cycle given the right conditions (223).

4.7.1. Regulation of the eukaryotic cell cycle

Normal dividing cells are subject to complex regulation at multiple steps in the cell cycle. This regulation is closely linked to sensing of damaged DNA which activates DNA damage response (DDR) pathways. Most lesions to DNA are repaired by one of the many DNA repair pathways (reviewed in (10,225)). In case of severe damage, the cell may enter the pathway of programmed cell death called apoptosis (226).

Transitions between the different cell cycle phases are critical points and therefore highly regulated (227). For example, a cell that contains damaged DNA should not enter M phase, as the damage will then be propagated to the next generation of cells. Cell cycle transitions are irreversible. This means that once the cell commits to enter a new cell cycle phase it cannot return to the previous phase, even when the transition signal disappears (228). This irreversibility is controlled by a feedback system that prevents cyclin resynthesis (229). The eukaryotic cell has three irreversible cell-cycle transitions: G_1–S, G_2–M and M–G_1. At each transition, specific cyclin-dependent kinases (CDKs) are ubiquitinated and degraded (230). CDKs play a central role in regulating transitions between cell cycle phases (227). The CDKs are highly conserved during evolution, indicating an indispensable role in regulation. The mammalian CDK family includes 13 proteins, but only CDK1, CDK2, CDK4 and CDK6 are directly involved in cell cycle regulation (Fig. 7) (reviewed in (231)). As the name indicates, CDKs are in complex with cyclins. Type-D cyclins (D1, D2 and D3) sense mitogenic signals and form complex with CDK4 and CDK6 during G_1 phase. These complexes are responsible

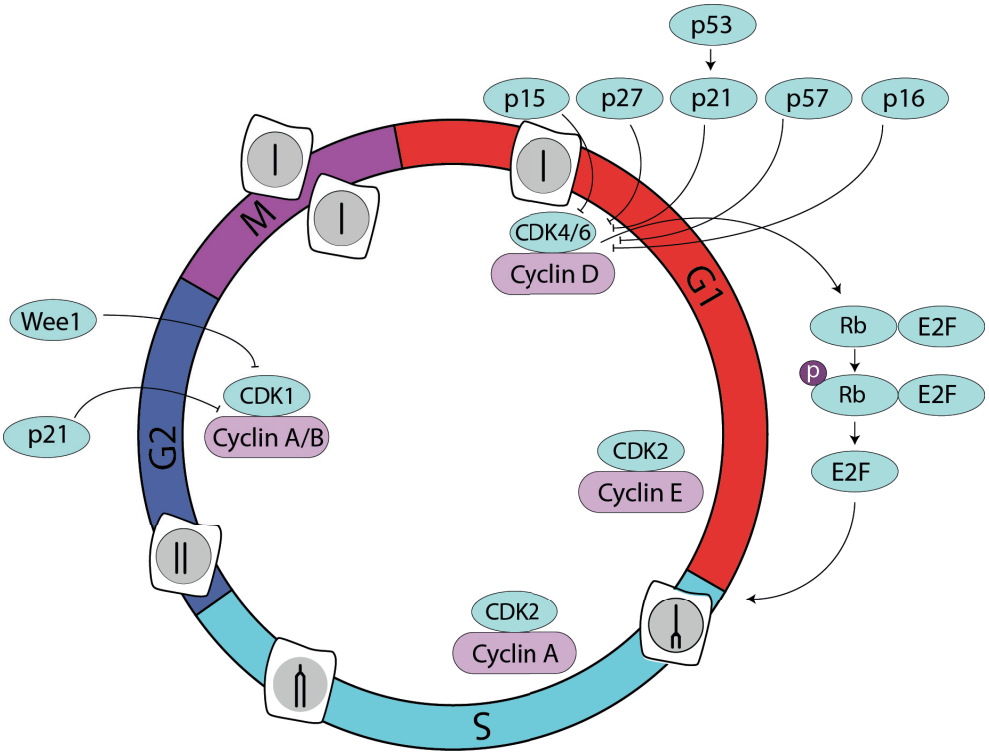


Figure 7: Overview of the human cell cycle and the main proteins involved in regulating the transition between phases. Cell cycle phases are indicated as S: S phase; M: Mitosis; G_1: G_1 phase and G2: G_2 phase.

for inactivation of the Retinoblastoma-associated protein (Rb), resulting in cyclin-E activation and CDK2/cyclin-E complex formation (232). In late S phase, CDK2 is activated by cyclin A2 which drives the cell from S phase to M phase. At the end of interphase, A-cyclins activate CDK1 thereby facilitating M phase progression (233). At the G_2–M transition, the nuclear kinase Wee1 is also degraded, leading to CDK1 activation (234). During mitosis the A-cyclins are degraded, causing CDK1–cyclin B complexes to be formed, which is part of the final regulation that drives the cells through cell division (233). The degradation of mitotic cyclins is ubiquitin-mediated and controlled by the anaphase promoting complex (APC), an E3 ubiquitin ligase.

Based on the mechanism described above, CDKs may seem essential for cell cycle progression. Surprisingly, knock-out of selected CDKs in mouse models leads to developmental defects, but does not lead to loss of cell cycle division (235,236). Moreover, mouse embryos lacking interphase CDKs develop normally until mid gestation (237). These

observations indicate that CDKs are required for some cell types at some developmental stages, but not for all cellular lineages.

4.7.2. When the regulation goes wrong

More than 100 genes are classified as so-called cancer predisposition genes (CPGs) (238). There is no clear definition of CPGs; however, these are generally genes that increase the risk of cancer development if mutated. CPGs are often associated with central cellular processes such as DNA repair and cell-cycle regulation, but also include genes in the chromatin-remodeling pathway.

Examples of proteins that are implicates in many cancers are CDKs. As described in the section above, CDKs control critical regulatory pathways in the cell cycle. A number of studies have shown that dys-regulation of CDKs is contributing to tumor development and tumor progression (reviewed in (239,240)). CDK activity is controlled by a wide set of inhibitor molecules, including the INK proteins (INK4A, INK4B, INK4C, INK4D) and the Cip and Kip proteins (p21, p27 and p57) (231). Proliferating stem cells are regulated by these inhibitors (233). Specifically, the INK proteins p21 and p27 are required to control self-renewal of neural, intestinal and hematopoietic progenitors, and similarly the INK proteins INK4A, INK4B and INK4C regulate self-renewal of hematopoietic, brain, lung and pancreatic stem and progenitor cells (reviewed in (233)).

The dys-regulation of CDKs can be either through overexpression of CDKs or loss of activity from CDK inhibitors (239). Mice resistant to the INK4 inhibitor develop hyperplasia in the same cells that require CDK4, as well as other tumor types. CDK4 is over-expressed in many tumors, including gliomas (241), sarcomas (242), breast tumors (243) and in uterine cervix cancer (244). Similar, CDK6 expression is elevated in squamous cell carcinomas and gliomas (245) and lymphoid tumors (246). Furthermore, deletions in the 9p genomic region, which overlaps with several cell cycle gene loci, including the INK4 family of inhibitors (p16/CDKN2A and p15/CDKN2B), are among the most frequent alteration in cancer (239). In lymphoid tissues, deletions in this regions are often linked to non-functional V(D)J recombination, an important step in immunoglobulin production in the immune system (247).

Since CDKs are dys-regulated in many cancers, either directly or indirectly through loss of inhibitors, one might ask if the tumors require CDKs to maintain their proliferating nature. Some tumor cell lines depend on interphase CDKs, whereas others seem to divide without CDK activity. For instance, colon carcinoma cell lines *do* proliferate in the absence of CDK2,

whereas cells from glioblastomas and osteosarcomas are sensitive to reduced CDK2 levels and proliferation is inhibited (248,249). In CDK4-deficient mice, resistance to mammary tumor development is observed (250). Similarly, mice lacking cyclin D, one of the co-factors of CDK4, are also resistant to breast tumors (251).

Drugs developed to target CDKs are generally less specific and they often target several similar proteins at the same time (239). Less specific drugs have higher anti-proliferative effects compared to more specific drugs with single targets (239).

Several transcription factors are dys-regulated in cancer, many of which promote metastasis (reviewed in (252)). Transcription factor c-Myc is predicted to target 15% of human genes, many of which are genes controlling cell cycle and apoptosis. C-Myc is classified as an oncogene and is very often found to be up-regulated in cancers in addition to promote metastasis (253). Members of the ETS family of Transcription factors (comprising more than 30 genes in humans (254)) are also commonly dys-regulated during tumor development. The expression of several members of the ETS family frequently correlate with tumor aggressiveness (252) and is elevated in migratory cells (254). Another large group of transcription factors are the forkhead box (Fox) proteins, with at least 41 proteins identified in humans (255). The number of Fox protein genes is correlated with tissue complexity of organism during evolution and the function of the different Fox proteins varies largely (reviewed in (255)). The FOXM1 transcription factors are positive regulators of G_1/S and G_2/SM cell cycle phase transition (255). Opposite, FoxO inhibit cell cycle progression (255) and deletion of FoxO is associated with cancer-predisposition (256). Moreover, FoxP, another Fox family member, is crucial for development of the immune system and is linked to hyper proliferative disorders in the immune cells (257). The Fox proteins are also regulating cancer progression and have therefore been suggested as potential prognostic markers (255). Specifically, FOXC1 and FOXC2 are responsible for activating essential growth factors during vascular development (258), and FOXO1 is important for angiogenesis during embryonic development in mouse (259). The NF-κB family of transcription factors is also a central player in cancer. They are in complex with other transcription factors such as p53, ETS and STAT, and have important roles in regulating inflammation and the immune system (reviewed in (260)). NF-κB has at least two main functions in relation to cancer. It is part of the immune system that targets cancerous cells, and functions as an oncogene when dys-regulated in cancer cells (260). Tumor suppressor p53 (or TP53 for gene name) is one of the hallmark tumor suppressor genes. The p53 protein is controlling cell cycle progression at the

transition between G_1 and S phase and initiates apoptosis if damaged DNA is not repaired. Most mutations affecting TP53 alter its wild type function, leading to loss-of-function. However, some studies have reported gain-of-function mutations leading to increased mutation rate and metastasis. As mutations in TP53 rarely affect its transcriptional domains, it is believed that mutated p53 still binds to DNA to alter gene expression, often in an unfavorable way for the cell. However, whereas several reports show that TP53 is mutated in cancer (reviewed in (261)), the exact mechanism of its function and the network of proteins that are involved around p53 is still not fully elucidated (262), which is why it is still frequently studied.

4.8. MicroRNAs and the cell cycle

The first hint about the involvement of miRNAs in cell cycle regulation came from a study on the miR-15-16 cluster showing that these miRNAs have anti-proliferative properties and that they were deleted in more than half of chronic lymphocytic leukemia (CLL) patients (263). The mapping of miR-15 and miR-16 showed that they were located at the chromosomal region 13q14.3, a region often altered in CLL and other cancers (264-267). Subsequently, the miR-15-16 cluster was shown to be deleted or down-regulated in a large number of cancers and to regulate critical cell cycle genes required for normal proliferation (reviewed in (268) and (269)). In the years following the miR-15-16 discoveries, many other interactions between cell cycle genes and miRNAs were described. The sections below review some of these interactions and describe the importance of miRNAs in cell cycle regulation.

4.8.1. MicroRNAs and p53

The p53 tumor-suppressor is part of a large network of proteins that sense DNA damage and various stress signals and respond by inducing growth arrest and promote apoptosis or DNA repair (270,271). Hannon et al. showed that the expression of the miR-34 family correlates perfectly with the expression of p53 and that these miRNAs were induced together with p53

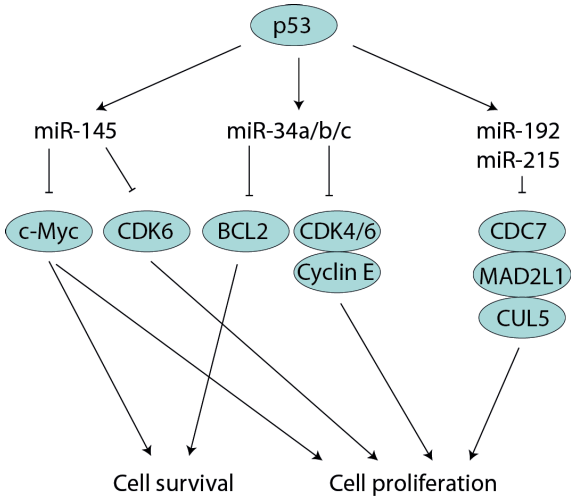


Figure 8: miRNA regulation by p53. Details on the interactions are found in the text.

when cells were treated with ionizing radiation (Fig. 8) (272). They further show that miR-34a/b/c are all direct targets of p53. When over-expressing miR-34 they observed growth arrest at G_1 combined with down-regulation of cell cycle genes, including cyclin E2 (CCNE2) and cyclin-dependent kinase 4 (CDK4). They concluded that p53 represses these genes by inducing miR-34. Other groups have reported similar results for other miRNAs, including miR-192, miR-215 and miR-145. Chau et al. showed that stress induced p53 causes up-regulation of miR-192 and miR-215, and like miR-34, these miRNAs have several down-stream targets important in cell cycle regulation (273). p53 also represses the proto-oncogene c-Myc by binding to and inducing miR-145 which is as a direct target for c-Myc. In addition to its role in inducing specific miRNAs, Miyazono et al. described a fascinating role of p53 in enhancing the maturation of several miRNAs, including miR-16-1, miR-143 and miR-145, by interacting with the Drosha processing complex (274). This raises the possibility of a more global function for the interaction between p53 and miRNAs.

4.8.2. MicroRNAs and the G_1/S transition

Cells have acquired a precise network of regulation at the G_1/S transition to prevent abnormal cell proliferation and tumorgenesis. The commitment to enter S phase depends to a large degree on cyclin–CDK transcription and E2F transcription factors play a central role in this regulation (reviewed in (275) and (276)). E2F itself and many of the other G_1/S associated genes are shown to be under regulation by miRNAs, either by direct targeting or through

suppression of important co-factors (reviewed in (277)). Embryonic stem (ES) cells that lack miRNA activity are shown to accumulate in G_1 phase, and surprisingly, by introducing only a few miRNAs from the miR-290 family the cells are able to rescue the defect and continue to proliferate (278). These miRNAs were shown to suppress several important regulators of the G_1/S transition. This shows the importance of miRNA regulation in at this specific cell cycle phase.

O'Donnell's group showed in 2005 that E2F is targeted by two miRNAs, miR-17-5p and miR-20a (279). C-Myc and E2F are both targeting each other's promoter (280,281) and miR-17-5p and miR-20a provides a mechanism that control this positive feedback loop and the expression of both c-Myc and E2F.

Other miRNAs are also shown to regulate E2F. E2F1 is a validated target for miR-149, miR-330 and miR-331-3p (282-284) and similar E2F3 is targeted by miR-125b, miR-210 and miR-195 (285-287). In most of these studies, de-regulation of these miRNAs is associated with altered cell growth or tumorigenesis.

The retinoblastoma protein (Rb) is tumor suppressor protein that is functionally connected with E2F. Rb binds and inhibits E2F and thereby prevents the cell from entering enter S phase (288). When Rb is phosphorylated by CDKs, E2F is released and the cell is able to progress from G_1 to S phase. Although Rb seems to be mainly regulated by phosphorylation, several reports show that miRNAs are functionally regulating Rb family members. The miRNA cluster miR-17-92 targets and down-regulates Rb2/p130, a member of the Retinoblastoma family, thereby promoting differentiation in adipocyte tissue (289). The miR-290 cluster targets Rbl2, thereby controlling DNA methylation through the interaction between Rbl2 and multiple DNA methyltransferases (290). Rb itself is also targeted by miR-106a (291).

MicroRNAs are regulating several genes encoding cyclin-dependent kinases inhibitors (CKIs). CKIs are divided into two families: INK4 and Cip/Kip (292). Some important members of the INK4 family are p15, p16, p18, and p19 whose function is to inhibit the activity of CDK4 and CDK6 (reviewed in (293)) (Fig 7). The Cip/Kip comprises p21, p27, and p57 and this family regulates many other CDK proteins (293). MiR-24 and miR-31 was shown to control proliferation by regulating the CDK4/6 specific inhibitor $p16^{INK4a}$ (294) and $p21^{Cip1}$ is targeted by several miRNAs, including miRNAs in the miR-17-92 family and miR-106b (295,296). One of the most studied miRNAs in the cell cycle, miR-221/222, together with miR-181, control the expression of two of the most important cell cycle inhibitors $p27^{Kip1}$ and $p57^{Kip2}$ (297,298) and reviewed in (299).

4.8.3. MicroRNAs and mitosis

Although most miRNA regulation seems to take place at G_1 and S phase of the cell cycle, some reports show that miRNAs are also active in later cell-cycle phases. CDK1 together with cyclin A and cyclin B is the master regulator in mitosis and its activation is dependent on de-phosphorylation (reviewed in (300)). The microRNAs miR-125b, miR-24 and let-7 regulates cyclin A and cyclin B, however most of these studies focuses on the role of these cyclins in G_1 and S phase (285,301,302). Furthermore, miR-195, miR-516-3p, miR-128a and miR-381 are all validated to regulate the CDK1 inhibitor Wee1, one of the important G_2/M regulators (303-305).

Taken together, given the large network of proteins involved in cell cycle regulation it is not unexpected that miRNAs contribute in the regulation. Focusing on the main cell cycle regulators we have seen that several miRNAs are targeting these genes, indicating that miRNAs are important molecules in cell division.

5. Aims of the study

Paper I:

- Identify cell cycle regulated transcripts in the human non-tumorigenic, keratinocytic cell line HaCaT.

Paper II:

- Provide an updated list of human DNA repair and chromatin remodeling genes.
- Characterize cell cycle regulated DNA repair pathways and chromatin remodeling genes.
- Make a database of cell cycle regulated human DNA repair and chromatin remodeling genes.

Paper III:

- Investigate how lost Ago2 affects miRNA and mRNA expression in normal cycling and serum stimulated mouse embryonic fibroblasts.

Paper IV:

- Investigate how lost Ago2 affects miRNA and isomiR expression in mouse embryonic fibroblasts.
- Identify sequence and structure characteristics that affect miRNA sorting in mammalian cells by analyzing Ago2 knock-out cells and available Ago IP small RNA sequencing data sets.

6. Main findings

Paper I - Transcription profiling during the cell cycle shows that a subset of Polycomb-targeted genes is upregulated during DNA replication.

Previous studies performing gene expression analysis on synchronized cells have shown that a large subset of genes are periodically expressed in the human cell cycle (306-308), as well as in other organisms (309-311). One study showed that there is significant overlap between cell cycle regulated genes in cancer and normal cells (306) and that some genes are unique for each cell type. **In Paper I**, we used whole transcriptome microarrays on double-thymidine and nocadazole synchronized cells to identify periodically expressed genes in human HaCaT keratinocytes as well as the human cervical cancer HeLa cells. We identified 1249 genes as cell cycle regulated which we divide in three groups. One group includes genes with housekeeping functions, characterized by high baseline expression and CpG-rich promoters carrying the histone H3 lysine 4 tri-methylation (H3K4me3) chromatin mark, associated with active transcription. The second category included genes specific for HaCaT cells, which were mostly unique compared to other cell types. The third category included genes carrying the histone H3 lysine 27 tri-methylation (H3K27me3) chromatin mark, associated with Polycomb silenced genes (312). Surprisingly, we identified a subset of the Polycomb-targeted genes to be up-regulated during G_1/S or S phase. This finding was confirmed in other cell lines and suggests that DNA replication offers a possibility for H3K27me3 silenced genes to be expressed by making the chromatin accessible for transcription factors.

Paper II - Cell cycle regulation of human DNA repair and chromatin remodeling genes.

To facilitate repair of damaged DNA, integration between DNA repair proteins and chromatin remodeling proteins is needed (66,101,111). Several DNA repair genes are cell cycle regulated (313,314), and proper regulation of these genes is needed to prevent incorrect transmission of genetic information (315). **In paper II**, we combined published and un-published cell cycle data sets to get an overview of cell cycle regulated DNA repair and chromatin remodeling genes. Each pathway is analyzed separately and we show that long patch base excision repair (BER) is enriched for S phase genes; whereas short patch BER uses genes expressed in all cell cycle phases. Furthermore, we show that MMR genes, Fanconi

anemia genes and homologous recombination genes are up-regulated during S phase. Other DNA repair pathways show less consistent cell cycle regulation. Several chromatin remodeling genes were found to be up-regulation during G_1/S- and S phase, including chromatin assembly factor 1 (CAF-1) major subunits CHAF1A and CHAF1B; the putative helicases HELLS and ATAD2, both involved in E2F activation and DNA double strand break repair, and RAD54L and RAD54B also involved in DNA double strand break repair. Finally, **paper II** provides an update of DNA repair and chromatin remodeling genes and shares the information through a novel Internet page where the expression profile for the genes can be viewed individually.

Paper III - Argonaute2 affects Hox gene expression and the late growth response in mouse embryonic fibroblasts.

Argonaute2 (Ago2) is the central protein in miRNA regulation. In contrast to the other Ago proteins, Ago2 possesses endonuclease activity, enabling cleavage of highly complementary mRNA targets (316) and miRNA duplexes (317). **In paper III**, we investigated time dependent changes in global gene expression between serum-stimulated Ago2 knock out (KO) and wild type (WT) MEF cells. We observed significant changes in the development-related Hox genes, a majority of which were up-regulated in Ago2 KO, suggesting that Ago2 plays an essential role in regulating this cluster of genes. In detail, we observed repression of posterior Hox genes in Ago2 KO whereas anterior Hox genes are highly up-regulated. This suggests that WT cells activate a posterior Hox gene program, which is lost when Ago2 is removed from the cells, leading to activation of the anterior Hox program. Furthermore, we found that Ago2 KO cells recover more slowly from the serum starvation than the WT cells, probably as a result of differences in the activation of transcriptional programs between the two cell types.

Paper IV - Deep sequencing of Ago2 knock-out cells reveals distinct sequence and structure features associated with miRNA sorting.

While much is known about the processing cascade and targeting mechanisms of miRNAs, a good model for how miRNAs are loaded into the different Ago protein remains to be described. In **paper IV** we analyzed miRNA sequences from Ago2 KO and WT cells to

determine the effects of Ago2 on the global miRNA population, and to identify sorting features for miRNAs. We identified several sequence and structure features that could influence miRNA sorting. Analysis of isomiR length showed that long isomiRs were up-regulated in Ago2 KO whereas short isomiRs were down-regulated in Ago2 KO. Moreover, we observed significant differences in base-pairing between Ago2 KO and WT. Specifically; base-pairing at the 5' nucleotide and in the central region of the miRNA duplex were associated with up-regulation in Ago2 KO cells. We also found that miRNAs down-regulated in Ago2 KO have less stable 5' ends. When analyzing a panel of Ago IP data sets, we observed similar base-pairing trends as in the Ago2 KO study. Here, base-pairing at the 5' nucleotide were associated with miRNAs down-regulated in Ago2 IP, which confirms the finding in Ago2 KO. The results indicate that Ago2 is more sensitive to base-pairing at the 5' nucleotide than the other Ago proteins, and that Ago2 prefers an unpaired 5' nucleotide. Finally, we extended the analysis of Ago2 KO cells with a meta-analysis of six datasets of small RNAs extracted from protein immunoprecipitation (IP) of the different Ago proteins. The meta-analysis confirmed the major features identified in Ago2 KO, and identified several additional Ago2 specific miRNA features. Together, the results from **paper IV** indicate that miRNAs associated with Ago2 have distinct structure and sequence specifics, involving base-paring, length and thermodynamics, which all may serve as sorting features for Ago2.

7. Discussion

Coordinated regulation of gene expression is fundamental for the cell to function properly and to prevent disease. Protein production is influenced by the rate of mRNA transcription from a given locus, which again is regulated by transcription factors' access to the chromatin. The mRNAs are subject to regulation by several mechanism including splicing, capping and polyadenylation, which controls stability of the mRNAs. The stability is again challenged by miRNAs that can bind and degrade mRNAs. Several studies have tried to integrate the different regulatory processes to understand when and to what degree they play a role in the cell (318-322), and there are dedicated journals covering this field (323,324). In the current thesis we have used high throughput sequence data to analyze changes in gene expression and we have demonstrated how time-series experiments can be used to identify changes that only become visible over time. The following sections discuss strengths and weaknesses of the methods we have used in the different studies as well as other available methods.

7.1. Does mRNA abundance mirror that of proteins?

In **paper I** and **II** we utilize high throughput gene expression arrays to explain the activity in biological pathways. In **paper II** we show that several DNA repair pathways are associated with S phase regulated genes, including long patch base excision repair, DNA mismatch repair, Fanconi anemia and homologous recombination, and we discuss these observations in the perspective of the role for these pathways in repairing lesions that occur during S phase. Likewise, we suggest that pathways that are not enriched in a particular cell cycle phase repairs lesion independently of cell cycle phases. In **paper I** and **II** we assume that transcript abundance is a major determinant of protein abundance. However, this may not always be the case.

The question of how well changes in mRNA concentration actually reflect the output and activities of the corresponding protein has previously been addressed. Recent advances in technology have made it possible to measure activities at basically every level of regulation, including DNA methylation (microarrays or sequencing), transcription (microarrays or RNA-Seq), translation (ribosomal profiling), protein and protein modification (mass spectrometry). However, some of these technologies have previously been associated with low sensitivity, especially those measuring protein abundance (325). New methods have increased the

sensitivity; however, some challenges still remain. For instance, stable isotope labeling by amino acids in cell culture (SILAC), which measures peptide abundance using mass spectrometry, are only able to measure a subset of the expressed proteins. In case of detecting cell cycle regulated proteins, if SILAC is only able to detect ~2000 proteins (326) and more that 1000 genes are predicted to be cell cycle regulated (306,313,314), many proteins will be missed using this approach However, because different methods are measuring different molecules, it does not always make sense to compare sensitivity without considering the molecules itself.

Studies from bacteria and mammalian cells have reported correlations between protein and mRNA with a Pearson correlation coefficient of ~0.40, thus saying that about 40% of the variation in protein abundance can be explained by variation in mRNA expression. This implies that ~60% of protein variation is due to differences in translation efficiency between mRNAs, post-transcriptional modifications or noise in the measurements (reviewed in (327) and original papers in (328-330)). These studies do not consider differences in the degradation rate between mRNA and protein, and it is known that mRNAs are produced at a much slower rate than proteins in addition to being more prone to degradation (reviewed in (327)). RNAs and metabolites are very stable and have a high protein-per-mRNA ratio, whereas mRNAs from transcription factors, signaling genes, chromatin modifying enzymes and genes with cell-cycle-specific functions are generally more unstable (328). These are genes known to be highly regulated, and higher fluctuations in mRNA levels are expected as a response to stimulus (328).

Other studies have tried to explain the remaining 60% of protein variation using mammalian cells (328,331). The conclusion from these studies was that translation efficiency and protein degradation are the main contributing factors for the observed variation, whereas experimental noise was almost negligible.

Although the correlation between mRNA levels and protein levels is not perfect, measuring the mRNA levels remains a good approximation for the abundance of a protein (332). In a biological system where the protein is either present or non-present at a given time, levels of mRNAs will most likely be the determining factor for protein abundance, which is regulated through transcriptional bursts (327). Such *on* and *off* states of mRNA levels have been described in mammalian and bacterial cells. Transcription factors are thought to be the main player to control the switch in protein abundance, and other regulatory mechanisms like miRNAs and post-transcriptional modifications work as fine tuners of (333,334).

In summary, measuring mRNA levels should give a good estimation of protein activity in closed biological systems where high fluctuations are expected, as for example synchronized cell cultures which are used in the current papers. However, to understand all the changes taking place in the cell, these methods have to be supplemented using matched data sets where multiple levels of regulation are considered in the same samples. In **paper II** we added a translational data set to the analysis. However, in this case more time points and cell lines need to be included to get a better picture of the translational changes in the cell cycle.

7.2. Use of cell synchronization to study cell cycle regulation

The most common way to measure changes in the cell cycle is to use synchronized cell cultures and harvest cells at given time points after the cells have been released from the synchronization block (335). Several such methods have been developed that block cells at mitosis (nocodazole and mitotic shake-off), G_0 phase (serum starvation) or at the G_1/S transition (double thymidine). Even if these are well-established methods, it has been debated whether they are suited to study cell cycle related events (336-341). A controversy between Stephen Cooper and Paul T. Spellmann highlighted problems associated with these methods, and Cooper proposed that whole-culture methods alter the size and genome composition of cell cultures in a way that deviates from the normal cell cycle. Spellmann and Sherlock replied to the critique by listing five arguments proposing that whole-culture synchronization methods are indeed suitable to study cell cycle gene expression (340,341): (i) whole-culture synchronized cells are dividing normally; (ii) The overlap between orthologous genes identified by these methods are good across organisms; (iii) the genes identified as cell cycle regulated frequently regulate cell cycle processes; (iv) studies from whole- and selection-synchronized cultures show good overlap in terms of significant cell cycle genes; (v) individual cells display the same significant cell cycle genes.

One of the main challenges with cell synchronization is the tendency for cells to loose synchronization over time. As seen from the flow cytometry charts in the **Paper I** and, **II**, a substantial fraction of cells have delayed entry into the second cell cycle. Because of this, cell cycle gene expression cannot be monitored for a consecutive number of cell cycles, which would be favorable to eliminate technical noise from the synchronization procedure. Therefore, a combination of synchronization methods are often used which block cells at different cell cycle phases (reviewed in (342)).

When cells are released from the synchronization block, new cell culture medium is added to the cells. This results in rapid activation of growth related genes and a cascade of serum-dependent gene expression is often observed (343). These genes can sometimes be incorrectly annotated as cell cycle regulated. We considered these effects by constructing two different models to monitor cell cycle gene expression. The cell cycle model identifies genes having oscillating expression profiles reminiscent of a sine/cosine curve. The serum-model identifies genes having transient expression changes. Transient expression changes are roughly divided into immediate, early, intermediate and late changes (described in (343)). Using two models, we were in **paper I** and **II** able to separate technical changes from biological changes and thereby reduce the number of false positive cell cycle genes.

7.3. Using time-series data to detect biological changes

Most biological processes are dynamic, meaning that they change over time. A common theme in the present thesis is the use of time-series data to identify dynamic changes in the cell. The common approach has been to use gene expression data, from either microarrays or high-throughput sequencing. In this way, changes can be monitored in a complete set of active genes. Systems that easily produce large populations of synchronized cells are ideal for this type of studies.

When designing a cell cycle synchronization experiment, one important question is how frequent the sampling should be to cover most of the changes. When measuring oscillating processes like the cell cycle, the sampling should be uniform and cover all cell cycle phases (reviewed in (344)). If the synchronization method involves medium replacement, which is the case for most synchronization methods, more sampling should be performed at early time points to control for early transient changes that may resemble cell cycle changes. In **paper I** and **II**, we have considered this in two ways. First by constructing two different PLS models, one model that captures transient serum related changes and one model that captures cell cycle related changes resembling a sine/cosine type of expression. Secondly, we increased the sampling rates for early time points. Separating between cell cycle and serum dependent expression will help identify changes that are directly related to the cell cycle but also changes that are related to proliferation.

Another consideration in time-series experiments is whether to include more replicates or more time points in the experimental setup. More replicates will increase statistical power which makes it easier to identify significant genes at the given time points. Adding more time

points will exclude more false positive cell cycle genes since genes for which the expression spikes at specific time points are removed (344). In **paper I** and **II**, adding more time points may have resulted in more significant transcripts. Another advantage of dense sampling rates is the possibility to identify causal relationships between genes. Amit et al. have demonstrated this when they characterized feedback regulation between transcription factors in EGF stimulated HeLa cells (343). Using 20 minute sampling rates they constructed a kinetic profile of growth factor-induced transcription, demonstrating that transcription takes place in clearly defined waves. We used a similar approach in **paper III** when analyzing growth responses in cells lacking Ago2. Here we showed that Ago2 knock out affects late serum responses while having limited effect on early biological functions. In **paper II**, we used the information on cell cycle regulated genes obtained from **paper I** in addition to public available data sets to map different DNA repair pathways to the cell cycle. Here we showed that genes related to the same DNA repair pathway are often expressed in the same cell cycle phase, indicating that there is a link between expression and function.

In summary, throughout the thesis we have shown how time-series experiments can be used to identify changes in dynamic systems and how transcriptome analysis elucidates interactions between genes in biological systems.

7.4. Challenges with immunoprecipitation and knock-out experiments

In **paper I** and **paper IV** we have analyzed DNA and RNA pulled out from the cell using specific antibodies. Such experiments rely strongly on the specificity of the antibodies used. High quality antibodies are specific to the target protein. Antibodies are classified into monoclonal and polyclonal, depending on the number of epitopes they recognize. Monoclonal antibodies recognize only one epitope on the target protein, whereas polyclonal antibodies recognize multiple epitopes. Because monoclonal antibodies are more specific to the target epitope, they usually have less background compared to polyclonal antibodies. When working on similar proteins, as for example protein homologs, monoclonal antibodies are generally preferred. High antibody specificity is particularly relevant for the Argonaute homologs analyzed in **paper IV**. The specificity of Argonaute antibodies have been discussed by Siomi et al. (345) and monoclonal antibodies against the human and mouse Argonaute2 protein are successfully developed (346,347), as well as for human Ago1 (348), human Ago3 (199) and human Ago4 (206). The availability of these antibodies makes it possible to identify small

RNAs specific for each Ago protein. In **paper IV** we used available Ago IP data sets to look for miRNA sorting in mammalian cells. Most of the studies included in the meta-analysis in **paper IV** fail to observe differences in miRNA association between the Ago proteins; however, they report to have used specific antibodies. Given that the antibodies are specific, one reason why the studies fail to detect Ago specific miRNAs could be lack of replicates in the IP experiment, making it difficult to perform statistical analysis on the data. When replicates are limited, only prominent differences will be detected.

Another way of detecting small RNAs associated with a particular Ago protein is to make Ago deficient cells. In this way, problems with unspecific antibody binding are eliminated. In **paper IV**, we have sequenced small RNAs from Ago2 deficient cells to investigate changes in the population of miRNAs. By looking at which miRNAs are affected by the KO, we may be able to detect potential miRNA preferences for Ago2. However, we need to be aware that Ago2 may have an effect on transcription in general; meaning that when Ago2 is removed from the cell, the transcription of miRNA-coding genes may change. One way to adjust for this is by looking at clustered miRNAs. These effects are discussed in more detail in **paper IV**.

8. Future perspectives

This section discusses the importance of the finding is the papers, in addition to suggestions on new experiments to support and get a deeper understanding of the findings. Some future perspectives are included in the papers and will therefore not be included here.

In **paper I** we show that a subset of Polycomb associated genes is up-regulated during DNA replication, thus providing a model for replication dependent transcription of these genes. Several studies have suggested that non-coding RNAs may play an important role in maintaining Polycomb associated chromatin marks (349-351). Little is known about the role of non-coding RNAs in cell cycle regulation, especially in the context of Polycomb mediated regulation. Furthermore, the exact mechanism of how the cell maintains the correct chromatin marks after DNA replication is not known. We speculate that non-coding RNAs may have a role in this regulation. Experiments are planned to investigate the expression of non-coding RNAs as well as Polycomb associated loci during the cell cycle. Specifically, we will perform IP of the two histone marks H3K27me3 and H3K4me3 at multiple time points during the cell cycle in HaCaT cells. The IP samples will be matched with RNA-Seq samples to detect non-coding RNAs at the same time points.

In **paper II** we investigated the role of DNA repair genes during the cell cycle and show that several DNA repair pathways are associated with specific cell cycle phases. These analyses were based on gene expression data sets, in addition to one data set on ribosomal profiling. Utilizing current methods measuring protein abundance and translation rates would be valuable to support the findings in **paper II**, and give a more complete picture of the network of DNA repair pathways in the cell cycle.

In **paper III and IV** we investigated the role of Argonaute protein 2 (Ago2) and show that knocking out Ago2 from mouse embryonic fibroblasts (MEFs) is associated with major changes in mRNA and miRNA levels. In **paper IV** we observed clear structure and sequence differences between up-and down-regulated miRNAs in Ago2 KO. To test if these structure and sequence features are indeed preferred by Ago2, transfecting RNA constructs harboring these features into Ago2 wild type cells followed by Ago2 IP and deep sequencing, will highlight such preferences. Finally, expressing the full-length Ago2 protein in MEF Ago2 knock-out cells will tell if the time dependent changes we observed in **paper III** were Ago2

specific. Similar, expressing a catalytic mutant of Ago2 would test if the changes in *Hox* gene expression were due to Ago2 cleavage, as proposed in **paper III**.

Recent publications have shown that Ago2 has a specific role in recruiting repair factors to DSBs, guided by a class of small RNAs termed DSB-induced small RNAs (diRNAs) produced in the vicinity of these sites (9,352). Other Ago proteins fail to do so, indicating that the recruitment is dependent on the catalytic activity of Ago2 (220,353). These findings show that the role of Ago2 goes beyond miRNA mediated gene repression, and small RNAs seem to have a conserved role in DSB repair. Further studies should be conducted to show how and where diRNAs are transcribed in the genome and why they are specific to Ago2.

9. References

1. Rokas, A. (2008) The molecular origins of multicellular transitions. *Current opinion in genetics & development*, **18**, 472-478.
2. Gregory, T.R. (2005) Synergy between sequence and size in large-scale genomics. *Nature reviews. Genetics*, **6**, 699-708.
3. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843-854.
4. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853-858.
5. Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*, **294**, 858-862.
6. Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in Caenorhabditis elegans. *Science*, **294**, 862-864.
7. Peterson, C.L. and Herskowitz, I. (1992) Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. *Cell*, **68**, 573-583.
8. Francia, S., Michelini, F., Saxena, A., Tang, D., de Hoon, M., Anelli, V., Mione, M., Carninci, P. and d'Adda di Fagagna, F. (2012) Site-specific DICER and DROSHA RNA products control the DNA-damage response. *Nature*, **488**, 231-235.
9. Wei, W., Ba, Z., Gao, M., Wu, Y., Ma, Y., Amiard, S., White, C.I., Rendtlew Danielsen, J.M., Yang, Y.G. and Qi, Y. (2012) A role for small RNAs in DNA double-strand break repair. *Cell*, **149**, 101-112.
10. Ciccia, A. and Elledge, S.J. (2010) The DNA damage response: making it safe to play with knives. *Molecular cell*, **40**, 179-204.
11. Bennett, G., Papamichos-Chronakis, M. and Peterson, C.L. (2013) DNA repair choice defines a common pathway for recruitment of chromatin regulators. *Nature communications*, **4**, 2084.
12. Delacote, F. and Lopez, B.S. (2008) Importance of the cell cycle phase for the choice of the appropriate DSB repair pathway, for genome stability maintenance: the trans-S double-strand break repair model. *Cell Cycle*, **7**, 33-38.
13. Escribano-Diaz, C., Orthwein, A., Fradet-Turcotte, A., Xing, M., Young, J.T., Tkac, J., Cook, M.A., Rosebrock, A.P., Munro, M., Canny, M.D. *et al.* (2013) A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Molecular cell*, **49**, 872-883.
14. Kato, K., Nakajima, K., Ui, A., Muto-Terao, Y., Ogiwara, H. and Nakada, S. (2014) Fine-tuning of DNA damage-dependent ubiquitination by OTUB2 supports the DNA repair pathway choice. *Molecular cell*, **53**, 617-630.
15. Tomimatsu, N., Mukherjee, B., Catherine Hardebeck, M., Ilcheva, M., Vanessa Camacho, C., Louise Harris, J., Porteus, M., Llorente, B., Khanna, K.K. and Burma, S. (2014) Phosphorylation of EXO1 by CDKs 1 and 2 regulates DNA end resection and repair pathway choice. *Nature communications*, **5**, 3561.
16. Krokan, H.E. and Bjoras, M. (2013) Base excision repair. *Cold Spring Harbor perspectives in biology*, **5**, a012583.

17. Sancar, A., Lindsey-Boltz, L.A., Unsal-Kacmaz, K. and Linn, S. (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual review of biochemistry*, **73**, 39-85.
18. Kraemer, K.H., Lee, M.M., Andrews, A.D. and Lambert, W.C. (1994) The role of sunlight and DNA repair in melanoma and nonmelanoma skin cancer. The xeroderma pigmentosum paradigm. *Archives of dermatology*, **130**, 1018-1021.
19. Yang, Z., Roginskaya, M., Colis, L.C., Basu, A.K., Shell, S.M., Liu, Y., Musich, P.R., Harris, C.M., Harris, T.M. and Zou, Y. (2006) Specific and efficient binding of xeroderma pigmentosum complementation group A to double-strand/single-strand DNA junctions with 3'- and/or 5'-ssDNA branches. *Biochemistry*, **45**, 15921-15930.
20. Batty, D.P. and Wood, R.D. (2000) Damage recognition in nucleotide excision repair of DNA. *Gene*, **241**, 193-204.
21. Diderich, K., Alanazi, M. and Hoeijmakers, J.H. (2011) Premature aging and cancer in nucleotide excision repair-disorders. *DNA repair*, **10**, 772-780.
22. Sugasawa, K., Akagi, J., Nishi, R., Iwai, S. and Hanaoka, F. (2009) Two-step recognition of DNA damage for mammalian nucleotide excision repair: Directional binding of the XPC complex and DNA strand scanning. *Molecular cell*, **36**, 642-653.
23. Min, J.H. and Pavletich, N.P. (2007) Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature*, **449**, 570-575.
24. Kamiuchi, S., Saijo, M., Citterio, E., de Jager, M., Hoeijmakers, J.H. and Tanaka, K. (2002) Translocation of Cockayne syndrome group A protein to the nuclear matrix: possible relevance to transcription-coupled DNA repair. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 201-206.
25. Fousteri, M., Vermeulen, W., van Zeeland, A.A. and Mullenders, L.H. (2006) Cockayne syndrome A and B proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled RNA polymerase II in vivo. *Molecular cell*, **23**, 471-482.
26. Mocquet, V., Laine, J.P., Riedl, T., Yajin, Z., Lee, M.Y. and Egly, J.M. (2008) Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *The EMBO journal*, **27**, 155-167.
27. Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L.H. and Fousteri, M.I. (2007) Sealing of chromosomal DNA nicks during nucleotide excision repair requires XRCC1 and DNA ligase III alpha in a cell-cycle-specific manner. *Molecular cell*, **27**, 311-323.
28. Fagbemi, A.F., Orelli, B. and Scharer, O.D. (2011) Regulation of endonuclease activity in human nucleotide excision repair. *DNA repair*, **10**, 722-729.
29. Marra, G., Chang, C.L., Laghi, L.A., Chauhan, D.P., Young, D. and Boland, C.R. (1996) Expression of human MutS homolog 2 (hMSH2) protein in resting and proliferating cells. *Oncogene*, **13**, 2189-2196.
30. Jiricny, J. (2006) The multifaceted mismatch-repair system. *Nature reviews. Molecular cell biology*, **7**, 335-346.
31. Lau, P.J. and Kolodner, R.D. (2003) Transfer of the MSH2.MSH6 complex from proliferating cell nuclear antigen to mispaired bases in DNA. *The Journal of biological chemistry*, **278**, 14-17.
32. Shell, S.S., Putnam, C.D. and Kolodner, R.D. (2007) The N terminus of Saccharomyces cerevisiae Msh6 is an unstructured tether to PCNA. *Molecular cell*, **26**, 565-578.
33. Pluciennik, A., Dzantiev, L., Iyer, R.R., Constantin, N., Kadyrov, F.A. and Modrich, P. (2010) PCNA function in the activation and strand direction of MutLalpha

endonuclease in mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 16066-16071.
34. Genschel, J. and Modrich, P. (2009) Functions of MutLalpha, replication protein A (RPA), and HMGB1 in 5'-directed mismatch repair. *The Journal of biological chemistry*, **284**, 21536-21544.
35. Schmutte, C., Marinescu, R.C., Sadoff, M.M., Guerrette, S., Overhauser, J. and Fishel, R. (1998) Human exonuclease I interacts with the mismatch repair protein hMSH2. *Cancer research*, **58**, 4537-4542.
36. Schmutte, C., Sadoff, M.M., Shim, K.S., Acharya, S. and Fishel, R. (2001) The interaction of DNA mismatch repair proteins with human exonuclease I. *The Journal of biological chemistry*, **276**, 33011-33018.
37. Kottemann, M.C. and Smogorzewska, A. (2013) Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature*, **493**, 356-363.
38. Wood, R.D. (2010) Mammalian nucleotide excision repair proteins and interstrand crosslink repair. *Environmental and molecular mutagenesis*, **51**, 520-526.
39. Hinz, J.M. (2010) Role of homologous recombination in DNA interstrand crosslink repair. *Environmental and molecular mutagenesis*, **51**, 582-603.
40. Taniguchi, T., Garcia-Higuera, I., Xu, B., Andreassen, P.R., Gregory, R.C., Kim, S.T., Lane, W.S., Kastan, M.B. and D'Andrea, A.D. (2002) Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways. *Cell*, **109**, 459-472.
41. Andreassen, P.R., D'Andrea, A.D. and Taniguchi, T. (2004) ATR couples FANCD2 monoubiquitination to the DNA-damage response. *Genes & development*, **18**, 1958-1963.
42. Moore, J.K. and Haber, J.E. (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in Saccharomyces cerevisiae. *Molecular and cellular biology*, **16**, 2164-2173.
43. McVey, M. and Lee, S.E. (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends in genetics : TIG*, **24**, 529-538.
44. Heyer, W.D., Ehmsen, K.T. and Liu, J. (2010) Regulation of homologous recombination in eukaryotes. *Annual review of genetics*, **44**, 113-139.
45. Karanam, K., Kafri, R., Loewer, A. and Lahav, G. (2012) Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. *Molecular cell*, **47**, 320-329.
46. Zhang, Y., Shim, E.Y., Davis, M. and Lee, S.E. (2009) Regulation of repair choice: Cdk1 suppresses recruitment of end joining factors at DNA breaks. *DNA repair*, **8**, 1235-1241.
47. Mao, Z., Bozzella, M., Seluanov, A. and Gorbunova, V. (2008) DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle*, **7**, 2902-2906.
48. Franchitto, A., Pichierri, P., Piergentili, R., Crescenzi, M., Bignami, M. and Palitti, F. (2003) The mammalian mismatch repair protein MSH2 is required for correct MRE11 and RAD51 relocalization and for efficient cell cycle arrest induced by ionizing radiation in G2 phase. *Oncogene*, **22**, 2110-2120.
49. Barlow, J.H., Lisby, M. and Rothstein, R. (2008) Differential regulation of the cellular response to DNA double-strand breaks in G1. *Molecular cell*, **30**, 73-85.
50. Jazayeri, A., Falck, J., Lukas, C., Bartek, J., Smith, G.C., Lukas, J. and Jackson, S.P. (2006) ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. *Nature cell biology*, **8**, 37-45.
51. West, S.C. (2003) Molecular views of recombination proteins and their control. *Nature reviews. Molecular cell biology*, **4**, 435-445.

52. San Filippo, J., Sung, P. and Klein, H. (2008) Mechanism of eukaryotic homologous recombination. *Annual review of biochemistry*, **77**, 229-257.
53. Yi, C. and He, C. (2013) DNA repair by reversal of DNA damage. *Cold Spring Harbor perspectives in biology*, **5**, a012575.
54. Eker, A.P., Quayle, C., Chaves, I. and van der Horst, G.T. (2009) DNA repair in mammalian cells: Direct DNA damage reversal: elegant solutions for nasty problems. *Cellular and molecular life sciences : CMLS*, **66**, 968-980.
55. Essen, L.O. and Klar, T. (2006) Light-driven DNA repair by photolyases. *Cellular and molecular life sciences : CMLS*, **63**, 1266-1277.
56. Carell, T., Burgdorf, L.T., Kundu, L.M. and Cichon, M. (2001) The mechanism of action of DNA photolyases. *Current opinion in chemical biology*, **5**, 491-498.
57. Kornberg, R.D. (1977) Structure of chromatin. *Annual review of biochemistry*, **46**, 931-954.
58. (1972) DNA structure. Geometrical versatility. *Nature*, **238**, 15-16.
59. Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G. and Zhao, K. (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell*, **132**, 887-898.
60. Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J. and Rando, O.J. (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. *Science*, **309**, 626-630.
61. Valouev, A., Johnson, S.M., Boyd, S.D., Smith, C.L., Fire, A.Z. and Sidow, A. (2011) Determinants of nucleosome organization in primary human cells. *Nature*, **474**, 516-520.
62. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B. *et al.* (2012) The accessible chromatin landscape of the human genome. *Nature*, **489**, 75-82.
63. Ahmad, K. and Henikoff, S. (2002) Histone H3 variants specify modes of chromatin assembly. *Proceedings of the National Academy of Sciences of the United States of America*, **99 Suppl 4**, 16477-16484.
64. Kamakaka, R.T. and Biggins, S. (2005) Histone variants: deviants? *Genes & development*, **19**, 295-310.
65. Probst, A.V., Dunleavy, E. and Almouzni, G. (2009) Epigenetic inheritance during the cell cycle. *Nature reviews. Molecular cell biology*, **10**, 192-206.
66. Falbo, K.B. and Shen, X. (2012) Function of the INO80 chromatin remodeling complex in DNA replication. *Front Biosci (Landmark Ed)*, **17**, 970-975.
67. Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science*, **293**, 1074-1080.
68. Strahl, B.D. and Allis, C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41-45.
69. Saha, A., Wittmeyer, J. and Cairns, B.R. (2002) Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes & development*, **16**, 2120-2134.
70. Bao, Y. and Shen, X. (2011) SnapShot: Chromatin remodeling: INO80 and SWR1. *Cell*, **144**, 158-158 e152.
71. Saha, A., Wittmeyer, J. and Cairns, B.R. (2006) Chromatin remodelling: the industrial revolution of DNA around histones. *Nature reviews. Molecular cell biology*, **7**, 437-447.
72. Grune, T., Brzeski, J., Eberharter, A., Clapier, C.R., Corona, D.F., Becker, P.B. and Muller, C.W. (2003) Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Molecular cell*, **12**, 449-460.

73. Corona, D.F. and Tamkun, J.W. (2004) Multiple roles for ISWI in transcription, chromosome organization and DNA replication. *Biochimica et biophysica acta*, **1677**, 113-119.
74. Deuring, R., Fanti, L., Armstrong, J.A., Sarte, M., Papoulas, O., Prestel, M., Daubresse, G., Verardo, M., Moseley, S.L., Berloco, M. *et al.* (2000) The ISWI chromatin-remodeling protein is required for gene expression and the maintenance of higher order chromatin structure in vivo. *Molecular cell*, **5**, 355-365.
75. Goldmark, J.P., Fazzio, T.G., Estep, P.W., Church, G.M. and Tsukiyama, T. (2000) The Isw2 chromatin remodeling complex represses early meiotic genes upon recruitment by Ume6p. *Cell*, **103**, 423-433.
76. Martens, J.A. and Winston, F. (2003) Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Current opinion in genetics & development*, **13**, 136-142.
77. Morillon, A., Karabetsou, N., O'Sullivan, J., Kent, N., Proudfoot, N. and Mellor, J. (2003) Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. *Cell*, **115**, 425-435.
78. Martens, J.A. and Winston, F. (2002) Evidence that Swi/Snf directly represses transcription in S. cerevisiae. *Genes & development*, **16**, 2231-2236.
79. Wu, W.H., Alami, S., Luk, E., Wu, C.H., Sen, S., Mizuguchi, G., Wei, D. and Wu, C. (2005) Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. *Nature structural & molecular biology*, **12**, 1064-1071.
80. Shen, X., Mizuguchi, G., Hamiche, A. and Wu, C. (2000) A chromatin remodelling complex involved in transcription and DNA processing. *Nature*, **406**, 541-544.
81. Conaway, R.C. and Conaway, J.W. (2009) The INO80 chromatin remodeling complex in transcription, replication and repair. *Trends in biochemical sciences*, **34**, 71-77.
82. Jonsson, Z.O., Jha, S., Wohlschlegel, J.A. and Dutta, A. (2004) Rvb1p/Rvb2p recruit Arp5p and assemble a functional Ino80 chromatin remodeling complex. *Molecular cell*, **16**, 465-477.
83. Fritsch, O., Benvenuto, G., Bowler, C., Molinier, J. and Hohn, B. (2004) The INO80 protein controls homologous recombination in Arabidopsis thaliana. *Molecular cell*, **16**, 479-485.
84. Cai, Y., Jin, J., Yao, T., Gottschalk, A.J., Swanson, S.K., Wu, S., Shi, Y., Washburn, M.P., Florens, L., Conaway, R.C. *et al.* (2007) YY1 functions with INO80 to activate transcription. *Nature structural & molecular biology*, **14**, 872-874.
85. Gerhold, C.B. and Gasser, S.M. (2014) INO80 and SWR complexes: relating structure to function in chromatin remodeling. *Trends in cell biology*.
86. Morrison, A.J., Kim, J.A., Person, M.D., Highland, J., Xiao, J., Wehr, T.S., Hensley, S., Bao, Y., Shen, J., Collins, S.R. *et al.* (2007) Mec1/Tel1 phosphorylation of the INO80 chromatin remodeling complex influences DNA damage checkpoint responses. *Cell*, **130**, 499-511.
87. Steger, D.J., Haswell, E.S., Miller, A.L., Wente, S.R. and O'Shea, E.K. (2003) Regulation of chromatin remodeling by inositol polyphosphates. *Science*, **299**, 114-116.
88. Morrison, A.J., Highland, J., Krogan, N.J., Arbel-Eden, A., Greenblatt, J.F., Haber, J.E. and Shen, X. (2004) INO80 and gamma-H2AX interaction links ATP-dependent chromatin remodeling to DNA damage repair. *Cell*, **119**, 767-775.
89. Okuno, Y., Satoh, H., Sekiguchi, M. and Masukata, H. (1999) Clustered adenine/thymine stretches are essential for function of a fission yeast replication origin. *Molecular and cellular biology*, **19**, 6699-6709.

90. Papamichos-Chronakis, M. and Peterson, C.L. (2008) The Ino80 chromatin-remodeling enzyme regulates replisome function and stability. *Nature structural & molecular biology*, **15**, 338-345.
91. Lopes, M., Cotta-Ramusino, C., Pellicioli, A., Liberi, G., Plevani, P., Muzi-Falconi, M., Newlon, C.S. and Foiani, M. (2001) The DNA replication checkpoint response stabilizes stalled replication forks. *Nature*, **412**, 557-561.
92. Bao, Y. and Shen, X. (2007) INO80 subfamily of chromatin remodeling complexes. *Mutation research*, **618**, 18-29.
93. Rubbi, C.P. and Milner, J. (2003) p53 is a chromatin accessibility factor for nucleotide excision repair of DNA damage. *The EMBO journal*, **22**, 975-986.
94. Goodarzi, A.A., Noon, A.T., Deckbar, D., Ziv, Y., Shiloh, Y., Lobrich, M. and Jeggo, P.A. (2008) ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Molecular cell*, **31**, 167-177.
95. Lorat, Y., Schanz, S., Schuler, N., Wennemuth, G., Rube, C. and Rube, C.E. (2012) Beyond repair foci: DNA double-strand break repair in euchromatic and heterochromatic compartments analyzed by transmission electron microscopy. *PloS one*, **7**, e38165.
96. Chapman, J.R., Taylor, M.R. and Boulton, S.J. (2012) Playing the end game: DNA double-strand break repair pathway choice. *Molecular cell*, **47**, 497-510.
97. Goodarzi, A.A., Jeggo, P. and Lobrich, M. (2010) The influence of heterochromatin on DNA double strand break repair: Getting the strong, silent type to relax. *DNA repair*, **9**, 1273-1282.
98. Symington, L.S. and Gautier, J. (2011) Double-strand break end resection and repair pathway choice. *Annual review of genetics*, **45**, 247-271.
99. Stanley, F.K., Moore, S. and Goodarzi, A.A. (2013) CHD chromatin remodelling enzymes and the DNA damage response. *Mutation research*, **750**, 31-44.
100. Kruhlak, M.J., Celeste, A., Dellaire, G., Fernandez-Capetillo, O., Muller, W.G., McNally, J.G., Bazett-Jones, D.P. and Nussenzweig, A. (2006) Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. *The Journal of cell biology*, **172**, 823-834.
101. Bao, Y. and Shen, X. (2007) Chromatin remodeling in DNA double-strand break repair. *Current opinion in genetics & development*, **17**, 126-131.
102. Bakkenist, C.J. and Kastan, M.B. (2003) DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature*, **421**, 499-506.
103. Noon, A.T., Shibata, A., Rief, N., Lobrich, M., Stewart, G.S., Jeggo, P.A. and Goodarzi, A.A. (2010) 53BP1-dependent robust localized KAP-1 phosphorylation is essential for heterochromatic DNA double-strand break repair. *Nature cell biology*, **12**, 177-184.
104. Lavin, M.F., Birrell, G., Chen, P., Kozlov, S., Scott, S. and Gueven, N. (2005) ATM signaling and genomic stability in response to DNA damage. *Mutation research*, **569**, 123-132.
105. Alvarez-Quilon, A., Serrano-Benitez, A., Lieberman, J.A., Quintero, C., Sanchez-Gutierrez, D., Escudero, L.M. and Cortes-Ledesma, F. (2014) ATM specifically mediates repair of double-strand breaks with blocked DNA ends. *Nature communications*, **5**, 3347.
106. Lobrich, M. and Jeggo, P.A. (2005) The two edges of the ATM sword: co-operation between repair and checkpoint functions. *Radiother Oncol*, **76**, 112-118.
107. van Attikum, H., Fritsch, O., Hohn, B. and Gasser, S.M. (2004) Recruitment of the INO80 complex by H2A phosphorylation links ATP-dependent chromatin remodeling with DNA double-strand break repair. *Cell*, **119**, 777-788.

108. Tsukuda, T., Fleming, A.B., Nickoloff, J.A. and Osley, M.A. (2005) Chromatin remodelling at a DNA double-strand break site in Saccharomyces cerevisiae. *Nature*, **438**, 379-383.
109. Yu, Y., Deng, Y., Reed, S.H., Millar, C.B. and Waters, R. (2013) Histone variant Htz1 promotes histone H3 acetylation to enhance nucleotide excision repair in Htz1 nucleosomes. *Nucleic acids research*, **41**, 9006-9019.
110. Reed, S.H. (2011) Nucleotide excision repair in chromatin: damage removal at the drop of a HAT. *DNA repair*, **10**, 734-742.
111. Palomera-Sanchez, Z. and Zurita, M. (2011) Open, repair and close again: chromatin dynamics and the response to UV-induced DNA damage. *DNA repair*, **10**, 119-125.
112. Yu, Y. and Waters, R. (2005) Histone acetylation, chromatin remodelling and nucleotide excision repair: hint from the study on MFA2 in Saccharomyces cerevisiae. *Cell Cycle*, **4**, 1043-1045.
113. Scrima, A., Konickova, R., Czyzewski, B.K., Kawasaki, Y., Jeffrey, P.D., Groisman, R., Nakatani, Y., Iwai, S., Pavletich, N.P. and Thoma, N.H. (2008) Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell*, **135**, 1213-1223.
114. Hanawalt, P.C. (2002) Subpathways of nucleotide excision repair and their regulation. *Oncogene*, **21**, 8949-8956.
115. Duan, M.R. and Smerdon, M.J. (2010) UV damage in DNA promotes nucleosome unwrapping. *The Journal of biological chemistry*, **285**, 26295-26303.
116. Bohr, V.A., Smith, C.A., Okumoto, D.S. and Hanawalt, P.C. (1985) DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell*, **40**, 359-369.
117. Guerrero-Santoro, J., Kapetanaki, M.G., Hsieh, C.L., Gorbachinsky, I., Levine, A.S. and Rapic-Otrin, V. (2008) The cullin 4B-based UV-damaged DNA-binding protein ligase binds to UV-damaged chromatin and ubiquitinates histone H2A. *Cancer research*, **68**, 5014-5022.
118. Wang, H., Zhai, L., Xu, J., Joo, H.Y., Jackson, S., Erdjument-Bromage, H., Tempst, P., Xiong, Y. and Zhang, Y. (2006) Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Molecular cell*, **22**, 383-394.
119. Hara, R. and Sancar, A. (2003) Effect of damage type on stimulation of human excision nuclease by SWI/SNF chromatin remodeling factor. *Molecular and cellular biology*, **23**, 4121-4125.
120. Jiang, Y., Wang, X., Bao, S., Guo, R., Johnson, D.G., Shen, X. and Li, L. (2010) INO80 chromatin remodeling complex promotes the removal of UV lesions by the nucleotide excision repair pathway. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 17274-17279.
121. Zhao, Q., Wang, Q.E., Ray, A., Wani, G., Han, C., Milum, K. and Wani, A.A. (2009) Modulation of nucleotide excision repair by mammalian SWI/SNF chromatin-remodeling complex. *The Journal of biological chemistry*, **284**, 30424-30432.
122. Smerdon, M.J., Lan, S.Y., Calza, R.E. and Reeves, R. (1982) Sodium butyrate stimulates DNA repair in UV-irradiated normal and xeroderma pigmentosum human fibroblasts. *The Journal of biological chemistry*, **257**, 13441-13447.
123. Ramanathan, B. and Smerdon, M.J. (1989) Enhanced DNA repair synthesis in hyperacetylated nucleosomes. *The Journal of biological chemistry*, **264**, 11026-11034.
124. Di Croce, L. and Helin, K. (2013) Transcriptional regulation by Polycomb group proteins. *Nature structural & molecular biology*, **20**, 1147-1155.
125. Kerppola, T.K. (2009) Polycomb group complexes--many combinations, many functions. *Trends in cell biology*, **19**, 692-704.

126. Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D. and Khorasanizadeh, S. (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes & development*, **17**, 1870-1881.
127. Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S. and Zhang, Y. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science*, **298**, 1039-1043.
128. Pasini, D., Bracken, A.P., Hansen, J.B., Capillo, M. and Helin, K. (2007) The polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Molecular and cellular biology*, **27**, 3769-3779.
129. Endoh, M., Endo, T.A., Endoh, T., Fujimura, Y., Ohara, O., Toyoda, T., Otte, A.P., Okano, M., Brockdorff, N., Vidal, M. *et al.* (2008) Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development*, **135**, 1513-1524.
130. Bracken, A.P. and Helin, K. (2009) Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nature reviews. Cancer*, **9**, 773-784.
131. Dietrich, N., Bracken, A.P., Trinh, E., Schjerling, C.K., Koseki, H., Rappsilber, J., Helin, K. and Hansen, K.H. (2007) Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus. *The EMBO journal*, **26**, 1637-1648.
132. Gil, J. and Peters, G. (2006) Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. *Nature reviews. Molecular cell biology*, **7**, 667-677.
133. Lowe, S.W. and Sherr, C.J. (2003) Tumor suppression by Ink4a-Arf: progress and puzzles. *Current opinion in genetics & development*, **13**, 77-83.
134. Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I. *et al.* (2007) Epigenetic stem cell signature in cancer. *Nature genetics*, **39**, 157-158.
135. Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E. *et al.* (2007) Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nature genetics*, **39**, 232-236.
136. Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W. *et al.* (2007) A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature genetics*, **39**, 237-242.
137. Bracken, A.P., Kleine-Kohlbrecher, D., Dietrich, N., Pasini, D., Gargiulo, G., Beekman, C., Theilgaard-Monch, K., Minucci, S., Porse, B.T., Marine, J.C. *et al.* (2007) The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes & development*, **21**, 525-530.
138. Sparmann, A. and van Lohuizen, M. (2006) Polycomb silencers control cell fate, development and cancer. *Nature reviews. Cancer*, **6**, 846-856.
139. Levine, M. and Tjian, R. (2003) Transcription regulation and animal diversity. *Nature*, **424**, 147-151.
140. Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**, 1311-1323.
141. Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E. and Chang, H.Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, **329**, 689-693.

142. Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A. *et al.* (2009) Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 11667-11672.
143. Cech, T.R. and Steitz, J.A. (2014) The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*, **157**, 77-94.
144. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281-297.
145. Griffiths-Jones, S. (2004) The microRNA Registry. *Nucleic acids research*, **32**, D109-111.
146. Friedman, R.C., Farh, K.K., Burge, C.B. and Bartel, D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome research*, **19**, 92-105.
147. Wightman, B., Ha, I. and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, **75**, 855-862.
148. Lee, C.T., Risom, T. and Strauss, W.M. (2007) Evolutionary conservation of microRNA regulatory circuits: an examination of microRNA gene complexity and conserved microRNA-target interactions through metazoan phylogeny. *DNA and cell biology*, **26**, 209-218.
149. Heimberg, A.M., Sempere, L.F., Moy, V.N., Donoghue, P.C. and Peterson, K.J. (2008) MicroRNAs and the advent of vertebrate morphological complexity. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 2946-2950.
150. Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F. and Hannon, G.J. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature*, **432**, 231-235.
151. Auyeung, V.C., Ulitsky, I., McGeary, S.E. and Bartel, D.P. (2013) Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell*, **152**, 844-858.
152. Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T. and Kim, V.N. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**, 887-901.
153. Chen, C.Z., Li, L., Lodish, H.F. and Bartel, D.P. (2004) MicroRNAs modulate hematopoietic lineage differentiation. *Science*, **303**, 83-86.
154. Ruby, J.G., Jan, C.H. and Bartel, D.P. (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature*, **448**, 83-86.
155. Yi, R., Qin, Y., Macara, I.G. and Cullen, B.R. (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & development*, **17**, 3011-3016.
156. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. and Kutay, U. (2004) Nuclear export of microRNA precursors. *Science*, **303**, 95-98.
157. Bohnsack, M.T., Czaplinski, K. and Gorlich, D. (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*, **10**, 185-191.
158. Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D. and Doudna, J.A. (2006) Structural basis for double-stranded RNA processing by Dicer. *Science*, **311**, 195-198.

159. MacRae, I.J., Zhou, K. and Doudna, J.A. (2007) Structural determinants of RNA recognition and cleavage by Dicer. *Nature structural & molecular biology*, **14**, 934-940.
160. Park, J.E., Heo, I., Tian, Y., Simanshu, D.K., Chang, H., Jee, D., Patel, D.J. and Kim, V.N. (2011) Dicer recognizes the 5' end of RNA for efficient and accurate processing. *Nature*, **475**, 201-205.
161. Gu, S., Jin, L., Zhang, Y., Huang, Y., Zhang, F., Valdmanis, P.N. and Kay, M.A. (2012) The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo. *Cell*, **151**, 900-911.
162. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215-233.
163. Cheloufi, S., Dos Santos, C.O., Chong, M.M. and Hannon, G.J. (2010) A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature*, **465**, 584-589.
164. Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D. *et al.* (2010) A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science*, **328**, 1694-1698.
165. Fang, Z., Du, R., Edwards, A., Flemington, E.K. and Zhang, K. (2013) The sequence structures of human microRNA molecules and their implications. *PloS one*, **8**, e54215.
166. Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P. and Bartel, D.P. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular cell*, **27**, 91-105.
167. Saetrom, P., Heale, B.S., Snove, O., Jr., Aagaard, L., Alluin, J. and Rossi, J.J. (2007) Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic acids research*, **35**, 2333-2342.
168. Shin, C., Nam, J.W., Farh, K.K., Chiang, H.R., Shkumatava, A. and Bartel, D.P. (2010) Expanding the microRNA targeting code: functional sites with centered pairing. *Molecular cell*, **38**, 789-802.
169. Schirle, N.T., Sheu-Gruttadauria, J. and MacRae, I.J. (2014) Gene regulation. Structural basis for microRNA targeting. *Science*, **346**, 608-613.
170. Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15-20.
171. Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R. and Bartel, D.P. (2009) RNAi in budding yeast. *Science*, **326**, 544-550.
172. Carmell, M.A., Xuan, Z., Zhang, M.Q. and Hannon, G.J. (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes & development*, **16**, 2733-2742.
173. Sasaki, T., Shiohama, A., Minoshima, S. and Shimizu, N. (2003) Identification of eight members of the Argonaute family in the human genome. *Genomics*, **82**, 323-330.
174. Jinek, M. and Doudna, J.A. (2009) A three-dimensional view of the molecular machinery of RNA interference. *Nature*, **457**, 405-412.
175. Kwak, P.B. and Tomari, Y. (2012) The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nature structural & molecular biology*, **19**, 145-151.
176. Elkayam, E., Kuhn, C.D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J. and Joshua-Tor, L. (2012) The structure of human argonaute-2 in complex with miR-20a. *Cell*, **150**, 100-110.

177. Khvorova, A., Reynolds, A. and Jayasena, S.D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell*, **115**, 209-216.
178. Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N. and Zamore, P.D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, **115**, 199-208.
179. Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., Jantsch, M.F., Konrat, R., Blasi, U. and Schroeder, R. (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA biology*, **4**, 118-130.
180. Matranga, C., Tomari, Y., Shin, C., Bartel, D.P. and Zamore, P.D. (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell*, **123**, 607-620.
181. Shin, C. (2008) Cleavage of the star strand facilitates assembly of some microRNAs into Ago2-containing silencing complexes in mammals. *Molecules and cells*, **26**, 308-313.
182. Ye, X., Huang, N., Liu, Y., Paroo, Z., Huerta, C., Li, P., Chen, S., Liu, Q. and Zhang, H. (2011) Structure of C3PO and mechanism of human RISC activation. *Nature structural & molecular biology*, **18**, 650-657.
183. Liu, Y., Ye, X., Jiang, F., Liang, C., Chen, D., Peng, J., Kinch, L.N., Grishin, N.V. and Liu, Q. (2009) C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. *Science*, **325**, 750-753.
184. Steiner, F.A., Okihara, K.L., Hoogstrate, S.W., Sijen, T. and Ketting, R.F. (2009) RDE-1 slicer activity is required only for passenger-strand cleavage during RNAi in Caenorhabditis elegans. *Nature structural & molecular biology*, **16**, 207-211.
185. Burroughs, A.M., Ando, Y., de Hoon, M.J., Tomaru, Y., Suzuki, H., Hayashizaki, Y. and Daub, C.O. (2011) Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. *RNA biology*, **8**, 158-177.
186. Dueck, A., Ziegler, C., Eichner, A., Berezikov, E. and Meister, G. (2012) microRNAs associated with the different human Argonaute proteins. *Nucleic acids research*, **40**, 9850-9862.
187. Wang, D., Zhang, Z., O'Loughlin, E., Lee, T., Houel, S., O'Carroll, D., Tarakhovsky, A., Ahn, N.G. and Yi, R. (2012) Quantitative functions of Argonaute proteins in mammalian development. *Genes & development*, **26**, 693-704.
188. Juvvuna, P.K., Khandelia, P., Lee, L.M. and Makeyev, E.V. (2012) Argonaute identity defines the length of mature mammalian microRNAs. *Nucleic acids research*, **40**, 6808-6820.
189. Ghildiyal, M., Xu, J., Seitz, H., Weng, Z. and Zamore, P.D. (2010) Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA*, **16**, 43-56.
190. Hu, H.Y., Yan, Z., Xu, Y., Hu, H., Menzel, C., Zhou, Y.H., Chen, W. and Khaitovich, P. (2009) Sequence features associated with microRNA strand selection in humans and flies. *BMC genomics*, **10**, 413.
191. Frank, F., Sonenberg, N. and Nagar, B. (2010) Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature*, **465**, 818-822.
192. Wang, Y., Juranek, S., Li, H., Sheng, G., Wardle, G.S., Tuschl, T. and Patel, D.J. (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature*, **461**, 754-761.
193. Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C. *et al.* (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide. *Cell*, **133**, 116-127.

194. Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E. and Carrington, J.C. (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell*, **133**, 128-141.
195. Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M. and Watanabe, Y. (2008) The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. *Plant & cell physiology*, **49**, 493-500.
196. Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R. *et al.* (2008) An endogenous small interfering RNA pathway in Drosophila. *Nature*, **453**, 798-802.
197. Valdmanis, P.N., Gu, S., Schuermann, N., Sethupathy, P., Grimm, D. and Kay, M.A. (2012) Expression determinants of mammalian argonaute proteins in mediating gene silencing. *Nucleic acids research*, **40**, 3704-3713.
198. Marie Lundbæk, R.M., Pål Sætrom. (2014) Argonaute2 affects Hox gene expression and the late growth response in mouse embryonic fibroblasts. *Submitted*.
199. Petri, S., Dueck, A., Lehmann, G., Putz, N., Rudel, S., Kremmer, E. and Meister, G. (2011) Increased siRNA duplex stability correlates with reduced off-target and elevated on-target effects. *RNA*, **17**, 737-749.
200. Meister, G. (2013) Argonaute proteins: functional insights and emerging roles. *Nature reviews. Genetics*, **14**, 447-459.
201. Morita, S., Horii, T., Kimura, M., Goto, Y., Ochiya, T. and Hatada, I. (2007) One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation. *Genomics*, **89**, 687-696.
202. Deshpande, G., Calhoun, G. and Schedl, P. (2005) Drosophila argonaute-2 is required early in embryogenesis for the assembly of centric/centromeric heterochromatin, nuclear division, nuclear migration, and germ-cell formation. *Genes & development*, **19**, 1680-1685.
203. Nan, Y., Han, L., Zhang, A., Wang, G., Jia, Z., Yang, Y., Yue, X., Pu, P., Zhong, Y. and Kang, C. (2010) MiRNA-451 plays a role as tumor suppressor in human glioma cells. *Brain research*, **1359**, 14-21.
204. Chu, Y., Yue, X., Younger, S.T., Janowski, B.A. and Corey, D.R. (2010) Involvement of argonaute proteins in gene silencing and activation by RNAs complementary to a non-coding transcript at the progesterone receptor promoter. *Nucleic acids research*, **38**, 7736-7748.
205. Huang, V., Zheng, J., Qi, Z., Wang, J., Place, R.F., Yu, J., Li, H. and Li, L.C. (2013) Ago1 Interacts with RNA polymerase II and binds to the promoters of actively transcribed genes in human cancer cells. *PLoS genetics*, **9**, e1003821.
206. Weinmann, L., Hock, J., Ivacevic, T., Ohrt, T., Mutze, J., Schwille, P., Kremmer, E., Benes, V., Urlaub, H. and Meister, G. (2009) Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell*, **136**, 496-507.
207. Robb, G.B., Brown, K.M., Khurana, J. and Rana, T.M. (2005) Specific and potent RNAi in the nucleus of human cells. *Nature structural & molecular biology*, **12**, 133-137.
208. Janowski, B.A., Huffman, K.E., Schwartz, J.C., Ram, R., Nordsell, R., Shames, D.S., Minna, J.D. and Corey, D.R. (2006) Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. *Nature structural & molecular biology*, **13**, 787-792.
209. Kim, D.H., Villeneuve, L.M., Morris, K.V. and Rossi, J.J. (2006) Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. *Nature structural & molecular biology*, **13**, 793-797.

210. Huang, V., Place, R.F., Portnoy, V., Wang, J., Qi, Z., Jia, Z., Yu, A., Shuman, M., Yu, J. and Li, L.C. (2012) Upregulation of Cyclin B1 by miRNA and its implications in cancer. *Nucleic acids research*, **40**, 1695-1707.
211. Moshkovich, N., Nisha, P., Boyle, P.J., Thompson, B.A., Dale, R.K. and Lei, E.P. (2011) RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. *Genes & development*, **25**, 1686-1701.
212. Ahlenstiel, C.L., Lim, H.G., Cooper, D.A., Ishida, T., Kelleher, A.D. and Suzuki, K. (2012) Direct evidence of nuclear Argonaute distribution during transcriptional silencing links the actin cytoskeleton to nuclear RNAi machinery in human cells. *Nucleic acids research*, **40**, 1579-1595.
213. Kim, D.H., Saetrom, P., Snove, O., Jr. and Rossi, J.J. (2008) MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 16230-16235.
214. Younger, S.T. and Corey, D.R. (2011) Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. *Nucleic acids research*, **39**, 5682-5691.
215. Place, R.F., Li, L.C., Pookot, D., Noonan, E.J. and Dahiya, R. (2008) MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 1608-1613.
216. Hwang, H.W., Wentzel, E.A. and Mendell, J.T. (2007) A hexanucleotide element directs microRNA nuclear import. *Science*, **315**, 97-100.
217. Liu, J., Hu, J. and Corey, D.R. (2012) Expanding the action of duplex RNAs into the nucleus: redirecting alternative splicing. *Nucleic acids research*, **40**, 1240-1250.
218. Ameyar-Zazoua, M., Rachez, C., Souidi, M., Robin, P., Fritsch, L., Young, R., Morozova, N., Fenouil, R., Descostes, N., Andrau, J.C. *et al.* (2012) Argonaute proteins couple chromatin silencing to alternative splicing. *Nature structural & molecular biology*, **19**, 998-1004.
219. Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A. *et al.* (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular cell*, **39**, 925-938.
220. Gao, M., Wei, W., Li, M.M., Wu, Y.S., Ba, Z., Jin, K.X., Li, M.M., Liao, Y.Q., Adhikari, S., Chong, Z. *et al.* (2014) Ago2 facilitates Rad51 recruitment and DNA double-strand break repair by homologous recombination. *Cell research*, **24**, 532-541.
221. Norbury, C. and Nurse, P. (1992) Animal cell cycles and their control. *Annual review of biochemistry*, **61**, 441-470.
222. O'Connor, C. (2008) Cell Division: Stages of Mitosis

Nature Education
223. Cooper, G.M.a.H., R.E. (2004) The cell: a molecular approach (3rd edition). *ASM Press*.
224. Ben-Shlomo, R. (2014) Chronodisruption, cell cycle checkpoints and DNA repair. *Indian journal of experimental biology*, **52**, 399-403.
225. Polo, S.E. and Jackson, S.P. (2011) Dynamics of DNA damage response proteins at DNA breaks: a focus on protein modifications. *Genes & development*, **25**, 409-433.
226. Roos, W.P. and Kaina, B. (2006) DNA damage-induced cell death by apoptosis. *Trends in molecular medicine*, **12**, 440-450.
227. Medema, R.H. and Macurek, L. (2012) Checkpoint control and cancer. *Oncogene*, **31**, 2601-2613.
228. Novak, B., Tyson, J.J., Gyorffy, B. and Csikasz-Nagy, A. (2007) Irreversible cell-cycle transitions are due to systems-level feedback. *Nature cell biology*, **9**, 724-728.

229. Lopez-Aviles, S., Kapuy, O., Novak, B. and Uhlmann, F. (2009) Irreversibility of mitotic exit is the consequence of systems-level feedback. *Nature*, **459**, 592-595.
230. Reed, S.I. (2003) Ratchets and clocks: the cell cycle, ubiquitylation and protein turnover. *Nature reviews. Molecular cell biology*, **4**, 855-864.
231. Malumbres, M. and Barbacid, M. (2005) Mammalian cyclin-dependent kinases. *Trends in biochemical sciences*, **30**, 630-641.
232. Harbour, J.W., Luo, R.X., Dei Santi, A., Postigo, A.A. and Dean, D.C. (1999) Cdk phosphorylation triggers sequential intramolecular interactions that progressively block Rb functions as cells move through G1. *Cell*, **98**, 859-869.
233. Malumbres, M. and Barbacid, M. (2009) Cell cycle, CDKs and cancer: a changing paradigm. *Nature reviews. Cancer*, **9**, 153-166.
234. Michael, W.M. and Newport, J. (1998) Coupling of mitosis to the completion of S phase through Cdc34-mediated degradation of Wee1. *Science*, **282**, 1886-1889.
235. Malumbres, M., Sotillo, R., Santamaria, D., Galan, J., Cerezo, A., Ortega, S., Dubus, P. and Barbacid, M. (2004) Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. *Cell*, **118**, 493-504.
236. Rane, S.G., Dubus, P., Mettus, R.V., Galbreath, E.J., Boden, G., Reddy, E.P. and Barbacid, M. (1999) Loss of Cdk4 expression causes insulin-deficient diabetes and Cdk4 activation results in beta-islet cell hyperplasia. *Nature genetics*, **22**, 44-52.
237. Santamaria, D., Barriere, C., Cerqueira, A., Hunt, S., Tardy, C., Newton, K., Caceres, J.F., Dubus, P., Malumbres, M. and Barbacid, M. (2007) Cdk1 is sufficient to drive the mammalian cell cycle. *Nature*, **448**, 811-815.
238. Rahman, N. (2014) Realizing the promise of cancer predisposition genes. *Nature*, **505**, 302-308.
239. Ortega, S., Malumbres, M. and Barbacid, M. (2002) Cyclin D-dependent kinases, INK4 inhibitors and cancer. *Biochimica et biophysica acta*, **1602**, 73-87.
240. Malumbres, M. and Barbacid, M. (2001) To cycle or not to cycle: a critical decision in cancer. *Nature reviews. Cancer*, **1**, 222-231.
241. Schmidt, E.E., Ichimura, K., Reifenberger, G. and Collins, V.P. (1994) CDKN2 (p16/MTS1) gene deletion or CDK4 amplification occurs in the majority of glioblastomas. *Cancer research*, **54**, 6321-6324.
242. Khatib, Z.A., Matsushime, H., Valentine, M., Shapiro, D.N., Sherr, C.J. and Look, A.T. (1993) Coamplification of the CDK4 gene with MDM2 and GLI in human sarcomas. *Cancer research*, **53**, 5535-5541.
243. An, H.X., Beckmann, M.W., Reifenberger, G., Bender, H.G. and Niederacher, D. (1999) Gene amplification and overexpression of CDK4 in sporadic breast carcinomas is associated with high tumor cell proliferation. *The American journal of pathology*, **154**, 113-118.
244. Cheung, T.H., Yu, M.M., Lo, K.W., Yim, S.F., Chung, T.K. and Wong, Y.F. (2001) Alteration of cyclin D1 and CDK4 gene in carcinoma of uterine cervix. *Cancer letters*, **166**, 199-206.
245. Costello, J.F., Plass, C., Arap, W., Chapman, V.M., Held, W.A., Berger, M.S., Su Huang, H.J. and Cavenee, W.K. (1997) Cyclin-dependent kinase 6 (CDK6) amplification in human gliomas identified using two-dimensional separation of genomic DNA. *Cancer research*, **57**, 1250-1254.
246. Chilosi, M., Doglioni, C., Yan, Z., Lestani, M., Menestrina, F., Sorio, C., Benedetti, A., Vinante, F., Pizzolo, G. and Inghirami, G. (1998) Differential expression of cyclin-dependent kinase 6 in cortical thymocytes and T-cell lymphoblastic lymphoma/leukemia. *The American journal of pathology*, **152**, 209-217.

247. Sasaki, S., Kitagawa, Y., Sekido, Y., Minna, J.D., Kuwano, H., Yokota, J. and Kohno, T. (2003) Molecular processes of chromosome 9p21 deletions in human cancers. *Oncogene*, **22**, 3792-3798.
248. Tetsu, O. and McCormick, F. (2003) Proliferation of cancer cells despite CDK2 inhibition. *Cancer cell*, **3**, 233-245.
249. van den Heuvel, S. and Harlow, E. (1993) Distinct roles for cyclin-dependent kinases in cell cycle control. *Science*, **262**, 2050-2054.
250. Reddy, H.K., Mettus, R.V., Rane, S.G., Grana, X., Litvin, J. and Reddy, E.P. (2005) Cyclin-dependent kinase 4 expression is essential for neu-induced breast tumorigenesis. *Cancer research*, **65**, 10174-10178.
251. Yu, Q., Sicinska, E., Geng, Y., Ahnstrom, M., Zagozdzon, A., Kong, Y., Gardner, H., Kiyokawa, H., Harris, L.N., Stal, O. *et al.* (2006) Requirement for CDK4 kinase function in breast cancer. *Cancer cell*, **9**, 23-32.
252. Ell, B. and Kang, Y. (2013) Transcriptional control of cancer metastasis. *Trends in cell biology*, **23**, 603-611.
253. Wolfer, A. and Ramaswamy, S. (2011) MYC and metastasis. *Cancer research*, **71**, 2034-2037.
254. Remy, P. and Baltzinger, M. (2000) The Ets-transcription factor family in embryonic development: lessons from the amphibian and bird. *Oncogene*, **19**, 6417-6431.
255. Myatt, S.S. and Lam, E.W. (2007) The emerging roles of forkhead box (Fox) proteins in cancer. *Nature reviews. Cancer*, **7**, 847-859.
256. Paik, J.H., Kollipara, R., Chu, G., Ji, H., Xiao, Y., Ding, Z., Miao, L., Tothova, Z., Horner, J.W., Carrasco, D.R. *et al.* (2007) FoxOs are lineage-restricted redundant tumor suppressors and regulate endothelial cell homeostasis. *Cell*, **128**, 309-323.
257. Gavin, M.A., Rasmussen, J.P., Fontenot, J.D., Vasta, V., Manganiello, V.C., Beavo, J.A. and Rudensky, A.Y. (2007) Foxp3-dependent programme of regulatory T-cell differentiation. *Nature*, **445**, 771-775.
258. Seo, S., Fujita, H., Nakano, A., Kang, M., Duarte, A. and Kume, T. (2006) The forkhead transcription factors, Foxc1 and Foxc2, are required for arterial specification and lymphatic sprouting during vascular development. *Developmental biology*, **294**, 458-470.
259. Potente, M., Urbich, C., Sasaki, K., Hofmann, W.K., Heeschen, C., Aicher, A., Kollipara, R., DePinho, R.A., Zeiher, A.M. and Dimmeler, S. (2005) Involvement of Foxo transcription factors in angiogenesis and postnatal neovascularization. *The Journal of clinical investigation*, **115**, 2382-2392.
260. Hoesel, B. and Schmid, J.A. (2013) The complexity of NF-kappaB signaling in inflammation and cancer. *Molecular cancer*, **12**, 86.
261. Rivlin, N., Brosh, R., Oren, M. and Rotter, V. (2011) Mutations in the p53 Tumor Suppressor Gene: Important Milestones at the Various Steps of Tumorigenesis. *Genes & cancer*, **2**, 466-474.
262. Mellert, H. and Espinosa, J.M. (2013) Tumor suppression by p53: is apoptosis important or not? *Cell reports*, **3**, 1335-1336.
263. Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K. *et al.* (2002) Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 15524-15529.
264. Chiorazzi, N., Rai, K.R. and Ferrarini, M. (2005) Chronic lymphocytic leukemia. *The New England journal of medicine*, **352**, 804-815.

265. Mian, M., Scandurra, M., Chigrinova, E., Shen, Y., Inghirami, G., Greiner, T.C., Chan, W.C., Vose, J.M., Testoni, M., Chiappella, A. *et al.* (2012) Clinical and molecular characterization of diffuse large B-cell lymphomas with 13q14.3 deletion. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, **23**, 729-735.
266. Bottoni, A., Piccin, D., Tagliati, F., Luchin, A., Zatelli, M.C. and degli Uberti, E.C. (2005) miR-15a and miR-16-1 down-regulation in pituitary adenomas. *Journal of cellular physiology*, **204**, 280-285.
267. Bonci, D., Coppola, V., Musumeci, M., Addario, A., Giuffrida, R., Memeo, L., D'Urso, L., Pagliuca, A., Biffoni, M., Labbaye, C. *et al.* (2008) The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities. *Nature medicine*, **14**, 1271-1277.
268. Pekarsky, Y. and Croce, C.M. (2014) Role of miR-15/16 in CLL. *Cell death and differentiation*.
269. Aqeilan, R.I., Calin, G.A. and Croce, C.M. (2010) miR-15a and miR-16-1 in cancer: discovery, function and future perspectives. *Cell death and differentiation*, **17**, 215-220.
270. Levine, A.J., Hu, W. and Feng, Z. (2006) The P53 pathway: what questions remain to be explored? *Cell death and differentiation*, **13**, 1027-1036.
271. Ko, L.J. and Prives, C. (1996) p53: puzzle and paradigm. *Genes & development*, **10**, 1054-1072.
272. He, L., He, X., Lim, L.P., de Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., Ridzon, D. *et al.* (2007) A microRNA component of the p53 tumour suppressor network. *Nature*, **447**, 1130-1134.
273. Georges, S.A., Biery, M.C., Kim, S.Y., Schelter, J.M., Guo, J., Chang, A.N., Jackson, A.L., Carleton, M.O., Linsley, P.S., Cleary, M.A. *et al.* (2008) Coordinated regulation of cell cycle transcripts by p53-Inducible microRNAs, miR-192 and miR-215. *Cancer research*, **68**, 10105-10112.
274. Suzuki, H.I., Yamagata, K., Sugimoto, K., Iwamoto, T., Kato, S. and Miyazono, K. (2009) Modulation of microRNA processing by p53. *Nature*, **460**, 529-533.
275. Wong, J.V., Dong, P., Nevins, J.R., Mathey-Prevot, B. and You, L. (2011) Network calisthenics: control of E2F dynamics in cell cycle entry. *Cell Cycle*, **10**, 3086-3094.
276. Cam, H. and Dynlacht, B.D. (2003) Emerging roles for E2F: beyond the G1/S transition and DNA replication. *Cancer cell*, **3**, 311-316.
277. Bueno, M.J. and Malumbres, M. (2011) MicroRNAs and the cell cycle. *Biochimica et biophysica acta*, **1812**, 592-601.
278. Wang, Y., Baskerville, S., Shenoy, A., Babiarz, J.E., Baehner, L. and Blelloch, R. (2008) Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nature genetics*, **40**, 1478-1483.
279. O'Donnell, K.A., Wentzel, E.A., Zeller, K.I., Dang, C.V. and Mendell, J.T. (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, **435**, 839-843.
280. Leone, G., DeGregori, J., Sears, R., Jakoi, L. and Nevins, J.R. (1997) Myc and Ras collaborate in inducing accumulation of active cyclin E/Cdk2 and E2F. *Nature*, **387**, 422-426.
281. Fernandez, P.C., Frank, S.R., Wang, L., Schroeder, M., Liu, S., Greene, J., Cocito, A. and Amati, B. (2003) Genomic targets of the human c-Myc protein. *Genes & development*, **17**, 1115-1129.
282. Lin, R.J., Lin, Y.C. and Yu, A.L. (2010) miR-149* induces apoptosis by inhibiting Akt1 and E2F1 in human cancer cells. *Molecular carcinogenesis*, **49**, 719-727.

283. Lee, K.H., Chen, Y.L., Yeh, S.D., Hsiao, M., Lin, J.T., Goan, Y.G. and Lu, P.J. (2009) MicroRNA-330 acts as tumor suppressor and induces apoptosis of prostate cancer cells through E2F1-mediated suppression of Akt phosphorylation. *Oncogene*, **28**, 3360-3370.
284. Guo, X., Guo, L., Ji, J., Zhang, J., Zhang, J., Chen, X., Cai, Q., Li, J., Gu, Q., Liu, B. *et al.* (2010) miRNA-331-3p directly targets E2F1 and induces growth arrest in human gastric cancer. *Biochemical and biophysical research communications*, **398**, 1-6.
285. Huang, L., Luo, J., Cai, Q., Pan, Q., Zeng, H., Guo, Z., Dong, W., Huang, J. and Lin, T. (2011) MicroRNA-125b suppresses the development of bladder cancer by targeting E2F3. *International journal of cancer. Journal international du cancer*, **128**, 1758-1769.
286. Giannakakis, A., Sandaltzopoulos, R., Greshock, J., Liang, S., Huang, J., Hasegawa, K., Li, C., O'Brien-Jenkins, A., Katsaros, D., Weber, B.L. *et al.* (2008) miR-210 links hypoxia with cell cycle regulation and is deleted in human epithelial ovarian cancer. *Cancer biology & therapy*, **7**, 255-264.
287. Xu, T., Zhu, Y., Xiong, Y., Ge, Y.Y., Yun, J.P. and Zhuang, S.M. (2009) MicroRNA-195 suppresses tumorigenicity and regulates G1/S transition of human hepatocellular carcinoma cells. *Hepatology*, **50**, 113-121.
288. Korenjak, M. and Brehm, A. (2005) E2F-Rb complexes regulating transcription of genes important for differentiation and development. *Current opinion in genetics & development*, **15**, 520-527.
289. Wang, Q., Li, Y.C., Wang, J., Kong, J., Qi, Y., Quigg, R.J. and Li, X. (2008) miR-17-92 cluster accelerates adipocyte differentiation by negatively regulating tumor-suppressor Rb2/p130. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 2889-2894.
290. Benetti, R., Gonzalo, S., Jaco, I., Munoz, P., Gonzalez, S., Schoeftner, S., Murchison, E., Andl, T., Chen, T., Klatt, P. *et al.* (2008) A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nature structural & molecular biology*, **15**, 998.
291. Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M. *et al.* (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 2257-2261.
292. Sherr, C.J. and Roberts, J.M. (1999) CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes & development*, **13**, 1501-1512.
293. Denicourt, C. and Dowdy, S.F. (2004) Cip/Kip proteins: more than just CDKs inhibitors. *Genes & development*, **18**, 851-855.
294. Malhas, A., Saunders, N.J. and Vaux, D.J. (2010) The nuclear envelope can control gene expression and cell cycle progression via miRNA regulation. *Cell Cycle*, **9**, 531-539.
295. Lin, J., Huang, S., Wu, S., Ding, J., Zhao, Y., Liang, L., Tian, Q., Zha, R., Zhan, R. and He, X. (2011) MicroRNA-423 promotes cell growth and regulates G(1)/S transition by targeting p21Cip1/Waf1 in hepatocellular carcinoma. *Carcinogenesis*, **32**, 1641-1647.
296. Conkrite, K., Sundby, M., Mukai, S., Thomson, J.M., Mu, D., Hammond, S.M. and MacPherson, D. (2011) miR-17~92 cooperates with RB pathway mutations to promote retinoblastoma. *Genes & development*, **25**, 1734-1745.
297. Kim, Y.K., Yu, J., Han, T.S., Park, S.Y., Namkoong, B., Kim, D.H., Hur, K., Yoo, M.W., Lee, H.J., Yang, H.K. *et al.* (2009) Functional links between clustered

microRNAs: suppression of cell-cycle inhibitors by microRNA clusters in gastric cancer. *Nucleic acids research*, **37**, 1672-1681.
298. Wang, X., Gocek, E., Liu, C.G. and Studzinski, G.P. (2009) MicroRNAs181 regulate the expression of p27Kip1 in human myeloid leukemia cells induced to differentiate by 1,25-dihydroxyvitamin D3. *Cell Cycle*, **8**, 736-741.
299. Garofalo, M., Quintavalle, C., Romano, G., Croce, C.M. and Condorelli, G. (2012) miR221/222 in cancer: their role in tumor progression and response to therapy. *Current molecular medicine*, **12**, 27-33.
300. Nigg, E.A. (2001) Mitotic kinases as regulators of cell division and its checkpoints. *Nature reviews. Molecular cell biology*, **2**, 21-32.
301. Schultz, J., Lorenz, P., Gross, G., Ibrahim, S. and Kunz, M. (2008) MicroRNA let-7b targets important cell cycle molecules in malignant melanoma cells and interferes with anchorage-independent growth. *Cell research*, **18**, 549-557.
302. Lal, A., Navarro, F., Maher, C.A., Maliszewski, L.E., Yan, N., O'Day, E., Chowdhury, D., Dykxhoorn, D.M., Tsai, P., Hofmann, O. *et al.* (2009) miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3'UTR microRNA recognition elements. *Molecular cell*, **35**, 610-625.
303. Qi, J., Yu, J.Y., Shcherbata, H.R., Mathieu, J., Wang, A.J., Seal, S., Zhou, W., Stadler, B.M., Bourgin, D., Wang, L. *et al.* (2009) microRNAs regulate human embryonic stem cell division. *Cell Cycle*, **8**, 3729-3741.
304. Butz, H., Liko, I., Czirjak, S., Igaz, P., Khan, M.M., Zivkovic, V., Balint, K., Korbonits, M., Racz, K. and Patocs, A. (2010) Down-regulation of Wee1 kinase by a specific subset of microRNA in human sporadic pituitary adenomas. *The Journal of clinical endocrinology and metabolism*, **95**, E181-191.
305. Chen, B., Duan, L., Yin, G., Tan, J. and Jiang, X. (2013) Simultaneously expressed miR-424 and miR-381 synergistically suppress the proliferation and survival of renal cancer cells---Cdc2 activity is up-regulated by targeting WEE1. *Clinics (Sao Paulo)*, **68**, 825-833.
306. Bar-Joseph, Z., Siegfried, Z., Brandeis, M., Brors, B., Lu, Y., Eils, R., Dynlacht, B.D. and Simon, I. (2008) Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 955-960.
307. Grant, G.D., Brooks, L., 3rd, Zhang, X., Mahoney, J.M., Martyanov, V., Wood, T.A., Sherlock, G., Cheng, C. and Whitfield, M.L. (2013) Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors. *Molecular biology of the cell*, **24**, 3634-3650.
308. Beyrouthy, M.J., Alexander, K.E., Baldwin, A., Whitfield, M.L., Bass, H.W., McGee, D. and Hurt, M.M. (2008) Identification of G1-regulated genes in normally cycling human cells. *PloS one*, **3**, e3943.
309. De Nisco, N.J., Abo, R.P., Wu, C.M., Penterman, J. and Walker, G.C. (2014) Global analysis of cell cycle gene expression of the legume symbiont Sinorhizobium meliloti. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 3217-3224.
310. de Almeida Engler, J., De Veylder, L., De Groodt, R., Rombauts, S., Boudolf, V., De Meyer, B., Hemerly, A., Ferreira, P., Beeckman, T., Karimi, M. *et al.* (2009) Systematic analysis of cell-cycle gene expression during Arabidopsis development. *The Plant journal : for cell and molecular biology*, **59**, 645-660.
311. Shedden, K. and Cooper, S. (2002) Analysis of cell-cycle gene expression in Saccharomyces cerevisiae using microarrays and multiple synchronization methods. *Nucleic acids research*, **30**, 2920-2929.

312. Beisel, C. and Paro, R. (2011) Silencing chromatin: comparing modes and mechanisms. *Nature reviews. Genetics*, **12**, 123-135.
313. Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O. *et al.* (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Molecular biology of the cell*, **13**, 1977-2000.
314. Pena-Diaz, J., Hegre, S.A., Anderssen, E., Aas, P.A., Mjelle, R., Gilfillan, G.D., Lyle, R., Drablos, F., Krokan, H.E. and Saetrom, P. (2013) Transcription profiling during the cell cycle shows that a subset of Polycomb-targeted genes is upregulated during DNA replication. *Nucleic acids research*, **41**, 2846-2856.
315. Branzei, D. and Foiani, M. (2008) Regulation of DNA repair throughout the cell cycle. *Nature reviews. Molecular cell biology*, **9**, 297-308.
316. Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular cell*, **15**, 185-197.
317. Wang, B., Li, S., Qi, H.H., Chowdhury, D., Shi, Y. and Novina, C.D. (2009) Distinct passenger strand and mRNA cleavage activities of human Argonaute proteins. *Nature structural & molecular biology*, **16**, 1259-1266.
318. Lackner, D.H., Beilharz, T.H., Marguerat, S., Mata, J., Watt, S., Schubert, F., Preiss, T. and Bahler, J. (2007) A network of multiple regulatory layers shapes gene expression in fission yeast. *Molecular cell*, **26**, 145-155.
319. Proudfoot, N.J., Furger, A. and Dye, M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501-512.
320. Orphanides, G. and Reinberg, D. (2002) A unified theory of gene expression. *Cell*, **108**, 439-451.
321. Moore, M.J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science*, **309**, 1514-1518.
322. Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature*, **416**, 499-506.
323. OMICS, J.o.I. http://www.jiomics.com/index.php/jio/index.
324. Biology, O.A.J.o.I. http://www.liebertpub.com/overview/omics-a-journal-of-integrative-biology/43/.
325. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B. (2007) Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry*, **389**, 1017-1031.
326. de Godoy, L.M., Olsen, J.V., de Souza, G.A., Li, G., Mortensen, P. and Mann, M. (2006) Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome biology*, **7**, R50.
327. Vogel, C. and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature reviews. Genetics*, **13**, 227-232.
328. Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature*, **473**, 337-342.
329. Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2013) Corrigendum: Global quantification of mammalian gene expression control. *Nature*, **495**, 126-127.
330. Maier, T., Guell, M. and Serrano, L. (2009) Correlation of mRNA and protein in complex biological samples. *FEBS letters*, **583**, 3966-3973.
331. Vogel, C., Abreu Rde, S., Ko, D., Le, S.Y., Shapiro, B.A., Burns, S.C., Sandhu, D., Boutz, D.R., Marcotte, E.M. and Penalva, L.O. (2010) Sequence signatures and

mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Molecular systems biology*, **6**, 400.
332. Ramakrishnan, S.R., Vogel, C., Prince, J.T., Li, Z., Penalva, L.O., Myers, M., Marcotte, E.M., Miranker, D.P. and Wang, R. (2009) Integrating shotgun proteomics and mRNA expression data to improve protein identification. *Bioinformatics*, **25**, 1397-1403.
333. Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R. and Rajewsky, N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58-63.
334. Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P. and Bartel, D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64-71.
335. Whitfield, M.L., George, L.K., Grant, G.D. and Perou, C.M. (2006) Common markers of proliferation. *Nature reviews. Cancer*, **6**, 99-106.
336. Cooper, S. (2004) Rejoinder: whole-culture synchronization cannot, and does not, synchronize cells. *Trends in biotechnology*, **22**, 274-276.
337. Cooper, S. (2004) Is whole-culture synchronization biology's 'perpetual-motion machine'? *Trends in biotechnology*, **22**, 266-269.
338. Cooper, S. (2003) Rethinking synchronization of mammalian cells for cell cycle analysis. *Cellular and molecular life sciences : CMLS*, **60**, 1099-1106.
339. Cooper, S. (1998) Mammalian cells are not synchronized in G1-phase by starvation or inhibition: considerations of the fundamental concept of G1-phase synchronization. *Cell proliferation*, **31**, 9-16.
340. Spellman, P.T. and Sherlock, G. (2004) Reply: whole-culture synchronization - effective tools for cell cycle studies. *Trends in biotechnology*, **22**, 270-273.
341. Spellman, P.T. and Sherlock, G. (2004) Final words: cell age and cell cycle are unlinked. *Trends in biotechnology*, **22**, 277-278.
342. Davis, P.K., Ho, A. and Dowdy, S.F. (2001) Biological methods for cell-cycle synchronization of mammalian cells. *BioTechniques*, **30**, 1322-1326, 1328, 1330-1321.
343. Amit, I., Citri, A., Shay, T., Lu, Y., Katz, M., Zhang, F., Tarcic, G., Siwak, D., Lahad, J., Jacob-Hirsch, J. *et al.* (2007) A module of negative feedback regulators defines growth factor signaling. *Nature genetics*, **39**, 503-512.
344. Bar-Joseph, Z., Gitter, A. and Simon, I. (2012) Studying and modelling dynamic biological processes using time-series gene expression data. *Nature reviews. Genetics*, **13**, 552-564.
345. Miyoshi, K., Okada, T.N., Siomi, H. and Siomi, M.C. (2011) Biochemical analyzes of endogenous argonaute complexes immunopurified with anti-Argonaute monoclonal antibodies. *Methods in molecular biology*, **725**, 29-43.
346. Rudel, S., Flatley, A., Weinmann, L., Kremmer, E. and Meister, G. (2008) A multifunctional human Argonaute2-specific monoclonal antibody. *RNA*, **14**, 1244-1253.
347. Zhu, J.Y., Strehle, M., Frohn, A., Kremmer, E., Hofig, K.P., Meister, G. and Adler, H. (2010) Identification and analysis of expression of novel microRNAs of murine gammaherpesvirus 68. *Journal of virology*, **84**, 10266-10275.
348. Beitzinger, M., Peters, L., Zhu, J.Y., Kremmer, E. and Meister, G. (2007) Identification of human microRNA targets from isolated argonaute protein complexes. *RNA biology*, **4**, 76-84.
349. Yap, K.L., Li, S., Munoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J. and Zhou, M.M. (2010) Molecular interplay of the noncoding RNA

ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Molecular cell*, **38**, 662-674.
350. Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D. and Kanduri, C. (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Molecular cell*, **32**, 232-246.
351. Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B. and Zhang, Y. (2003) Role of histone H3 lysine 27 methylation in X inactivation. *Science*, **300**, 131-135.
352. Tsukioka, H., Takahashi, M., Mon, H., Okano, K., Mita, K., Shimada, T., Lee, J.M., Kawaguchi, Y., Koga, K. and Kusakabe, T. (2006) Role of the silkworm argonaute2 homolog gene in double-strand break repair of extrachromosomal DNA. *Nucleic acids research*, **34**, 1092-1101.
353. Yamanaka, S. and Siomi, H. (2014) diRNA-Ago2-RAD51 complexes at double-strand break sites. *Cell research*, **24**, 511-512.

Papers I-IV

Paper I

2846–2856 *Nucleic Acids Research, 2013, Vol. 41, No. 5* *Published online 15 January 2013*
doi:10.1093/nar/gks1336

Transcription profiling during the cell cycle shows that a subset of Polycomb-targeted genes is upregulated during DNA replication

Javier Peña-Diaz[1], Siv A. Hegre[1,2], Endre Anderssen[1], Per A. Aas[1], Robin Mjelle[1], Gregor D. Gilfillan[3], Robert Lyle[3], Finn Drabløs[1], Hans E. Krokan[1] and Pål Sætrom[1,4,*]

[1]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway, [2]St. Olavs Hospital, NO-7006 Trondheim, Norway, [3]Department of Medical Genetics and Norwegian Sequencing Centre, Oslo University Hospital, NO-0407 Oslo, Norway and [4]Department of Computer and Information Science, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

Received June 10, 2012; Revised November 29, 2012; Accepted November 30, 2012

ABSTRACT

Genome-wide gene expression analyses of the human somatic cell cycle have indicated that the set of cycling genes differ between primary and cancer cells. By identifying genes that have cell cycle dependent expression in HaCaT human keratinocytes and comparing these with previously identified cell cycle genes, we have identified three distinct groups of cell cycle genes. First, housekeeping genes enriched for known cell cycle functions; second, cell type-specific genes enriched for HaCaT-specific functions; and third, Polycomb-regulated genes. These Polycomb-regulated genes are specifically upregulated during DNA replication, and consistent with being epigenetically silenced in other cell cycle phases, these genes have lower expression than other cell cycle genes. We also find similar patterns in foreskin fibroblasts, indicating that replication-dependent expression of Polycomb-silenced genes is a prevalent but unrecognized regulatory mechanism.

INTRODUCTION

Genome-wide studies of gene expression throughout the cell division cycle have revealed several genes that are differentially expressed (1–9), but have also indicated that the set of cycling genes differs between primary and cancer cells (3). Primary cells are, however, inherently difficult to synchronize, for example, only 40–50% of foreskin fibroblast cells in culture can be synchronized by serum starvation or double thymidine block (3). Although sophisticated statistics may partially overcome lack of synchronization (3), a large population of asynchronous or arrested cells results in high background gene expression noise. Consequently, more cycling genes can be detected in a highly synchronous culture than in a culture where at most 50% of the cells are synchronized. Moreover, as the only human cell line—in addition to primary fibroblasts (1,3,4)—profiled for cell cycle expression so far is the cervical cancer cell line HeLa (2,5), it is unclear to what extent cell type-specific factors affect reported differences in cycling genes.

We have used the human keratinocyte cell line HaCaT to address this question. Specifically, by measuring the gene expression profiles of double thymidine synchronized HaCaT cells, we identified three major groups of cycling genes. First, a set of genes with housekeeping characteristics, strong enrichment for known cell cycle functions and overlap with previously identified cell cycle genes. Second, a set of genes with cell type-specific characteristics, enrichment for HaCaT-specific functions and poor overlap with previously identified cell cycle genes. Third, a set of genes that has the mark for Polycomb silencing: histone H3 lysine 27 tri-methylation (H3K27me3). We show that this third set of genes is expressed in a replication-dependent manner, as the genes are upregulated during S phase in a pattern related to DNA replication timing. Consistent with being epigenetically silenced in other cell cycle phases, these genes are generally lower expressed than are other cell cycle expressed genes. We also find

*To whom correspondence should be addressed. Tel: +47 735 94 796; Fax: +47 735 94 466; Email: pal.satrom@ntnu.no
Present address:
Javier Peña-Diaz, Institute of Molecular Cancer Research (IMCR), University of Zurich, 8057 Zurich, Switzerland.

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

similar patterns in foreskin fibroblasts synchronized by serum starvation, indicating that replication-dependent expression of Polycomb-silenced genes is a prevalent but unrecognized regulatory mechanism.

MATERIALS AND METHODS

HaCaT cell culture and synchronization

HaCaT cells were plated at 10% confluence (1×10^6 cells) in 150-mm tissue culture dishes in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS). Cells were arrested in the interphase G_1/S by double thymidine block; briefly, cells were treated with 2 mM of thymidine for 18 h, released from the arrest for 10 h and arrested a second time with 2 mM of thymidine for additional 18 h. After treatment, media was replaced, and cells were collected at 3-h intervals for up to 33 h, covering approximately two cell cycles. Synchrony was monitored by flow cytometry analysis of propidium iodide-stained cells and by cell counting. Quantification of cells in each phase was done with the MultiCycle DNA cell cycle analysis software (Phoenix Flow Systems Inc., San Diego, CA, USA) combined with the cell counting results.

HeLa cell culture and synchronization

Adherent HeLa cells were plated in 150-mm culture dishes in DMEM with 10% of FBS, 2 mM of glutamine, 0.1 mg/ml of gentamicin and 1.25 μg/ml of fungizone. Cells at 60–70% confluence were arrested in the G_2/M transition with 100 ng/ml of nocodazole for 17 h. The mitotic cells were then collected by manual shake-off, washed twice and re-plated in fresh DMEM to progress through the cell cycle. Cells were harvested from culture dishes by trypsinization every 30 min for the first 2 h and then every 3 h from 3 to 24 h after release. Phosphate-buffered saline containing 3% of FBS was added to inactivate the trypsin.

HeLa cells were pelleted and resuspended in 100 μl of RNAlater (Applied Biosystems/Ambion, Austin, TX, USA). All pellets were kept at 4°C overnight and were stored at −80°C before use. Verification of the cell cycle stage was determined by analysing the DNA content of propidium iodide-stained cells by a BD FACSAria (BD Biosciences, San Jose, CA, USA) flow cytometer. Quantification of cells in each phase was done with FlowJo (Tree Star Inc., OR, USA).

cRNA synthesis and microarray hybridization—HaCaT

Total RNA was extracted using the High Pure RNA Isolation Kit (Roche Applied Science, Indianapolis, IN, USA) and the manufacturer's protocol. RNA from synchronous cells was reverse transcribed into cDNA (cDNA synthesis Kit, Invitrogen, Carlsbad, CA, USA), which was used as a template for the RNA polymerase Enzo (Affymetrix, Santa Clara, CA, USA) to synthesize dUTP–dCTP biotinylated cRNA. The labelled cRNA was hybridized to Affymetrix oligonucleotide arrays (HG-U133 Set) under conditions specified by the manufacturer.

Microarray analyses—HeLa

Total RNA was prepared using the *mir*Vana miRNA Isolation Kit (Ambion) according to the manufacturer's protocol. The integrity and stability of RNA samples were assessed by Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). We used the Illumina TotalPrep RNA amplification Kit (Ambion) to amplify RNA for hybridization on Illumina BeadChips. Total RNA from each sample was used to synthesize first strand cDNA by reverse transcription. After the second strand cDNA synthesis and cDNA purification steps, the *in vitro* transcription to synthesize cRNA was prepared overnight for 12 h. The gene expression profiles were measured using Illumina HumanHT-12 v3 Expression BeadChip (Illumina, San Diego, CA, USA).

Identification of periodically expressed genes—statistical analysis

Normalization across arrays was performed by robust multi-array average (RMA) from the Bioconductor package affy (10). As there are no probes specific for splice variants in the arrays used, for probes that hybridized with mRNAs corresponding to the same gene, only the probe having the highest variation was used for further analyses. The normalized data were then used to identify genes that showed cyclic behaviour. Genes that matched a cell cycle profile were identified using partial least squares (PLS) regression. Specifically, the gene expression matrix was regressed on to a sine and cosine function with periods equal to the estimated duration of the cell cycle. False discovery rates were computed using a modified Hotelling's T-square statistic and resampling. Resampling ($n = 1000$) was done by assigning phase angles randomly to the time-steps, thus preserving the correlation within replicates of the same time-steps.

Assigning cell cycle phases

Assignment of cell cycle phases used the following protocol. First, each gene in the PLS model was assigned a phase angle by computing the arctangent of the gene's first and second principle component. Second, we used a list of genes previously published to be predominantly expressed in G_1/S, S, G_2, G_2/M and M/G_1 phase [(2): Table 2 and the four M/G_1 associated genes from Figure 2] and identified which of these genes were significant ($q < 0.05$) in our PLS model (Supplementary Figure S3A). Third, for each set of significant G_1/S, S, G_2, G_2/M and M/G_1 phase genes, we computed the set's average phase angle and the corresponding standard deviation and used these values as parameters in a normal distribution to construct a phase angle probability density function for each of the five cell cycle phases (Supplementary Figure S3A). Fourth, we used these probability density functions to assign the most probable phase IDs to each of the significant ($q < 0.05$) genes in the PLS model. Fifth, if necessary, the standard deviation parameters in the phase angle probability density functions were

increased to ensure that the phase assignments had the same order as the average phase angles for the five groups. Specifically, if any out of order phase assignments were detected, the lowest of the five standard deviations was increased by 25%, and the phase assignments were redone. This process was repeated until all phase assignments were in the order defined by the averages.

Promoter CpG analyses

Gene annotations were downloaded from the refGene table in the University of California, Santa Cruz (UCSC) Genome Browser database (human genome assembly version hg18) (11). Promoter regions were defined as the 1000 nt upstream of annotated transcription start sites. Overlapping regions corresponding to genes with multiple transcripts having identical or alternative start sites were merged into one region to create a non-redundant set of promoter regions. Based on this non-redundant set, CpG ratios and CpG promoter classes were calculated and assigned as previously described (12).

Promoter histone analyses

Processed peak data from chromatin immunoprecipitation sequencing (ChIP-seq) experiments of H3K4me3 and H3K27me3 in seven cell lines (GM12878, HUVEC, K562, NHEK, H1-hESC, HMEC and NHLF) were downloaded from UCSC (hg18) (13,14). Non-redundant promoter regions that contained overlapping H3K4me3 and H3K27me3 peaks in all seven cell lines were classified as H3K4me3 and H3K27me3 enriched, respectively; promoter regions that contained both marks in all seven cell lines were classified as bivalent; and promoter regions that did not contain any of the marks or only contained the marks in some of the cell lines were classified as having no or inconsistent marks.

Processed peak data from ChIP-seq experiments of H3K4me3 and H3K27me3 in HeLa and normal adult dermal fibroblasts (NHDF-Ad) were downloaded from UCSC (hg19; Broad Histone track) and remapped to hg18 by the UCSC liftOver tool.

HaCaT ChIP-seq

ChIP, library preparation and sequencing were done as described in the Supplementary Methods. The resulting input chromatin and H3K4me3 and H3K27me3 IP libraries contained ~64 million (M), ~42 M and ~52 M sequence reads, respectively. We used cutadapt (15) to trim adapter sequences, bowtie (16) to align sequence reads from the human genome (version hg18) and discarded all reads that aligned to more than one genomic location. The ~50 M, ~32 M and ~39 M aligned reads for the input, H3K4me3 and H3K27me3 libraries, respectively, were processed by SICER (17) to identify genomic regions that were significantly enriched for H3K4me3 and H3K27me3 in HaCaT (SICER used the input chromatin library as control to identify significant regions; see Supplementary Dataset S3 and S4).

Gene ontology and Kyoto Encyclopedia of Genes and Genomes analyses

Gene ontology (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment was assessed by GOstats (18). For GO terms, the test was conditioned on the structure of the GO graph.

Data access

The microarray data from this publication have been submitted to the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/) and were assigned the identifiers E-MTAB-454 (HaCaT) and E-TABM-1152 (HeLa). HaCaT ChIP-seq data are available from http://tang.medisin.ntnu.no/~palsat/HaCaT-ChIP.tgz; bed-files with significant H3K4me3 and H3K27me3 regions are available as Supplementary Dataset S3 and S4.

RESULTS

Partial least squares regression identifies cell cycle-regulated transcripts

To overcome the problem of low synchronization of primary cells and to investigate potential cell type-related differences in cycling genes, we turned to the human keratinocyte cell line HaCaT—a non-tumorigenic, spontaneously immortalized cell line that exhibits apparently normal differentiation (19). Using a double thymidine block to arrest and, subsequently, release HaCaT cells at the G_1/S boundary resulted in ~90% of the cells re-entering the cell cycle and continuing as a uniform cohort through the cell cycle (Supplementary Figure S1). Based on the DNA content profile, we estimated the cell cycle duration to be ~20 h. Although the cells gradually lost synchrony, such that 70% instead of 90% of the cells were synchronously entering the second S phase, the results were reproducible across three independent experiments. We, therefore, considered synchronized HaCaT cells a good starting point for gene expression analyses.

We were primarily interested in genes that showed consistent cyclic expression with a period equal to the estimated cell cycle duration of 20 h. We, therefore, used PLS (20) to identify genes that followed a sine or cosine pattern with a period of 20 h. Conceptually, this PLS analysis projects the time profile for all genes onto a plane described by the sine and cosine of the cell cycle. As a result, genes that are periodically expressed with high amplitude will be far from the origin, whereas genes that vary little or randomly will be close to the origin. Moreover, the phase angle of the cyclic variation will determine the direction from the origin such that genes that have the same temporal expression patterns will lie in the same direction from the origin on the plane.

By profiling RNA isolated from the synchronized cells at 3 h intervals for 33 h, and combining PLS analyses with resampling to compute false discovery rates (21), we identified 1249 Entrez genes that had significant periodic expression patterns during the HaCaT cell cycle (Figure 1, Supplementary Dataset S1). The synchronization experiments' trajectory through the PLS cell cycle plane

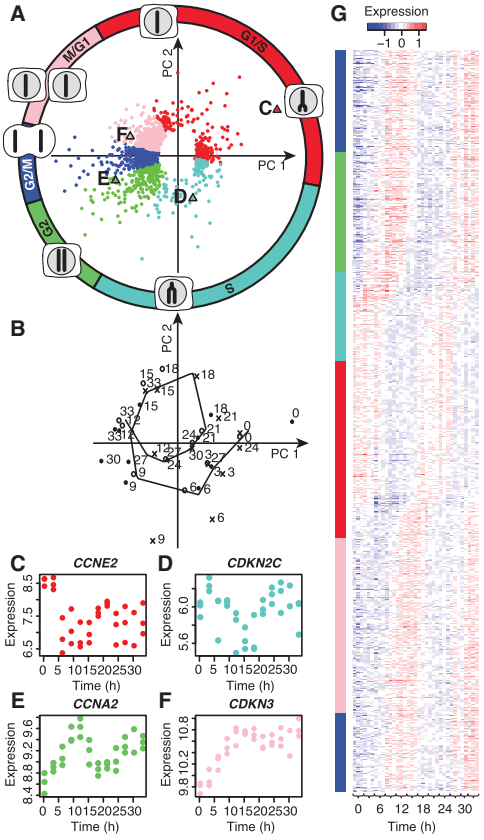


Figure 1. PLS regression identifies cyclic gene expression patterns. The loadings (**A**) and scores (**B**) plots show the contribution of each gene (loadings) to the location of each sample in the subspace of the PLS model (20) (scores; here, sine and cosine, period of 20 h). (A) Points show significant genes [$q < 0.05$; permutation test (21)] and their contribution to the PLS cell cycle model's first and second components (PC1 and PC2). Colours show the cell cycle phase where the gene is predominantly expressed (see 'Materials and Methods' section). The circle and cell diagrams indicate early and late DNA replication, G_2, mitosis and G_1 as determined by the scores plot (B) and flow cytometry (Supplementary Figure 1). (B) Points show the location of the samples from the three replicates (point, circle and cross) within the PLS model's first and second components; the line shows the replicates' average trajectory; numbers show the sample time (h). (**C–F**) Expression profiles for selected genes from the loadings plot [labelled triangles in (A)]. (C) Cyclin E2 (*CCNE2*); (D) Cyclin-dependent kinase inhibitor 2C (*CDKN2C*); (E) Cyclin A2 (*CCNA2*); and (F) Cyclin-dependent kinase inhibitor 3 (*CDKN3*). (**G**) Heatmap showing the expression changes of the cell cycle genes relative to their median expression. The profiles are ordered according the genes' phase angle in the PLS model in (A). Margin colours show the genes' assigned cell cycle phase.

confirmed that the three replicates gave similar and reproducible results (Figure 1B). The trajectory also indicated the cells' gradual loss of synchrony, as the samples were gradually approaching the origin—especially at later time points.

To further analyse the cell cycle genes, we used existing annotations (2) to subdivide the genes into five main groups that represented phases G_1/S, S, G_2, G_2/M and M/G_1 of the cell cycle (Figure 1; see 'Materials and Methods' section). The assigned phases showed good correspondence with previously published phases (Supplementary Table S1); the genes missing from our list had expression patterns that were either inconsistent between replicates, not cyclic, or consistently expressed in the first cell cycle only (Supplementary Figures S2–S4; Supplementary Discussion S1). We, therefore, concluded that our PLS model could robustly identify genes that showed consistent and reproducible cyclic expression patterns within the synchronized HaCaT cells.

HaCaT cell cycle genes include both generally active housekeeping and keratinocyte-specific genes

Having established a set of genes with cell cycle-dependent expression, we set out to characterize these genes further. As cell cycle control is a central housekeeping function, we expected our set of genes to be enriched for high-CpG (HCG) promoters—a hallmark of developmentally regulated genes and housekeeping genes (12,22). We also expected the cell cycle genes to be enriched for histone H3 lysine 4 tri-methylation (H3K4me3) within their promoters, as this epigenetic mark is associated with actively transcribed genes (23,24). Indeed, although the complete set of genes followed a bimodal distribution of HCG and low-CpG (LCG) promoters (12,22), the distribution for the cell cycle genes was shifted, such that a larger percentage had HCG promoters (Figure 2A and B; $P = 2e\text{-}11$, binomial test). Similarly, ChIP-seq of H3K4me3 and H3K27me3 marks in the HaCaT cells (see 'Materials Methods' section; Supplementary Dataset S2 and S3) showed that the cell cycle genes described here were enriched for H3K4me3 (Figure 2C). Moreover, the genes were also enriched for H3K4me3 marks that were present in and common for all of seven other cell lines (13,14) (Figure 2D), which is consistent with these genes having housekeeping functions common to many cell types.

The cell cycle genes that showed inconsistent or no histone modifications within their promoters were highly enriched among LCG genes (Figure 2E and F). Tissue-specific genes are commonly enriched among LCGs, which suggested that this subset of cell cycle expressed genes that have LCG promoters with no or inconsistent H3K4me3 modifications might be cell type-specific for HaCaT. To test this possibility, we first investigated whether all the human LCG genes without consistent H3K4me3 within the seven cell lines (LCG/NoH3K4me3) were enriched within GO categories or KEGG pathways (18). As expected, the set of all LCG/NoH3K4me3 genes was highly enriched for tissue specific functions, whereas the set of all HCG/H3K4me3 genes was enriched for housekeeping categories and pathways (Supplementary Figure S5, Supplementary Dataset S4). Second, by accounting for such background enrichment within the nine subgroups, we found both the HCG/H3K4me3 cell cycle genes and the intermediate CpG (ICG)/H3K4me3 cell cycle genes to be specifically enriched for cell cycle-related terms (Figure 2G). Moreover, the LCG/NoH3K4me3 cell cycle genes were

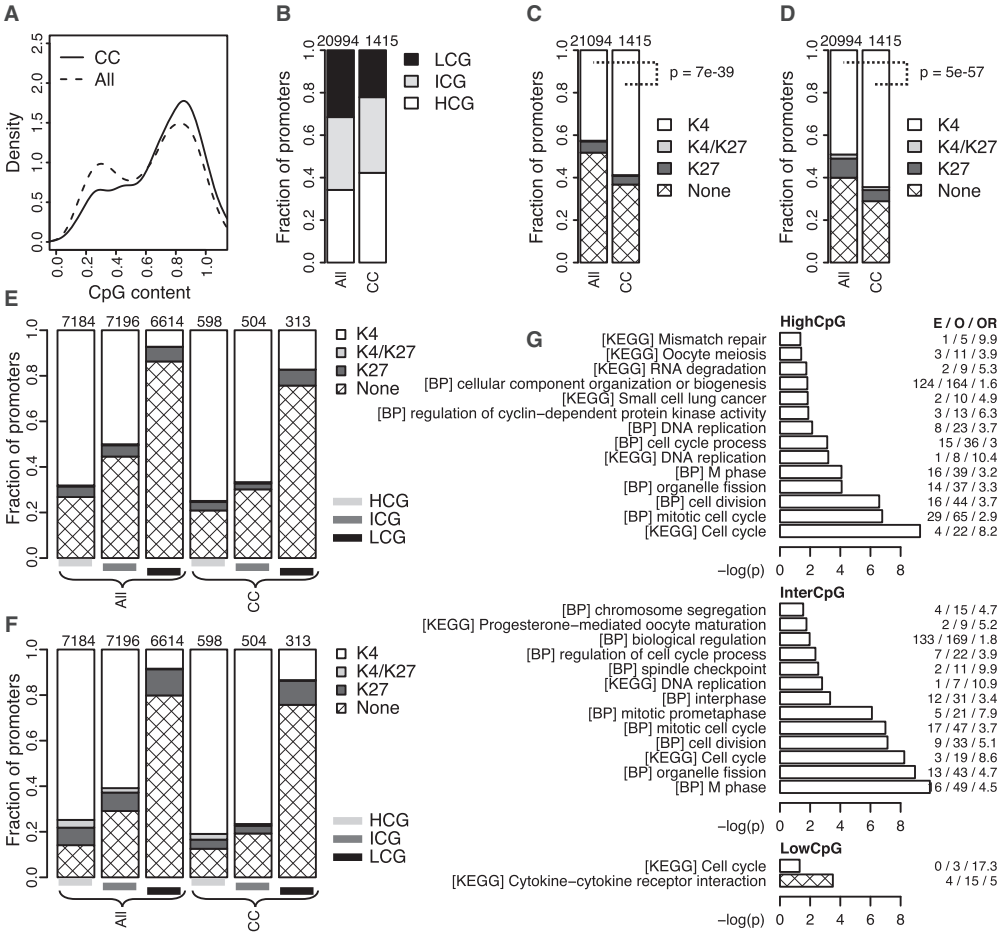


Figure 2. HaCaT cell cycle genes include both generally active housekeeping and keratinocyte-specific genes. (**A**) Distributions of CpG content in promoters for all genes included in the PLS model (all, dashed line) and cell cycle expressed genes (CC, solid line). The lines are the kernel density estimates of the two CpG content distributions. (**B**) Fraction of LCG (black), ICG (grey) and HCG (white) promoters for the two groups in (A). Numbers above bars are the total number of non-redundant promoter regions. (**C and D**) Fraction of promoters that in HaCaT (C) or in seven different cell lines (D) contain exclusive H3K4me3 or H3K27me3 marks (K4, white; K27, dark grey), bivalent K4 and K27 marks (K4/K27, light grey), or no or inconsistent marks (None, diamonds). *P*-values are the results of binomial tests comparing the fraction of H3K4me3-marked genes in the two sets. (**E and F**) Fractions from (C: E) and (D: F) subdivided into LCG, ICG and HCG (black, dark grey and light grey lines) promoters. (**G**) GO biological process (BP) terms or KEGG pathways significantly enriched ($P < 0.05$) among HCG, ICG or LCG cell cycle genes subdivided into promoters with consistent H3K4me3 (white) or H3K27me3 (no significant terms) marks, or no or inconsistent marks (diamonds). Horizontal axes show Benjamini–Hochberg corrected *P*-values ($-\log_{10}$); numbers on the right show the expected (E) and observed (O) number of genes and the corresponding odds ratios (OR).

specifically enriched for the KEGG cytokine–cytokine receptor interaction pathway, which is consistent with cytokines being important regulators of wound healing in general and keratinocyte proliferation in particular (25). Thus, at least some of the cell cycle genes seem to be cell type-specific for HaCaT cells.

Distinct cell types have exclusive cell cycle regulated genes related to the cells' specific characteristics and functions

To further determine whether cell type-specific factors could explain differences in reported cycling genes, we analysed the overlap among the cell cycle-regulated genes identified in our and previous human studies (Figure 3). Although our study identified more cell cycle genes than did the foreskin fibroblast (FF) (3) and HeLa (2) studies (Figure 3A), the cell cycle genes we identified had clear cycling patterns in HaCaT. In contrast, the genes identified in the two previous studies but missed in our study had weak or inconsistent patterns in HaCaT; compare Figure 3B and 3C.

Some of the differences between the studies were likely related to differences in microarray technologies and study

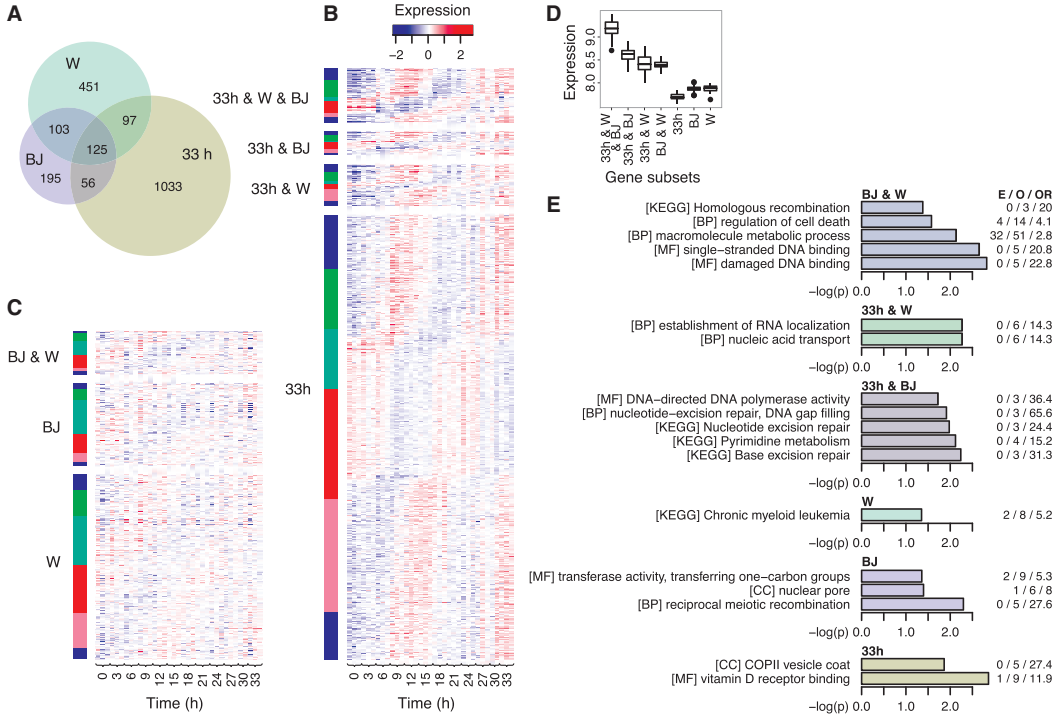


Figure 3. Cell cycle genes that are common or exclusive for foreskin fibroblasts (FF) (3), HeLa (2) and HaCaT reveal characteristics that are specific for the three cell types. (**A**) Venn diagram showing the overlap in cell cycle genes between FF (BJ), HeLa (W) and HaCaT (33 h). (**B**) Heatmap showing the expression changes of the HaCaT cell cycle genes relative to their median expression. The genes are grouped according to their overlap with the cell cycle genes from the FF and HeLa studies and are ordered according the genes' phase angle (from $-\pi$ to π) in the HaCaT 33 h PLS model. Margin colours show the genes' assigned cell cycle phase (red, G_1/S; turquoise, S; green, G_2; blue, G_2/M; and pink, M/G_1). (**C**) Heatmap showing the expression changes in HaCaT for genes identified as cell cycle regulated in FF or HeLa but not in HaCaT. The genes are grouped according to their overlap between the two studies and are ordered according to the genes' phase angle in the HaCaT 33 h PLS model. (**D**) Median expression level in HaCaT of the genes common and exclusive for the three studies. The genes are grouped according to their overlap between the three studies. Boxes, horizontal black line, and circles show the first and third quartiles, the median and outliers; whiskers show the most extreme data points up to 1.5 times the interquartile range from the box. (**E**) KEGG pathways and GO molecular function (MF), biological process (BP) and cellular component (CC) terms that are overrepresented ($P < 0.05$) and unique for six of the seven subsets of cell cycle genes common or exclusive for FF, HeLa and HaCaT; see Supplementary Table S2 for pathways and terms overrepresented among cell cycle genes common for FF, HeLa and HaCaT.

designs. Microarray results are generally reproducible across platforms and laboratories for genes with medium and high expression (26), and the genes common for the three studies were generally highly expressed in HaCaT (Figure 3D). These genes were also strongly enriched for cell cycle functions (Supplementary Table S2), indicating that these common genes represent a core set of cell cycle genes. As for design, although previous studies have either not included or used poorly matched biological replicates, we included three direct biological replicates. This, combined with the reproducibly high level of synchronization for the HaCaT cells, gave our study better statistical power to identify cyclically expressed genes.

Although the cell cycle genes common for all three studies represented a core set of human cell cycle genes, the other subsets of cell cycle genes were at least partially related to cell type-specific functions and characteristics (Figure 3E). To illustrate, the GO term 'vitamin D receptor binding' was overrepresented among the cell cycle genes exclusive for HaCaT, which is consistent with vitamin D and the vitamin D receptor's important functions in keratinocytes (27). In contrast, and consistent with HeLa's cancer origin, 'chronic myeloid leukaemia' was the only term exclusive for the HeLa cell cycle genes. Further supporting this conclusion, the KEGG pathways 'nucleotide excision repair' and 'base excision repair' were exclusively overrepresented among the cell cycle genes common for FF and HaCaT, which suggests that these two DNA repair pathways are partially dysregulated in HeLa. Although we cannot exclude that technical differences underlie some of the differences in gene expression (see 'Discussion' section), distinct cell types do seem to have distinct cell cycle regulated genes. For some cells, such as HeLa and likely other cancers (3), these distinct cell cycle genes may reflect that the cells are in an abnormal dysregulated state. Nevertheless, for more

normal non-tumorigenic cells, such as HaCaT, at least some of the distinct cell cycle genes are related to the cells' specific functions.

Cell cycle genes marked with H3K27me3 are transcribed during DNA replication

Although the cell cycle genes were generally enriched for H3K4me3, a fraction of the cell cycle genes were marked with H3K27me3 (Figure 2B)—the mark for genes silenced by Polycomb group proteins (28,29). Replication opens compacted chromatin, and although H3K27me3 marks are copied to daughter chromatin, the marks may initially be diluted, such that the normally silenced DNA is transiently accessible for transcription during this replication process. If this model was correct, we would expect that cell cycle genes marked with H3K27me3 were enriched among genes upregulated during DNA replication.

Indeed, the H3K27me3 marks in the HaCaT cells were differentially distributed among the cell cycle genes (Figure 4A). Of genes predominantly expressed in the G_1/S and S phases, 37% had H3K4me3 marks and 12% had H3K27me3 marks. In contrast, of genes expressed in G_2, G_2/M and M/G_1, 80% contained H3K4me3, whereas only 1% contained H3K27me3 marks. In total, 87% of the H3K27me3 marked cell cycle genes were expressed in G_1/S or S (P = 5e-17; one-tailed binomial test). Similarly, a larger percentage of the G_1/S and S phase genes was LCG (Figure 4B). The enrichment of H3K27me3 was not related to CpG content, as the G_1/S and S phase genes showed similar enrichment for H3K27me3 independent of CpG content (Figure 4C). These patterns were also consistent within the previously published data from seven different cell lines (13,14) (Supplementary Figure S6).

Genes upregulated in early S-phase tend to be replicated earlier than late S-phase genes

Replication occurs in distinct domains largely characterized by differences in GC content, and although these domains are partly reorganized during development, domains with high and low GC content are generally replicated early and late in S phase, respectively (30–35). If DNA replication facilitated transient transcription of silenced genes, we would, therefore, expect the genes upregulated during G_1/S to generally have a higher GC content than the S phase and other cell cycle genes. Indeed, the G_1/S genes had the highest overall GC content of the cell cycle genes (Figure 4D; P-value of 7e-3, two-sided un-paired Student's t-test with unequal variance of difference in mean between G_1/S and S phase genes) and replication-timing data from three different cell lines (30,33) indicated that the G_1/S genes on average tended to be replicated earlier than the S phase genes (Figure 4E, Student's t-test P-value of 1e-3; Supplementary Figure S7). Importantly, 63% of the cell cycle regulated genes reside in genomic regions previously shown to have stable replication timing across multiple cell types (31), and these genes had the same pattern in replication timing as the complete set of cell cycle genes (Supplementary Figure S8).

Consistent with some of the genes residing in silenced chromatin and being transiently expressed during replication, the G_1/S and S genes had a general lower expression level than the genes expressed in the other phases (Figure 4F). Moreover, the genes marked with H3K27me3 in HaCaT had the lowest expression level of the G_1/S and S phase genes (Figure 4G). The differences in expression between H3K27me3-marked genes and other subgroups were significant for all G_1/S phase groups (P-values of 2e-16, 2e-9, 3e-5 and 0.01 for K4, HCG, ICG and LCG groups, respectively; two-sided un-paired Student's t-test with unequal variance) and for the K4 and HCG S phase groups (P-values of 7e-4, 0.02, respectively; P-values of 0.1 and 0.3 for ICG and LCG groups, respectively).

Foreskin fibroblasts show similar replication-related expression of H3K27me3 marked genes

To investigate whether this replication-related cell cycle expression pattern could be found in other cell types, we first used our computational framework to reanalyse the previously published primary foreskin fibroblast data (3) (Supplementary Figure S9). As in the HaCaT cells, the majority (64%) of the genes that had significant cell cycle expression and were marked with H3K27me3 in fibroblasts were upregulated during DNA replication (Figure 5A); the mark was especially enriched for the G_1/S genes (P = 0.009; one-tailed binomial test for enrichment). In contrast to HaCaT, the majority of the fibroblast cell cycle genes marked with H3K27me3 was also marked with H3K4me3. The fibroblast cell cycle genes also showed a similar trend in GC content as the HaCaT genes, although the difference between the genes expressed early (G_1/S) and late (S) during DNA replication was not significant (Figure 5B; P = 0.1, Student's t-test).

Second, we measured the gene expression profiles in HeLa cells synchronized in the G_2/M transition by nocodazole and mitotic shake-off and used PLS to identify genes with cell cycle-dependent expression (Supplementary Figures S10 and S11). Again, the majority (52%) of the cell cycle genes that were marked with H3K27me3 in HeLa was upregulated during DNA replication, but the differences between the phases were not significant (Figure 5C; P = 0.2). Moreover, the cell cycle genes again showed a similar trend in GC content as the HaCaT, such that the G_1/S genes had a higher GC content than the S genes (Figure 5D; P = 0.01, Student's t-test). Thus, replication-related expression of H3K27me3-marked genes does not seem to be a strong characteristic of all cell types. It is relevant to note, however, that the characteristic is absent in the most abnormal of the three cells, whereas HaCaT and the primary fibroblasts both have significant replication-related expression of H3K27me3-marked genes.

DISCUSSION

Previous studies of the gene expression patterns during the human cell division cycle have indicated marked

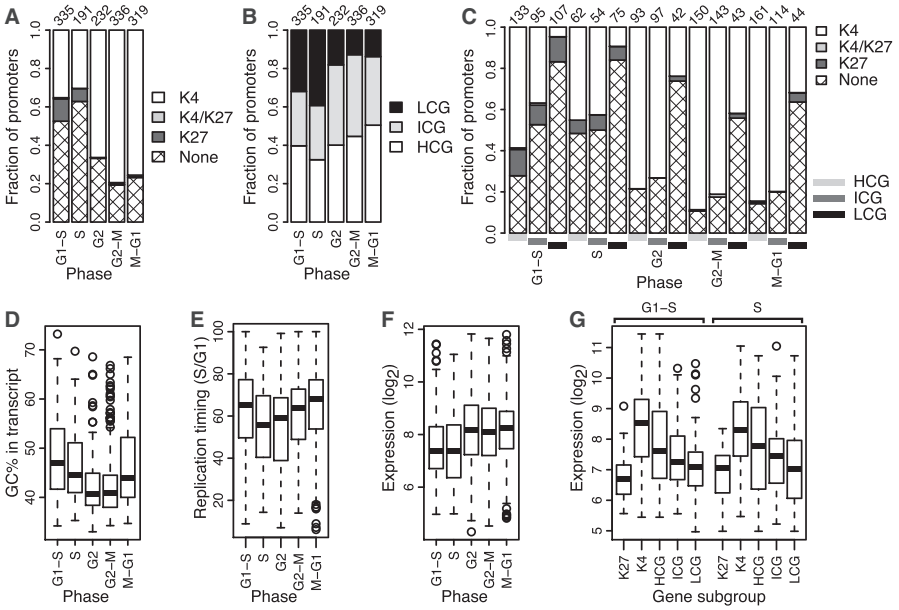


Figure 4. G_1/S and S phase genes are enriched for H3K27me3 marks and show replication-related expression. (**A**) Distribution of HaCaT H3K4me3 and H3K27me3 marks in promoters of genes upregulated in G_1/S, S, G_2, G_2/M and M/G_1 phases; see Figure 2. (**B**) Fractions of LCG, ICG and HCG promoters for genes upregulated in the five cell cycle phases. (**C**) Distribution of histone marks from (A) subdivided by promoter CpG content. (**D–F**) GC% (D), replication timing (E) and expression level (F) of genes expressed in the five cell cycle phases. GC% is the GC% in the region defined by the gene's annotated transcription start and end sites; replication timing is the median S/G_1 ratio within embryonic stem cells (33) of regions overlapping each gene, such that values close to 100 and 0 represent early and late replication, respectively; expression is the gene's average microarray expression. Boxes, horizontal black line, and circles show the first and third quartiles, the median and outliers; whiskers show the most extreme data points up to 1.5 times the interquartile range from the box. (**G**) Expression levels of specific subgroups of G_1/S and S phase genes.

differences between primary and cancer cells (3), but our results show that distinct cell types have distinct cell cycle expressed genes that are related to the cells' functions. Specifically, by profiling the gene expression in synchronized human HaCaT keratinocytes, we have identified a set of cell cycle expressed genes containing known marks for cell type-specific expression, such as low CpG content in their promoters and few or inconsistent promoter histone marks. These genes were enriched for functions important for keratinocyte proliferation, such as cytokine–cytokine receptor interaction.

Cell cycle regulation is an important housekeeping function and accordingly, most of the cell cycle expressed genes identified in our study have high CpG promoters and the active chromatin mark H3K4me3 in multiple cell lines. Nevertheless, we have also shown that some genes with low CpG promoters or the silent chromatin mark H3K27me3 have cell cycle dependent expression. These genes are mostly expressed during S phase in a replication-related pattern, which suggests that DNA replication can open some epigenetically silenced loci for transcription.

Replication-dependent transcription is a mechanism that can achieve precise transcriptional regulation during the cell cycle, as normally silenced genes will only be transcribed when their loci are replicated in S-phase. Compared with the other cell cycle genes, the Polycomb-regulated cell cycle genes in HaCaT were significantly enriched for the GO terms 'GO:0007267 cell–cell signalling' and 'GO:0046903 secretion' (Benjamini–Hochberg-corrected *P*-values of 0.007 and 0.02, respectively). In HaCaT cells, replication-dependent transcription may, therefore, be a regulatory mechanism for coordinating cell–cell signalling with the cell cycle.

Lanzuolo *et al.* (36) have earlier reported that two Polycomb-repressed homeotic genes show replication-dependent expression in *Drosophila* S2 cells. Our results extend this observation to humans and show that many other genes with Polycomb-repressive marks have replication-dependent expression. We also note that cis-encoded RNAs do play a role in Polycomb-mediated targeting and regulation (37–39). Given that Polycomb regulates all transcripts in targeted loci, replication-dependent expression may be a mechanism for expressing such cis-encoded Polycomb-associated RNAs at existing Polycomb-regulated loci.

When comparing our list of HaCaT cell cycle genes with the cell cycle genes from the previous HeLa (2) and primary foreskin fibroblast studies (3), we found several differences, such that only 125 genes were common for all three studies (Figure 3A). Several technical factors, such as differences in microarray technology, analytical

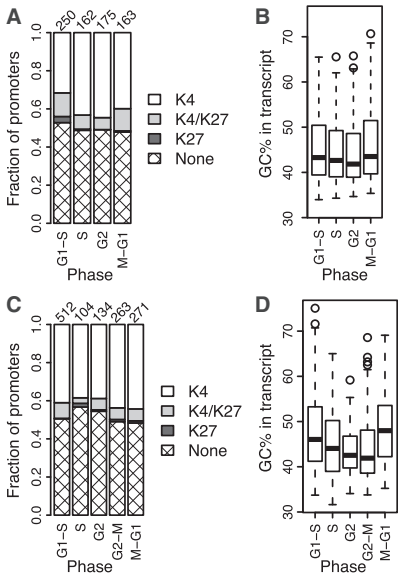


Figure 5. H3K27me3 marks are enriched in genes upregulated during DNA replication in primary foreskin fibroblasts. (**A**) Distribution of H3K4me3 and H3K27me3 marks in promoters of genes upregulated in G_1/S, S, G_2 and M/G_1 phases in primary foreskin fibroblasts synchronized by serum starvation [expression data from (3); ChIP-seq data from normal adult dermal fibroblasts (13,14)]. (**B**) GC% of the genes from (A). (**C**) Distribution of H3K4me3 and H3K27me3 marks in promoters of genes upregulated in G_1/S, S, G_2, G_2/M and M/G_1 phases in HeLa cells synchronized by nocodazole and mitotic shake-off [ChIP-seq data from HeLa (13,14)]. (**D**) GC% of the genes from (C). See Figures 2 and 4 for details on graphs.

approach and experimental design, likely contributed to these differences, for example, the genes that were common for all three studies were highly expressed, whereas the genes that were exclusive for our study were lowly expressed in HaCaT. We could detect these HaCaT-specific genes because our three matched biological replicates gave increased statistical power to identify weak, but consistent, cyclic expression patterns.

In addition to these technical differences, our results also show that underlying biological differences between the cells can explain some of the differences in cell cycle expressed genes. Specifically, our results show that in addition to the core genes that have cell cycle-dependent expression in multiple cell types, some genes related to the cell type's specific functions and characteristics also have cyclic expression. This is in contrast to a previous report, which indicated that such cell-type specific expression mostly reflects differences between primary and cancer cells (3). These cancer-related differences are evident in our analyses as well, but we also find that genes and pathways important to keratinocyte function, such as cytokine–cytokine receptor interaction (25) and vitamin D receptor binding, (27) have cell cycle-dependent expression. Both functions are important for normal keratinocyte proliferation, which makes it unlikely that these results are an artifact of HaCaT's transformed state or technical differences between the studies. Instead, the cell cycle-dependent expression of these genes likely reflects that processing of proliferative signals from the extracellular environment should be coordinated with the cell cycle.

We should note that as we have measured expression changes in total cellular RNA, some of the expression changes we have detected could potentially be explained by cell cycle-dependent changes in RNA processing or RNA stability instead of RNA transcription. Moreover, our approach of chemically synchronizing cells can likely explain some of the changes in gene expression in our data, as the synchronization procedure likely triggers stress-related responses that could both give false-positive cyclic profiles and mask the cell-cycle-dependent expression of some true cell cycle genes. This masking effect could potentially explain why we did not detect the known cell-cycle genes CDKN1A and VEGFC to have cell cycle-dependent expression in HaCaT (Supplementary Figure S2C and D). Our analyses, including the good overlap between the genes we identified and known cell-cycle-expressed genes, strong enrichment for cell-cycle-related functions and overlap with previous studies, do suggest, however, that the majority of the genes we identified are true cell cycle genes.

In summary, by combining robust statistics with the power of matched biological replicates, we have identified a set of genes that show consistent cell cycle-dependent expression in HaCaT cells. These genes include both general cell cycle genes and genes specific for HaCaT function. Some of the cell cycle genes also have H3K27me3 marks—a histone modification commonly associated with Polycomb-silenced genes. We show that in both HaCaT and primary foreskin fibroblasts, these genes are upregulated during DNA replication, which is consistent with a mechanism where DNA replication can open some Polycomb-repressed loci for transcription.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2, Supplementary Figures 1–11, Supplementary Methods, Supplementary Datasets 1–4 and Supplementary References [2,3,18,30,31,33,40].

ACKNOWLEDGEMENTS

The authors thank N.B. Liabakk for technical assistance and A.D. Riggs and O. Snøve Jr for critical reading. The sequencing service was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the 'Functional Genomics' and 'Infrastructure' programs of the Norwegian Research Council and the Southeastern Regional Health Authorities of Norway.

FUNDING

Norwegian Functional Genomics Program of the Norwegian Research Council (to H.E.K. and P.S.); Norwegian Cancer Association (to H.E.K.); Cancer Fund at St. Olav's Hospital Trondheim (to H.E.K. and P.S.); Svanhild and Arne Must Fund for Medical Research (to H.E.K.); European Community program IHP-MCIF-99-1, contract HPMF-CT-2001-01290 (to J.P.-D.). Funding for open access charge: Norwegian Research Council.

Conflict of interest statement. None declared.

REFERENCES

1. Cho,R.J., Huang,M., Campbell,M.J., Dong,H., Steinmetz,L., Sapinoso,L., Hampton,G., Elledge,S.J., Davis,R.W. and Lockhart,D.J. (2001) Transcriptional regulation and function during the human cell cycle. *Nat. Genet.*, **27**, 48–54.
2. Whitfield,M.L., Sherlock,G., Saldanha,A.J., Murray,J.I., Ball,C.A., Alexander,K.E., Matese,J.C., Perou,C.M., Hurt,M.M., Brown,P.O. *et al.* (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell.*, **13**, 1977–2000.
3. Bar-Joseph,Z., Siegfried,Z., Brandeis,M., Brors,B., Lu,Y., Eils,R., Dynlacht,B.D. and Simon,I. (2008) Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proc. Natl Acad. Sci. USA*, **105**, 955–960.
4. Iyer,V.R., Eisen,M.B., Ross,D.T., Schuler,G., Moore,T., Lee,J.C., Trent,J.M., Staudt,L.M., Hudson,J. Jr, Boguski,M.S. *et al.* (1999) The transcriptional program in the response of human fibroblasts to serum. *Science*, **283**, 83–87.
5. Beyrouthy,M.J., Alexander,K.E., Baldwin,A., Whitfield,M.L., Bass,H.W., McGee,D. and Hurt,M.M. (2008) Identification of G1-regulated genes in normally cycling human cells. *PLoS One*, **3**, e3943.
6. Spellman,P.T., Sherlock,G., Zhang,M.Q., Iyer,V.R., Anders,K., Eisen,M.B., Brown,P.O., Botstein,D. and Futcher,B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell*, **9**, 3273–3297.
7. Cho,R.J., Campbell,M.J., Winzeler,E.A., Steinmetz,L., Conway,A., Wodicka,L., Wolfsberg,T.G., Gabrielian,A.E., Landsman,D., Lockhart,D.J. *et al.* (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell*, **2**, 65–73.
8. Laub,M.T., McAdams,H.H., Feldblyum,T., Fraser,C.M. and Shapiro,L. (2000) Global analysis of the genetic network controlling a bacterial cell cycle. *Science*, **290**, 2144–2148.
9. Rustici,G., Mata,J., Kivinen,K., Lio,P., Penkett,C.J., Burns,G., Hayles,J., Brazma,A., Nurse,P. and Bahler,J. (2004) Periodic gene expression program of the fission yeast cell cycle. *Nat. Genet.*, **36**, 809–817.
10. Bolstad,B.M., Irizarry,R.A., Astrand,M. and Speed,T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
11. Rhead,B., Karolchik,D., Kuhn,R.M., Hinrichs,A.S., Zweig,A.S., Fujita,P.A., Diekhans,M., Smith,K.E., Rosenbloom,K.R., Raney,B.J. *et al.* (2010) The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.*, **38**, D613–D619.
12. Weber,M., Hellmann,I., Stadler,M.B., Ramos,L., Paabo,S., Rebhan,M. and Schubeler,D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.*, **39**, 457–466.
13. Rosenbloom,K.R., Dreszer,T.R., Pheasant,M., Barber,G.P., Meyer,L.R., Pohl,A., Raney,B.J., Wang,T., Hinrichs,A.S., Zweig,A.S. *et al.* (2010) ENCODE whole-genome data in the UCSC Genome Browser. *Nucleic Acids Res.*, **38**, D620–D625.
14. Birney,E., Stamatoyannopoulos,J.A., Dutta,A., Guigo,R., Gingeras,T.R., Margulies,E.H., Weng,Z., Snyder,M., Dermitzakis,E.T., Thurman,R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
15. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.*, **17**, 10–12.
16. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
17. Zang,C., Schones,D.E., Zeng,C., Cui,K., Zhao,K. and Peng,W. (2009) A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*, **25**, 1952–1958.
18. Falcon,S. and Gentleman,R. (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics*, **23**, 257–258.
19. Boukamp,P., Petrussevska,R.T., Breitkreutz,D., Hornung,J., Markham,A. and Fusenig,N.E. (1988) Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J. Cell Biol.*, **106**, 761–771.
20. Johansson,D., Lindgren,P. and Berglund,A. (2003) A multivariate approach applied to microarray data for identification of genes with cell cycle-coupled transcription. *Bioinformatics*, **19**, 467–473.
21. Gidskehaug,L., Anderssen,E., Flatberg,A. and Alsberg,B.K. (2007) A framework for significance analysis of gene expression data using dimension reduction methods. *BMC Bioinformatics*, **8**, 346.
22. Saxonov,S., Berg,P. and Brutlag,D.L. (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl Acad. Sci. USA*, **103**, 1412–1417.
23. Mikkelsen,T.S., Ku,M., Jaffe,D.B., Issac,B., Lieberman,E., Giannoukos,G., Alvarez,P., Brockman,W., Kim,T.K., Koche,R.P. *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448**, 553–560.
24. Kouzarides,T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.
25. Werner,S. and Grose,R. (2003) Regulation of wound healing by growth factors and cytokines. *Physiol. Rev.*, **83**, 835–870.
26. Kuo,W.P., Liu,F., Trimarchi,J., Punzo,C., Lombardi,M., Sarang,J., Whipple,M.E., Maysuria,M., Serikawa,K., Lee,S.Y. *et al.* (2006) A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies. *Nat. Biotechnol.*, **24**, 832–840.
27. Cianferotti,L., Cox,M., Skorija,K. and Demay,M.B. (2007) Vitamin D receptor is essential for normal keratinocyte stem cell function. *Proc. Natl Acad. Sci. USA*, **104**, 9428–9433.
28. Beisel,C. and Paro,R. (2011) Silencing chromatin: comparing modes and mechanisms. *Nat. Rev. Genet.*, **12**, 123–135.
29. Margueron,R. and Reinberg,D. (2011) The Polycomb complex PRC2 and its mark in life. *Nature*, **469**, 343–349.
30. Woodfine,K., Fiegler,H., Beare,D.M., Collins,J.E., McCann,O.T., Young,B.D., Debernardi,S., Mott,R., Dunham,I. and Carter,N.P. (2004) Replication timing of the human genome. *Hum. Mol. Genet.*, **13**, 191–202.
31. Hansen,R.S., Thomas,S., Sandstrom,R., Canfield,T.K., Thurman,R.E., Weaver,M., Dorschner,M.O., Gartler,S.M. and Stamatoyannopoulos,J.A. (2010) Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA*, **107**, 139–144.
32. Farkash-Amar,S., Lipson,D., Polten,A., Goren,A., Helmstetter,C., Yakhini,Z. and Simon,I. (2008) Global organization of replication time zones of the mouse genome. *Genome Res.*, **18**, 1562–1570.
33. Desprat,R., Thierry-Mieg,D., Lailler,N., Lajugie,J., Schildkraut,C., Thierry-Mieg,J. and Bouhassira,E.E. (2009) Predictable dynamic program of timing of DNA replication in human cells. *Genome Res.*, **19**, 2288–2299.
34. Costantini,M. and Bernardi,G. (2008) Replication timing, chromosomal bands, and isochores. *Proc. Natl Acad. Sci. USA*, **105**, 3433–3437.

35. Hiratani,I., Ryba,T., Itoh,M., Yokochi,T., Schwaiger,M., Chang,C.W., Lyou,Y., Townes,T.M., Schubeler,D. and Gilbert,D.M. (2008) Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.*, **6**, e245.
36. Lanzuolo,C., Lo Sardo,F., Diamantini,A. and Orlando,V. (2011) PcG complexes set the stage for epigenetic inheritance of gene silencing in early S phase before replication. *PLoS Genet.*, **7**, e1002370.
37. Plath,K., Fang,J., Mlynarczyk-Evans,S.K., Cao,R., Worringer,K.A., Wang,H., de la Cruz,C.C., Otte,A.P., Panning,B. and Zhang,Y. (2003) Role of histone H3 lysine 27 methylation in X inactivation. *Science*, **300**, 131–135.
38. Zhao,J., Sun,B.K., Erwin,J.A., Song,J.J. and Lee,J.T. (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science*, **322**, 750–756.
39. Pandey,R.R., Mondal,T., Mohammad,F., Enroth,S., Redrup,L., Komorowski,J., Nagano,T., Mancini-Dinardo,D. and Kanduri,C. (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell*, **32**, 232–246.
40. Ernst,J., Kheradpour,P., Mikkelsen,T.S., Shoresh,N., Ward,L.D., Epstein,C.B., Zhang,X., Wang,L., Issner,R., Coyne,M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43–49.

Supporting Online Material

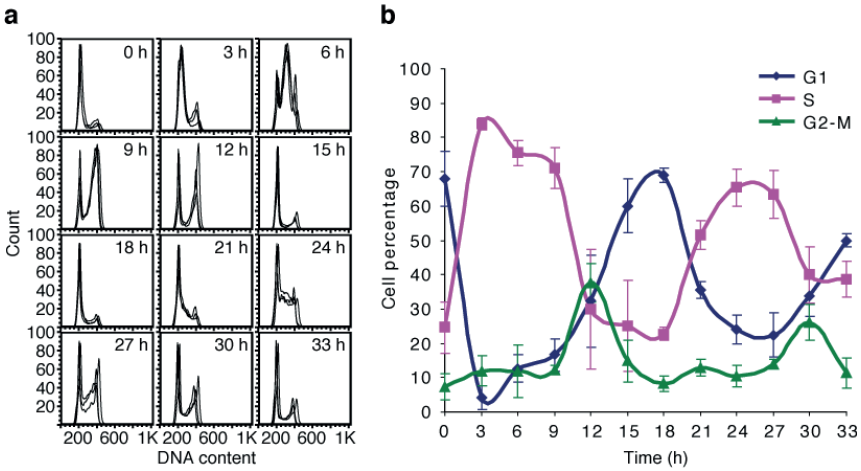
Supplementary Methods 1 – ChIP-seq

Plates (75cm^2) with ~6 million HaCaT cells were washed twice with PBS, then cross-linked with 0.5% formaldehyde at room temperature for 10 mins, following which cross-linking was quenched by the addition of glycine to 125 mM. Cells were again washed with PBS, then scraped from plates in PBS supplemented with 1mM EDTA and protease inhibitors. Cross-linked cells were collected by centrifugation at 1500 G, and resuspended in cold RIPA buffer (10 mM Tris pH 8.0, 1 mM EDTA, 140 mM NaCl, 1 % Triton X-100, 0.1% SDS, 0.1 % Na-Deoxycholate) supplemented with protease inhibitors at a density of 1 x 10^7 cells/ml. Lysed cells were then sonicated in TPX tubes (Diagenode) using a Bioruptor (Diagenode) for 12 mins on full power, then debris was removed by centrifugation in a chilled mircofuge at 16000 G for 15 mins. Soluble chromatin supernatant equivalent to 3.8 x 10^6 cells / IP was pre-cleared using 25 µl of a 1:1 mix of protein A and protein G Dynabeads (Invitrogen), then used for overnight immunoprecipitation using either 4 µg anti-H3K4me3 antisera (Diagenode s.a., Cat. # pAb-003-050; Lot # A2-002P) or 2 µg anti H3K27me3 (Upstate Biotechnology, Cat. # 07-449; Lot # DAM1387952). Antibody-protein complexes were immunoprecipitated with protein A/G Dynabeads, then washed 5 times with RIPA buffer supplemented with protease inhibitors, once with LiCl wash buffer (250 mM LiCl, 10 mM Tris pH 8.0, 1 mM EDTA, 0.5 % Igepal CA-630, 0.5 % Na-deoxycholate) supplemented with protease inhibitors, and once in TE. Immunoprecipitated DNA was isolated by successive RNaseA and proteinase K treatment, followed by purification over Genomic DNA cleanup & concentrator columns (Zymo Research Corp.). Two ng of immunoprecipitated material, or input chromatin isolated as above, was used for library preparation using TruSeq™ DNA Sample Preparation reagents (Illumina), and submitted to the Norwegian Sequencing Centre for sequencing on a HiSeq 2000 (Illumina). IP and control input libraries were clustered on a single lane of an Illumina flowcell using TruSeq™ SR Cluster Kit v3 reagents and subjected to 50 bases of sequencing using TruSeq™ SBS v3 reagents (Illumina). Image analysis and base calling was performed using Illumina's RTA software version 1.12.4.2. Demultiplexing and conversion from bcl file to fastq file was performed using CASAVA v1.8.2. Reads were filtered to remove those with low base call quality using Illumina's default chastity criteria.

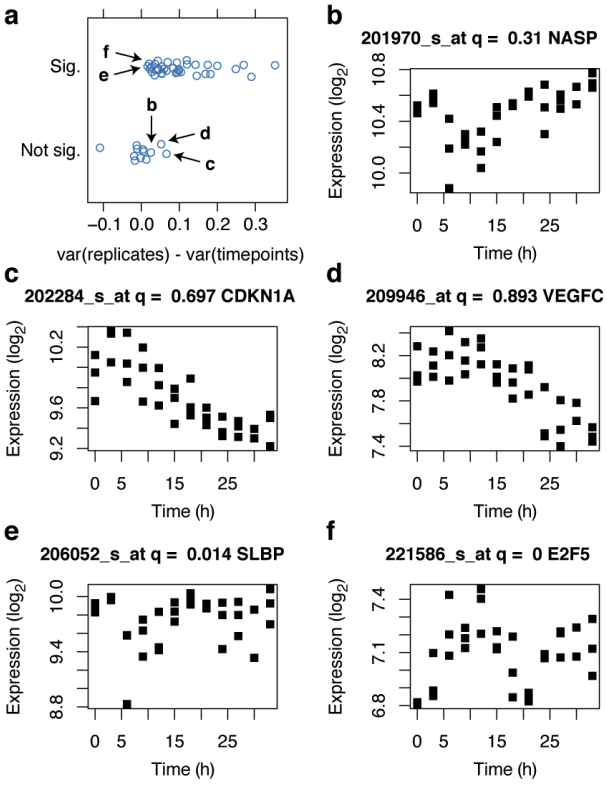
Supplementary Discussion 1 – A PLS model based only on the first 24 hours after release identifies additional cell cycle regulated genes at the expense of additional false positives

When studying the list of genes with previously published phases and the genes on that list not detected by our PLS model to be cell cycle expressed in HaCaT, we noticed that 7 of the 12 missing genes were previously published to be expressed in S phase (Supplementary Table 1). Closer inspection of the expression profiles indicated that some

of these genes appeared to have a cyclic expression pattern during the first cell division. We therefore created a new PLS model based on only the first 24 hours after release (Supplementary Fig. 3). This model identified several additional cell cycle genes, including 4 of the 12 genes missed by the PLS model based on all the 33 hours after release (Supplementary Fig. 4, Supplementary Table 1). Although the 24 hours PLS model gave an increased coverage of known cell cycle genes of about 8%, this was at the expense of about 9% additional significant cell cycle genes than the 33 hours model. Using all the available probes on the chip instead of only the probe that showed the highest variation for each Entrez gene further increased the coverage of known cell cycle genes to 42 of 48 genes, but also increased the total number of significant cell cycle genes to 2081 (data not shown). To limit the influence of potential false positives and focus our analyses on genes that showed consistent cyclic expression across multiple cell divisions, we therefore decided to use the 33 hours PLS model as basis for further analyses.

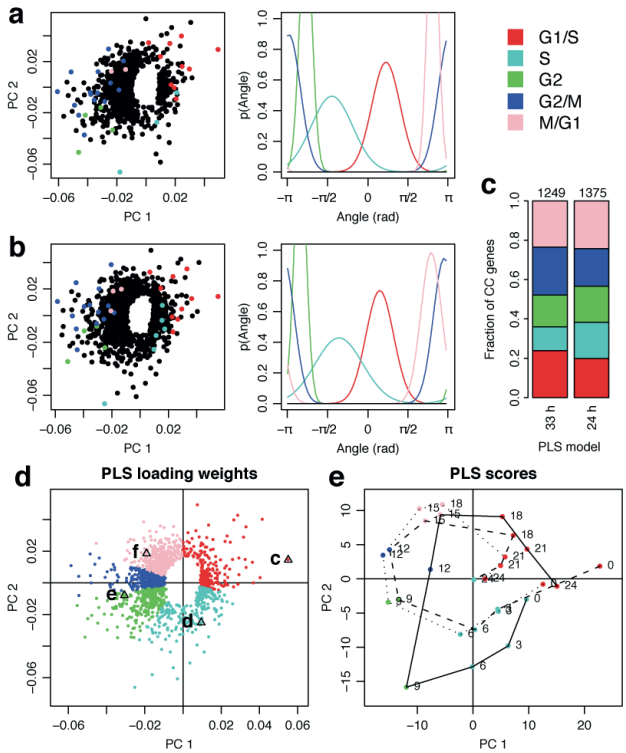


Supplementary Figure 1 - Synchrony of double thymidine blocked HaCaT cells. (**a**) Cell synchrony was monitored by flow cytometry of propidium iodide-stained cells. The figures show for each time point, superimposed profiles of three replicate synchronization experiments. Horizontal axes show DNA content (arbitrary units) and vertical axes show the number of events (cells) with the corresponding DNA content. (**b**) Percentage of cells assigned to G1, S, and G2/M phases for each of the time points analyzed. Values and error bars are averages and standard deviations (n = 3).



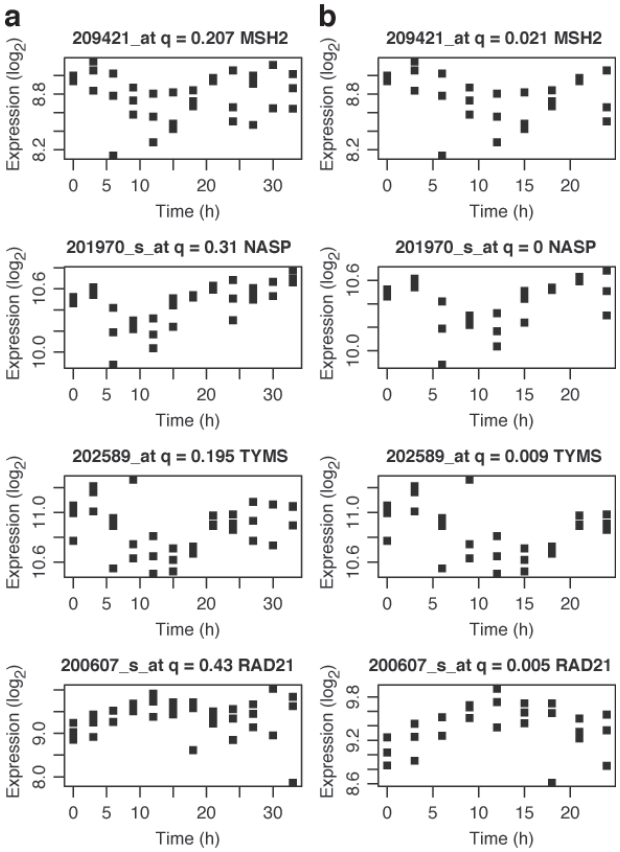
Supplementary Figure 2 - "Missing" cell cycle genes either show no cyclic expression or have more variance between replicates than "detected" cell cycle genes. (**a**) The figure shows the average expression variance across the 12 time points (var(replicates); variance averaged over the 3 replicates) minus the average expression variance across the 3 replicates (var(timepoints); variance averaged over the 12 time points) for the previously known cell cycle regulated genes (table S2) detected ("Sig.") and not detected ("Not sig.") to be significantly cell cycle regulated in HaCaT by this study. All the significant genes have higher average variance within a replicate than within time points. In contrast, 6 of the 12 "missing" genes (*BRCA1, BRCA2, DHFR, MSH2, RRM1*, and *RAD21*) have lower average variance within a replicate than within time points; that is, their expression was inconsistent and varied as much between the replicates as within the same replicate. Only three of the "missing" genes (*CDKN1A*, *NASP*, and *VEGFC*) had a difference in average variance that was greater than any of the significant genes. Arrows and labels show these three "missing" genes with the largest difference in average variance (**b-d**) and the two significant genes with the smallest difference in average variance (**e-f**). (b-f) Expression patterns for the five genes annotated in (a). The "missing" genes (b-d) do not show cyclic expression patterns

through both cell cycles, whereas the significant genes do (e-f). (b) *NASP* showed an expression pattern consistent with being up-regulated during S phase and down-regulated during G2/M phase in the first cell cycle but not in the second; that is, *NASP* was down-regulated in the first G2/M phase (9 and 12 hours after release), but not in the second G2/M phase (27 and 30 hours after release). (c-d) *CDKN1A* and *VEGFC* were highly expressed in the first G1/S transition and S phase, but were not up-regulated in the second cell cycle. Figure titles show the probe id, the false discovery rate (q), and the gene name.

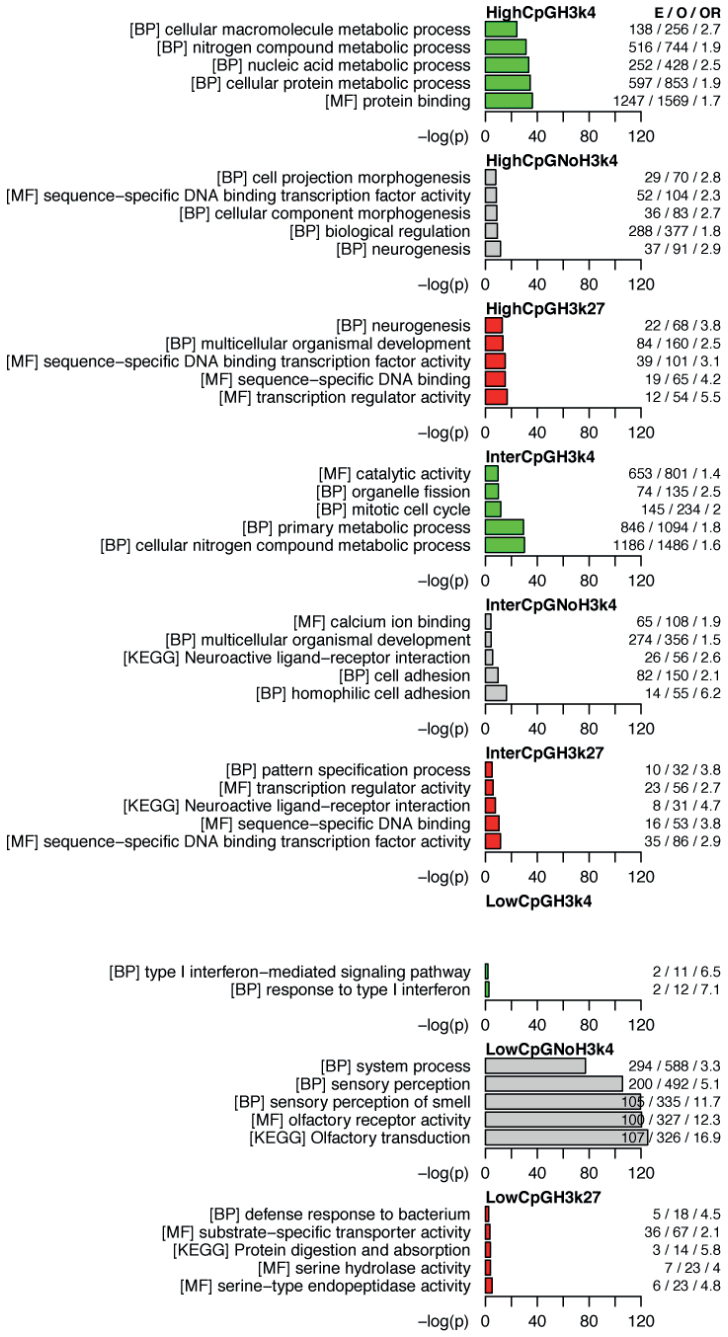


Supplementary Figure 3 – A cyclic PLS model based on the first 24 hours after release identifies additional genes expressed in S phase in HaCaT. (**a-b**) Known cell cycle genes cluster according to their published cell cycle phase. (Left) Loadings plot showing significant genes ($q < 0.05$; permutation test) and the genes with known cell cycle expression (Supplementary Table 2) in the cyclic PLS models based on all the 33 hours (a) and only the first 24 hours (b) after release from double thymidine block. The genes with known cell cycle expression are color-coded according to their known cell cycle phase (red, G1/S; turquoise, S; green, G2; blue, G2/M; and pink, M/G1). (Right) Phase angle probability distributions for the five cell cycle phases estimated based on the phase angles of the known cell cycle genes in the loadings plot (see Methods). Horizontal axis shows the phase angle (radians); vertical axis shows the probability. Note that because the horizontal axis represents a linearization of a circle, the distributions "wrap around" from π to $-\pi$. (**c**) The 24 hours PLS model identifies more genes as significant cell cycle genes than the 33 hours PLS model and the largest increase is for genes expressed in S phase. The bar graphs show the relative distribution of significant genes ($q < 0.05$) in the five cell cycle phases for the 33 hours and 24 hours PLS models. The numbers above the bars show the total number of significant genes in each model. Note that the 24 hours model gave poorer separation of the G2/M and M/G1 phases than the 33

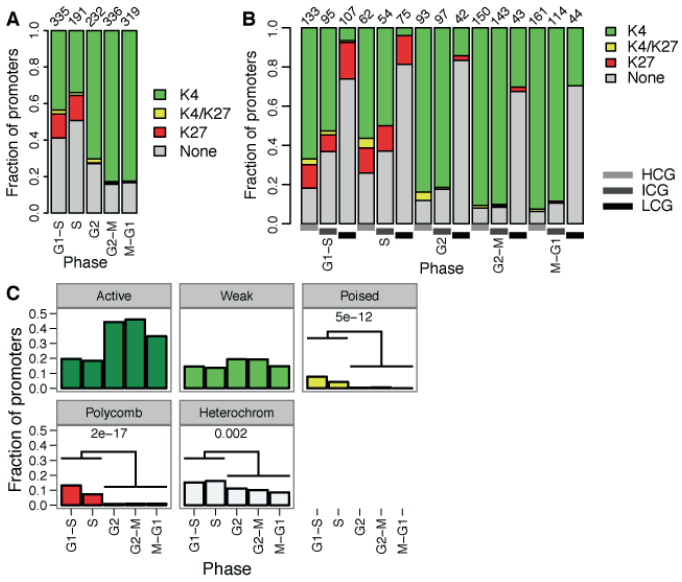
hours model (compare phase angle probability distributions in (a) and (b)), which resulted in the 24 hours model having fewer genes in G2/M phase than the 33 hours model had. (**d-e**) Loadings (d) and scores (e) plots for the cyclic PLS model based on only the first 24 hours after release from double thymidine block. The solid, dashed, and dotted lines in (e) show the trajectories for the three biological replicates. Labeled triangles in (d) show the location of the four selected genes in Fig. 1: (c) *CCNE2*; (d) *CDKN2C*; (e) *CCNA2*; and (f) *CDKN3*. See Figure 1 for the four genes' expression profiles and additional details regarding loadings and scores plots.



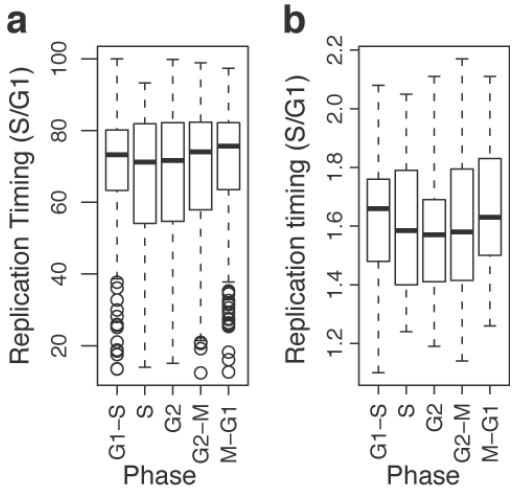
Supplementary Figure 4 – Expression patterns of the four known cell cycle regulated genes identified by the 24 hours PLS model (b) but missed by the 33 hours PLS model (a). Figure titles show the probe id, the false discovery rate (q), and the gene name. Except for *RAD21*, the additional genes detected by the 24 hours model were predominantly expressed in G1/S or S phase.


HighCpGH3k4
E / O / OR
[BP] cellular macromolecule metabolic process
138 / 256 / 2.7
[BP] nitrogen compound metabolic process
516 / 744 / 1.9
[BP] nucleic acid metabolic process
252 / 428 / 2.5
[BP] cellular protein metabolic process
597 / 853 / 1.9
[MF] protein binding
1247 / 1569 / 1.7
−log(p) 0 40 80 120
HighCpGNoH3k4
[BP] cell projection morphogenesis
29 / 70 / 2.8
[MF] sequence−specific DNA binding transcription factor activity
52 / 104 / 2.3
[BP] cellular component morphogenesis
36 / 83 / 2.7
[BP] biological regulation
288 / 377 / 1.8
[BP] neurogenesis
37 / 91 / 2.9
−log(p) 0 40 80 120
HighCpGH3k27
[BP] neurogenesis
22 / 68 / 3.8
[BP] multicellular organismal development
84 / 160 / 2.5
[MF] sequence−specific DNA binding transcription factor activity
39 / 101 / 3.1
[MF] sequence−specific DNA binding
19 / 65 / 4.2
[MF] transcription regulator activity
12 / 54 / 5.5
−log(p) 0 40 80 120
InterCpGH3k4
[MF] catalytic activity
653 / 801 / 1.4
[BP] organelle fission
74 / 135 / 2.5
[BP] mitotic cell cycle
145 / 234 / 2
[BP] primary metabolic process
846 / 1094 / 1.8
[BP] cellular nitrogen compound metabolic process
1186 / 1486 / 1.6
−log(p) 0 40 80 120
InterCpGNoH3k4
[MF] calcium ion binding
65 / 108 / 1.9
[BP] multicellular organismal development
274 / 356 / 1.5
[KEGG] Neuroactive ligand−receptor interaction
26 / 56 / 2.6
[BP] cell adhesion
82 / 150 / 2.1
[BP] homophilic cell adhesion
14 / 55 / 6.2
−log(p) 0 40 80 120
InterCpGH3k27
[BP] pattern specification process
10 / 32 / 3.8
[MF] transcription regulator activity
23 / 56 / 2.7
[KEGG] Neuroactive ligand−receptor interaction
8 / 31 / 4.7
[MF] sequence−specific DNA binding
16 / 53 / 3.8
[MF] sequence−specific DNA binding transcription factor activity
35 / 86 / 2.9
−log(p) 0 40 80 120
LowCpGH3k4
[BP] type I interferon−mediated signaling pathway
2 / 11 / 6.5
[BP] response to type I interferon
2 / 12 / 7.1
−log(p) 0 40 80 120
LowCpGNoH3k4
[BP] system process
294 / 588 / 3.3
[BP] sensory perception
200 / 492 / 5.1
[BP] sensory perception of smell
105 / 335 / 11.7
[MF] olfactory receptor activity
100 / 327 / 12.3
[KEGG] Olfactory transduction
107 / 326 / 16.9
−log(p) 0 40 80 120
LowCpGH3k27
[BP] defense response to bacterium
5 / 18 / 4.5
[MF] substrate−specific transporter activity
36 / 67 / 2.1
[KEGG] Protein digestion and absorption
3 / 14 / 5.8
[MF] serine hydrolase activity
7 / 23 / 4
[MF] serine−type endopeptidase activity
6 / 23 / 4.8
−log(p) 0 40 80 120

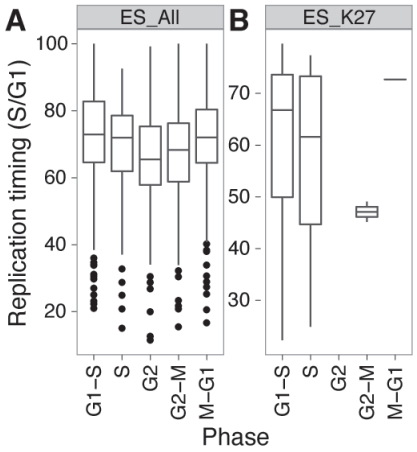
Supplementary Figure 5 – High-CpG and consistent H3K4me3 are associated with housekeeping functions whereas low-CpG and no or inconsistent H3K4me3 are associated with cell type-specific functions. The barplots show the five most significantly KEGG pathways (KEGG), or GO molecular function (MF) or biological process (BP) terms overrepresented among all genes with high, intermediate, or low CpG promoters and consistent H3K4me3, H3K27me3, or lack of consistent H3K4me3 modifications. Horizontal axes show Benjamini-Hochberg corrected p-values (–log10); numbers on the right show the expected (E) and observed (O) number of genes and the corresponding odds ratios (OR). Cell type-specific functions such as "Olfactory transduction" are strongly overrepresented among LCG/NoH3K4me3 genes whereas housekeeping functions such as "cellular metabolic process" are overrepresented among HCG/H3K4me3 genes.



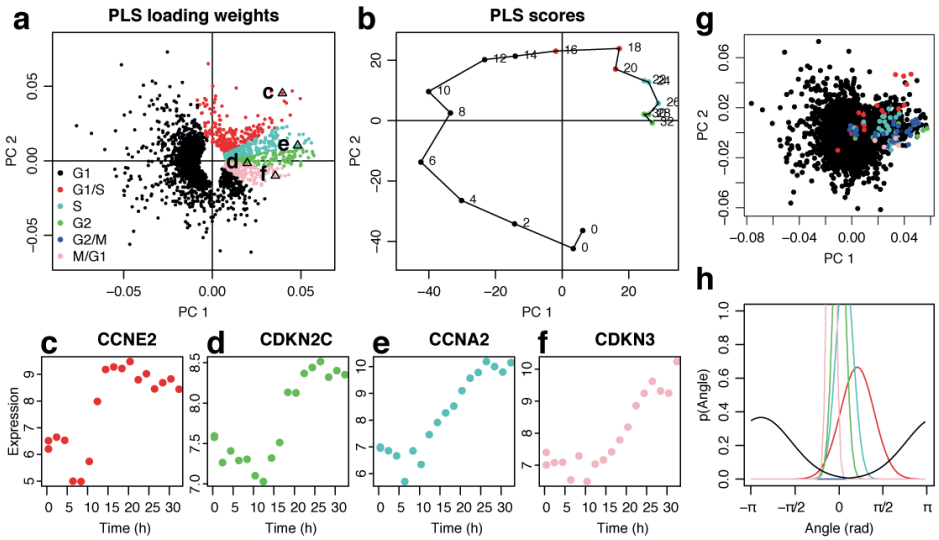
Supplementary Figure 6 - G1/S and S phase genes are consistently enriched for H3K27me3 marks within multiple cell lines. (**A**) Distribution of consistent H3K4me3 and H3K27me3 marks in promoters of genes upregulated in G1/S, S, G2, G2/M, and M/G1 phases; see Fig. 4. (**B**) Fractions of low-CpG (LCG), intermediate-CpG (ICG), and high-CpG (HCG) promoters for genes upregulated in the five cell cycle phases. (**C**) Promoter chromatin states for cell cycle genes within the NHEK keratinocyte cell line show that poised and Polycomb-regulated promoters are strongly enriched whereas promoters with heterochromatin regions are weakly enriched among genes expressed during DNA replication. The bar graphs show the fraction of promoters that contain regions classified by a published "Chromatin-state" Hidden Markov Model (1) to be Active, Weak, and Poised promoters and Polycomb-repressed and Heterochromatic regions. Lines and numbers within the graphs indicate chromatin states (Poised, Polycomb, and Heterochrom) that are enriched among genes expressed during DNA replication (G1/S and S phases; numbers are p-values from one-tailed binomial tests).



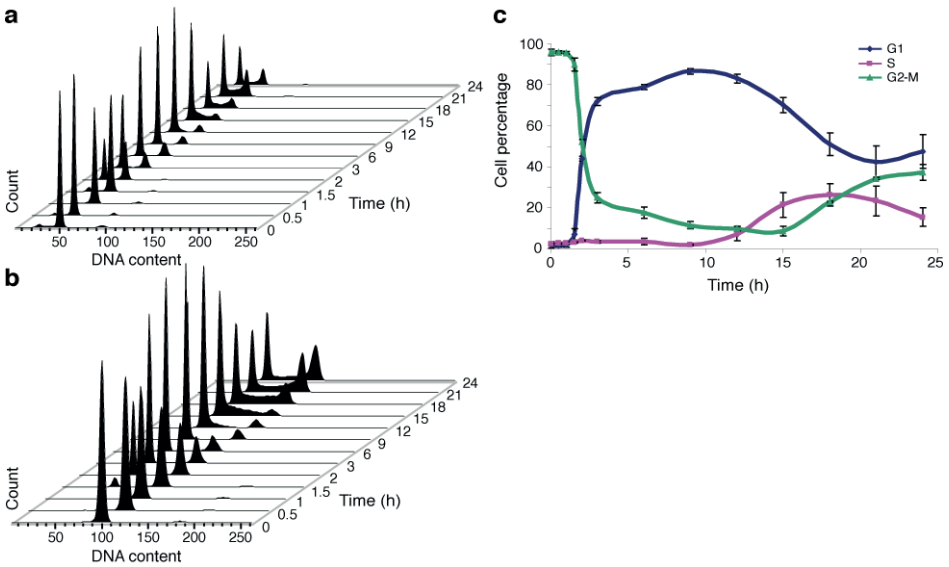
Supplementary Figure 7 – Data from different cell lines show that HaCaT G1/S genes tend to be replicated earlier than S genes. The box plots show the replication timing of cell cycle genes from HaCaT in (**a**) basophilic erythroblasts (2) and (**b**) lymphoblastoid cells (3). The replication timing is the median S/G1 ratio of regions overlapping each gene; data from sequencing-based TimEX (**a**) are normalized S/G1 ratios (see (2)); data from microarray-based hybridization experiments (**b**) are probe S/G1 ratios (see (3)). For (**a**) early and late replicating regions have values close 100 and 0, respectively, whereas for (**b**) early and late replicating regions have values close to 2 and 1, respectively. Boxes, horizontal black line, and circles show the first and third quartiles, the median, and outliers; whiskers show the most extreme datapoints up to 1.5 times the interquartile range from the box.



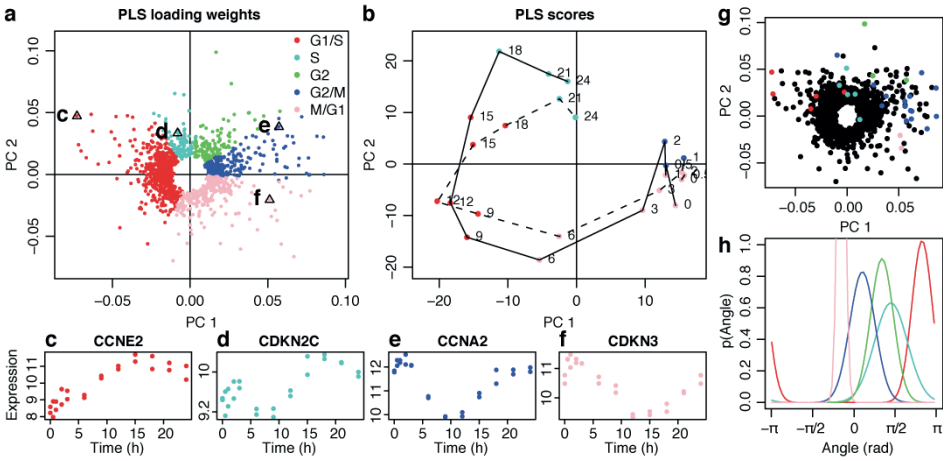
Supplementary Figure 8 – Replication timing for genes residing in regions with stable replication timing. (**A**) Replication timing within all cell cycle genes that do not overlap regions found by Hansen and colleagues to exhibit replication timing plasticity (4). The data is a subset of Figure 4E where all transcripts that overlap regions with replication timing plasticity have been removed (regions of replication timing plasticity were defined in Supplementary Table S2 in (4)) (**B**) Replication timing within the subset of genes from (**A**) that are marked with H3K27me3 within their promoters. (**A**-**B**) Replication timing is the median S/G1 ratio within embryonic stem cells (2) of regions overlapping each gene. See Supplementary Figure 7 for details on the box-and-whisker plots.



Supplementary Figure 9 – Partial least squares (PLS) regression reanalysis of serum starved primary foreskin fibroblasts identifies cell cycle genes. (**a-b**) Loadings (a) and scores (b) plots for a cyclic PLS model of the gene expression profiles from Bar-Joseph and colleagues' primary foreskin fibroblast experiment (5). The period used was 30 hours estimated based on the cells' DNA content profile (Fig. 1a in (5)). Labeled triangles in (a) show the location of the four selected genes from Figure 1 and panels (**c-f**) show their expression profiles. The genes are (c) *CCNE2*; (d) *CDKN2C*; (e) *CCNA2*; and (f) *CDKN3*. See Figure 1 for additional details regarding loadings and scores plots. (**g**) Loadings plot showing the genes with known cell cycle expression (Supplementary Table 2) in the PLS model (a). (**h**) Phase angle probability distributions for four of the five cell cycle phases. Specifically, we excluded the G2/M phase from the model, as the experiment, in contrast to the HaCaT double thymidine block experiment (Fig. 1; Supplementary Figure 3a), did not give sufficient temporal resolution to accurately separate the genes in the G2/M phase from the G2 and M/G1 phases; compare the patterns for the genes with known cell cycle expression in Supplementary Figure 3a with their pattern in (g). In Supplementary Figure 3a, these known genes are well separated, whereas in the primary foreskin fibroblast data, the genes are clustered at the end of the series and the G2/M phase genes completely overlap the G2 and M/G1 genes. Also shown in (h) is the phase angle probability distribution for the G1 phase, as estimated by the cells' DNA content profile. These phase angles had a marked lack of known cell cycle genes (see g), and were therefore excluded from the model. The final cell cycle model, which included the G1/S, S, G2, and M/G1 phases, consisted of 627 distinct genes and included 60% of the genes originally reported by Bar-Joseph and colleagues. See Methods and Supplementary Figure 3 for additional details.



Supplementary Figure 10 – Synchrony by nocodazole and mitotic shake-off blocked HeLa cells. (**a** and **b**) Cell synchrony was monitored by flow cytometry of propidium iodide-stained cells. The figures show for each of the two biological replicates, DNA profiles at each time point of the synchronization experiment. Horizontal axes show DNA content (arbitrary units) and vertical axes show the number of events (cells) with the corresponding DNA content. (**c**) Percentage of cells assigned to G1, S, and G2/M phases for each of the time points analyzed. Values and error bars are averages and standard deviations (n = 2).



Supplementary Figure 11 – Nocodazole, mitotic shake-off, and partial least squares (PLS) regression identifies HeLa cell cycle genes. (**a-b**) Loadings (a) and scores (b) plots for a cyclic PLS model of the gene expression profiles from HeLa cells synchronized by nocodazole treatment and mitotic shake-off. The period used was 25 hours estimated based on the cells' DNA content profile (Supplementary Fig. 10). Labeled triangles in (a) show the location of the four selected genes from Figure 1 and panels (**c-f**) show their expression profiles. The genes are (c) *CCNE2*; (d) *CDKN2C*; (e) *CCNA2*; and (f) *CDKN3*. See Figure 1 for additional details regarding loadings and scores plots. (**g**) Loadings plot showing significant genes ($q < 0.05$; permutation test) and the genes with known cell cycle expression (Supplementary Table 2) in the PLS model (a). (**h**) Phase angle probability distributions for the five cell cycle phases. See Methods and Supplementary Figure 3 for additional details.

Supplementary Table 1 - List of genes with previously published phases. The genes are from Table 2 in (6). "Predicted (33 h)" are the cell cycle phases assigned by this study's main PLS model for the HaCaT cells, which is based on all the time points up to 33 h after release; "Predicted (24 h)" are the cell cycle phases assigned by a PLS model based on only the time points up to 24 h after release (in HaCaT); "Whitfield" are the cell cycle phases reported based on microarray analyses of synchronized HeLa cells (6); "Actual" are previously reported cell cycle phases as compiled by (6).

Gene	Predicted (33 h)	Predicted (24 h)	Whitfield	Actual
E2F5	G2	G2	G2/M	G1
CCNE1	G1/S	G1/S	G1/S	G1/S
CCNE2	G1/S	G1/S	G1/S	G1/S
CDC25A	G1/S	G1/S	G1/S	G1/S
CDC45L	S	S	S	G1/S
CDC6	G1/S	G1/S	G1/S	G1/S
CDKN1A	ND	ND	ND	G1/S
E2F1	G1/S	G1/S	G1/S	G1/S
MCM2	G1/S	G1/S	G1/S	G1/S
MCM6	G1/S	G1/S	G1/S	G1/S
NPAT	ND	ND	G1/S	G1/S
PCNA	G1/S	G1/S	G1/S	G1/S
SLBP	G1/S	G1/S	G1/S	G1/S
BRCA1	ND	ND	S	S
BRCA2	ND	ND	ND	S
CCNG2	S	S	ND	S
CDKN2C	S	S	G2	S
DHFR	ND	ND	S	S
MSH2	ND	G1/S	G1/S	S
NASP	ND	G1/S	G1/S	S
RRM1	ND	ND	S	S
RRM2	G1/S	S	S	S
TYMS	ND	S	S	S
CCNA2	G2	G2	G2	G2
CCNF	G2	G2	G2	G2
CENPF	G2	G2	G2/M	G2
TOP2A	G2	G2	G2	G2
BIRC5	M/G1	ND	G2/M	G2/M
BUB1	G2/M	G2/M	G2/M	G2/M
BUB1B	G2/M	G2/M	G2/M	G2/M
CCNB1	G2/M	G2/M	G2/M	G2/M
CCNB2	M/G1	M/G1	G2/M	G2/M
CDK1	G2	G2	G2	G2/M
CDC20	M/G1	M/G1	G2/M	G2/M
CDC25B	G2/M	G2/M	G2/M	G2/M

CDC25C	G2/M	G2	G2	G2/M
CDKN2D	ND	ND	G2/M	G2/M
CENPA	G2	G2	G2/M	G2/M
CKS1B	G2/M	G2/M	G2	G2/M
CKS2	G2/M	G2/M	G2/M	G2/M
PLK1	G2/M	G2/M	G2/M	G2/M
AURKA	G2	G2	G2/M	G2/M
RACGAP1	G2/M	G2/M	NM	G2/M
KIF20A	G2/M	M/G1	NM	G2/M
RAD21	ND	G2/M	M/G1	M/G1
PTTG1	M/G1	M/G1	M/G1	M/G1
VEGFC	ND	ND	ND	M/G1
CDKN3	M/G1	M/G1	M/G1	M/G1

ND: Not determined; the gene was not detected as having a significant cell cycle expression pattern. NM: Not measured; the gene was not present on the microarray used by (6).

Supplementary Table 2 - Cell cycle genes common for foreskin fibroblasts (FF) (5), HeLa (6), and HaCaT are strongly enriched for cell cycle-related terms. The table shows all KEGG pathways and GO molecular function (MF), biological process (BP), and cellular component (CC) terms that are overrepresented ($p < 0.05$) among the 125 cell cycle Entrez genes common for FF, HeLa, and HaCaT. The table is sorted based on the Benjamini-Hochberg-corrected p-value (p) of a hypergeometric test for overrepresentation, which for the GO terms was conditioned on the structure of the GO graph (7). Exp. and Count are the expected and observed number of genes; OR is the corresponding odds ratio; Size is the size of the background group used in the test. Rows shaded in gray indicate terms that were not significant ($p < 0.05$) among any of the other subsets of cell cycle genes common or exclusive for FF, HeLa, or HaCaT (see fig. S6).

Ontology	Term	ID	p	OR	Exp.	Count	Size
BP	organelle fission	GO:0048285	2e-45	43.1	2.0	44	250
BP	cell division	GO:0051301	1e-33	34.6	1.7	35	230
BP	M phase	GO:0000279	3e-33	30.0	2.2	37	309
BP	cell cycle process	GO:0022402	8e-31	27.3	2.4	36	359
KEGG	Cell cycle	04110	3e-30	72.9	1.1	26	129
BP	mitotic cell cycle	GO:0000278	1e-29	28.0	2.1	34	309
BP	mitosis	GO:0007067	1e-26	40.1	1.1	26	157
CC	non-membrane-bounded organelle	GO:0043228	2e-19	6.6	17.3	61	2351
CC	nucleoplasm	GO:0005654	1e-17	9.2	5.7	36	770
CC	organelle lumen	GO:0043233	2e-17	6.6	12.3	50	1670
BP	DNA replication	GO:0006260	3e-16	22.8	1.2	20	158
CC	intracellular organelle	GO:0043229	2e-15	8.7	62.3	104	8480
BP	cellular component organization	GO:0016043	2e-14	6.5	13.4	48	1960
CC	intracellular	GO:0005622	9e-14	18.2	76.7	111	10435
CC	membrane-bounded organelle	GO:0043227	2e-12	5.4	55.3	95	7530
CC	organelle part	GO:0044422	7e-12	7.2	9.8	36	2115
BP	chromosome organization	GO:0051276	1e-11	8.9	3.7	25	460
CC	condensed chromosome	GO:0000793	1e-11	25.5	0.7	13	91
KEGG	DNA replication	03030	3e-10	48.9	0.3	9	36
CC	spindle	GO:0005819	2e-9	29.6	0.4	10	65
CC	chromosome	GO:0005694	1e-8	12.4	1.4	14	221
BP	phosphoinositide-mediated signaling	GO:0048015	2e-8	22.4	0.6	11	78
CC	condensed chromosome kinetochore	GO:0000777	3e-8	39.8	0.3	8	38
BP	DNA packaging	GO:0006323	7e-8	16.5	0.9	12	112
BP	spindle organization	GO:0007051	1e-7	138.5	0.1	6	12
BP	regulation of cyclin-dependent protein kinase activity	GO:0000079	2e-7	28.1	0.4	9	52
BP	chromosome segregation	GO:0007059	5e-7	24.9	0.5	9	59
CC	nucleus	GO:0005634	9e-7	4.9	12.8	33	3157
CC	microtubule cytoskeleton	GO:0015630	1e-6	12.6	1.0	11	180
MF	protein binding	GO:0005515	3e-6	3.6	44.1	71	6555

CC	centrosome	GO:0005813	6e-6	10.4	1.2	11	175
MF	ATP binding	GO:0005524	1e-5	3.7	10.8	31	1426
BP	G2 phase of mitotic cell cycle	GO:0000085	1e-5	513.4	0.0	4	5
BP	cell cycle	GO:0007049	2e-5	16.3	0.7	9	175
MF	microtubule motor activity	GO:0003777	2e-5	16.6	0.6	8	76
MF	adenyl nucleotide binding	GO:0030554	4e-5	3.4	11.5	31	1518
CC	chromosome passenger complex	GO:0032133	6e-5	Inf	0.0	3	3
MF	nucleoside binding	GO:0001882	6e-5	3.3	11.7	31	1544
BP	DNA repair	GO:0006281	1e-4	7.9	1.8	12	230
BP	cellular response to stimulus	GO:0051716	2e-4	4.5	5.3	20	656
CC	midbody	GO:0030496	4e-4	56.0	0.1	4	14
BP	regulation of mitosis	GO:0007088	4e-4	24.2	0.3	6	40
CC	microtubule	GO:0005874	4e-4	7.4	1.5	10	216
BP	cell proliferation	GO:0008283	6e-4	3.8	7.3	23	914
KEGG	p53 signaling pathway	04115	8e-4	12.7	0.6	6	69
CC	cytosol	GO:0005829	8e-4	3.4	7.6	22	1035
CC	chromosome, centromeric region	GO:0000775	1e-3	22.2	0.3	5	41
BP	positive regulation of nuclear division	GO:0051785	1e-3	32.3	0.2	5	25
CC	DNA replication factor C complex	GO:0005663	1e-3	138.7	0.0	3	6
MF	purine ribonucleotide binding	GO:0032555	1e-3	2.9	13.4	31	1772
BP	mitotic sister chromatid segregation	GO:0000070	1e-3	66.6	0.1	4	12
KEGG	Mismatch repair	03430	1e-3	27.0	0.2	4	23
BP	microtubule-based movement	GO:0007018	1e-3	11.0	0.8	8	104
BP	establishment of chromosome localization	GO:0051303	1e-3	388.6	0.0	3	4
BP	nucleosome assembly	GO:0006334	2e-3	13.4	0.6	7	77
CC	outer kinetochore of condensed chromosome	GO:0000940	2e-3	104.0	0.1	3	7
BP	DNA replication initiation	GO:0006270	3e-3	52.3	0.1	4	14
BP	regulation of organelle organization	GO:0033043	3e-3	10.0	0.9	8	117
CC	spindle microtubule	GO:0005876	3e-3	31.1	0.2	4	22
CC	nuclear chromosome	GO:0000228	4e-3	8.8	0.9	7	120
BP	chromatin assembly or disassembly	GO:0006333	5e-3	9.2	1.0	8	124
MF	DNA-dependent ATPase activity	GO:0008094	6e-3	16.3	0.4	5	47
BP	mitotic cell cycle spindle assembly checkpoint	GO:0007094	1e-2	95.3	0.1	3	7
MF	5'-flap endonuclease activity	GO:0017108	1e-2	Inf	0.0	2	2
BP	cytokinesis	GO:0000910	2e-2	17.5	0.3	5	42
CC	nucleolus	GO:0005730	2e-2	3.4	5.0	15	675
MF	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	GO:0016818	2e-2	3.4	5.4	16	712
BP	negative regulation of mitosis	GO:0045839	2e-2	76.3	0.1	3	8
MF	nucleotide binding	GO:0000166	2e-2	2.4	16.3	32	2156
CC	cytoplasm	GO:0005737	2e-2	2.0	50.4	70	6857
CC	spindle pole centrosome	GO:0031616	2e-2	275.0	0.0	2	3

BP	regulation of ubiquitin-protein ligase activity	GO:0051438	2e-2	11.3	0.6	6	75
BP	positive regulation of cell cycle	GO:0045787	2e-2	15.8	0.4	5	46
KEGG	Small cell lung cancer	05222	2e-2	8.0	0.7	5	86
BP	mitotic cell cycle checkpoint	GO:0007093	3e-2	25.6	0.2	4	25
CC	kinesin complex	GO:0005871	3e-2	32.0	0.1	3	16
BP	cellular biopolymer metabolic process	GO:0034960	3e-2	2.2	43.9	65	5477
BP	regulation of catalytic activity	GO:0050790	4e-2	3.4	5.7	17	716
BP	regulation of phosphate metabolic process	GO:0019220	4e-2	4.4	3.1	12	383
BP	cellular macromolecular complex subunit organization	GO:0034621	4e-2	4.3	3.1	12	385
BP	establishment or maintenance of microtubule cytoskeleton polarity	GO:0030951	5e-2	Inf	0.0	2	2
BP	DNA replication-dependent nucleosome assembly	GO:0006335	5e-2	Inf	0.0	2	2
BP	positive regulation of mitotic metaphase/anaphase transition	GO:0045842	5e-2	Inf	0.0	2	2
BP	mitotic chromosome movement towards spindle pole	GO:0007079	5e-2	Inf	0.0	2	2
BP	mitotic spindle elongation	GO:0000022	5e-2	Inf	0.0	2	2
BP	DNA replication, removal of RNA primer	GO:0043137	5e-2	Inf	0.0	2	2
BP	regulation of kinase activity	GO:0043549	5e-2	5.1	2.2	10	272

References

1. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M. *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, **473**, 43-49.
2. Desprat, R., Thierry-Mieg, D., Lailler, N., Lajugie, J., Schildkraut, C., Thierry-Mieg, J. and Bouhassira, E.E. (2009) Predictable dynamic program of timing of DNA replication in human cells. *Genome Res*, **19**, 2288-2299.
3. Woodfine, K., Fiegler, H., Beare, D.M., Collins, J.E., McCann, O.T., Young, B.D., Debernardi, S., Mott, R., Dunham, I. and Carter, N.P. (2004) Replication timing of the human genome. *Hum Mol Genet*, **13**, 191-202.
4. Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gartler, S.M. and Stamatoyannopoulos, J.A. (2010) Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci U S A*, **107**, 139-144.
5. Bar-Joseph, Z., Siegfried, Z., Brandeis, M., Brors, B., Lu, Y., Eils, R., Dynlacht, B.D. and Simon, I. (2008) Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proc Natl Acad Sci U S A*, **105**, 955-960.
6. Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O. *et al.* (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell*, **13**, 1977-2000.
7. Falcon, S. and Gentleman, R. (2007) Using GOstats to test gene lists for GO term association. *Bioinformatics*, **23**, 257-258.

Paper II

Cell cycle regulation of human DNA repair and chromatin remodeling genes

Robin Mjelle[1], Siv Anita Hegre[1], Per Arne Aas[1], Geir Slupphaug[1], Finn Drabløs[1], Pål Sætrom[1,2]*, Hans E. Krokan[1]*

[1]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway.

[2] Department of Computer and Information Science, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway.

*Corresponding authors:

Hans E. Krokan, hans.krokan@ntnu.no

Pål Sætrom, e-mail address: pal.satrom@ntnu.no

ABSTRACT

Maintenance of a genome requires DNA repair integrated with chromatin remodeling. We have analyzed six transcriptome data sets and one data set on translational regulation of known DNA repair and remodeling genes in synchronized human cells. These data are available through our new database: http://bigr.medisin.ntnu.no/data/website_dnarep/. Genes that have similar transcription profiles in at least two of our data sets generally agree well with known protein profiles. In brief, long patch base excision repair (BER) is enriched for S phase genes, whereas short patch BER uses genes essentially equally expressed in all cell cycle phases. Furthermore, most genes related to DNA mismatch repair, Fanconi anemia and homologous recombination have their highest expression in the S phase. In contrast, genes specific for direct repair, nucleotide excision repair, as well as non-homologous end joining do not show cell cycle-related expression. Cell cycle regulated chromatin remodeling genes were most frequently confined to G1/S and S. These include e.g. genes for chromatin assembly factor 1 (CAF-1) major subunits CHAF1A and CHAF1B; the putative helicases HELLS and ATAD2 that both co-activate E2F transcription factors central in G1/S-transition and recruit DNA repair and chromatin-modifying proteins and DNA double strand break repair proteins; and RAD54L and RAD54B involved in double strand break repair. *TOP2A* was consistently most highly expressed in G2, but also expressed in late S phase, supporting a role in regulating entry into mitosis. Translational regulation complements transcriptional regulation and appears to be a relatively common cell cycle regulatory mechanism for DNA repair genes. Our results identify cell cycle phases in which different pathways have highest activity, and demonstrate that periodically expressed genes in a pathway are frequently co-expressed. Furthermore, the data suggest that S phase expression and over-expression of some multifunctional chromatin remodeling proteins may set up feedback loops driving cancer cell proliferation.

Keywords

Cell cycle; DNA repair genes; chromatin remodeling genes

Abbreviations

MMR, mismatch repair; NER, nucleotide excision repair; BER, base excision repair; DSB, DNA double-strand break; NHEJ, non-homologous end joining; HR, homologous recombination; FA, Fanconi anemia; SS, serum starvation; TT, double thymidine block; NZ, nocodazole; MS, mitotic shake off.

Highlights

- Novel database for cell cycle-regulated DNA repair and chromatin remodeling genes
- Analyses of genome wide regulation of DNA repair and chromatin remodeling genes
- Correlated expression of DNA repair genes with similar functions

1. Introduction

Repair of damage to DNA requires complex biological mechanisms that are tightly regulated and integrated. Damage that is not corrected prior to replication may be cytotoxic and mutagenic, making DNA damage responses during the cell cycle of particular interest. Some proteins have DNA repair as their sole or main function. However, a number of DNA repair proteins also have additional functions, e.g. in adaptive immunity [1], transcription [2] and in replication [3]. Progression of cell cycle processes is normally monitored by distinct checkpoints in G1/S, intra-S and G2/M. These checkpoints control the progression through the various phases of the cell cycle. Although the checkpoints are distinct, they all respond to lesions in DNA and share several proteins [4]. Checkpoint activation and recruitment of DNA damage response proteins depend on the type of lesion [5]. Below we have briefly outlined characteristics of the major DNA repair mechanisms.

The most versatile DNA repair mechanism is perhaps base excision repair (BER) that corrects a very large number of small base lesions caused by oxidation, deamination and alkylation. The pathway is initiated by one among 11 known mammalian DNA glycosylases that recognize and remove the damaged base. BER is completed by a short patch or a long patch repair route that in part use different proteins in the downstream steps [6]. Some BER factors are known to be regulated in a cell cycle specific manner. Specifically, expression of uracil-DNA glycosylase encoded by the *UNG* gene peaks in late G1 and S phase both at transcript and protein levels [7-10], whereas thymine/uracil mismatch glycosylase TDG peaks throughout G1 phase and declines in the S phase [9]. However, upregulation of nuclear UNG in G1 phase and strong downregulation in S phase was reported by one group [11]. The reason for this inconsistency is not known.

Mismatch repair (MMR) corrects base-base mismatches and insertion/deletion loops generated during DNA replication and recombination. Since MMR is an immediate post-replicative correction mechanism, a prediction would be that the proteins involved are cell cycle regulated. Some key factors in MMR, e.g. MSH2 and MLH1 proteins were indeed reported to be upregulated when quiescent cells were stimulated to proliferate, but fluctuation through cell cycle phases was less clear [12, 13]. MMR also requires several replication factors, such as DNA polymerases, EXO1, RFC and PCNA [14].

Nucleotide excision repair (NER) is the principal repair mechanism for DNA damage causing helix distortion, most commonly pyrimidine dimers from ultraviolet light, and requires more than 30 gene products, some unique and other shared with DNA replication and transcription [15].

Double strand breaks (DSBs) are highly toxic lesions that may cause various mutations, deletions and oncogenic translocations. They may be caused directly by ionizing radiation, or indirectly by endogenous or exogenous challenges that cause single stranded breaks that, unless repaired, are converted to DSBs upon replication. Repair of DSBs in mammalian cells takes place by non-homologous end joining (NHEJ) or related alternative mechanisms in all cell cycle phases and in addition by homologous recombination (HR) repair during S phase and G2 phase, when a sister chromatid is available [16, 17].

Interstrand crosslinks (ICLs) may be formed by environmental mutagens, as well as a number of chemotherapeutic drugs. ICLs represent a highly toxic type of lesion, the processing of which requires input from several repair mechanisms, including NER, HR, MMR, NHEJ, translesion polymerases and, importantly, a number of Fanconi anemia (FA) proteins. Deficiency in FANC-proteins is associated with high sensitivity to agents that cause ICLs. Furthermore, repair of ICLs may use different mechanisms in different cell cycle phases [18].

Methods for cell cycle synchronization include temporary growth arrest in the G1/G0-phase at high cell density or serum starvation, as well as block of cell cycle progression at the G1/S phase transition or in the G2 phase. Alternatively, enrichment of cells in certain cell cycle phases without chemical treatment may be obtained by mitotic shake off, elutriation centrifugation or by other means. These methods all have advantages and drawbacks, including limited synchrony, cell line-specific differences, intra-culture heterogeneity, toxicity of chemicals changing expression, small yield of cells and requirement for specialized instrumentation. It is advantageous if cultures can be followed through two or more rounds of the cell cycle and the method should ideally fulfill a set of defined criteria [19]. To investigate genome-wide expression of cell cycle-associated genes, microarray analyses of synchronized cells have been carried out using different eukaryotic cell lines [20-25]. Results from cell cycle studies show that functionally distinct classes of genes are expressed at the highest level when they are

needed [25-27]. Whereas the regulation of protein complexes frequently has evolved differently in different species, regulated subunits of proteins or complexes are usually expressed just before their time of action. Furthermore, changes in transcriptional regulation have frequently co-evolved with post-translational control [26].

Here we have used available data sets to examine transcription profiles and translational regulation of all known human DNA repair and chromatin remodeling genes during the cell cycle. These data sets use different synchronization procedures and cell lines, which reduce the risk of method-associated errors. Periodic expression of transcripts does not prove that protein levels fluctuate correspondingly. However, from a number of earlier studies, expression at transcript and protein levels of several DNA repair and remodeling genes are known. Reassuringly, our results are in good agreement with these results, as discussed below. Furthermore, DNA repair and remodeling often use multi-subunit proteins or complexes to carry out the task. Based on available information [26], it may be fair to hypothesize that if at least one of the required and critical subunits in a protein or protein complex is clearly cell cycle regulated, the function of the protein or complex is likely to be cell cycle regulated. Our studies identify a number of DNA repair and chromatin remodeling genes that are cell cycle regulated, most commonly peaking in the G1/S or S phase. Furthermore, the information identifies gene sets that likely contribute to overall regulation of pathways and subpathways.

2. Materials and methods

2.1 DNA repair and chromatin remodeling genes

A list of 177 DNA-repair genes published previously [28, 29] was used and supplemented with additional *bona fide* DNA repair genes identified by PubMed searches, to give a sum of 345 DNA repair genes. The list of established and putative genes for chromatin remodeling proteins contains 99 genes. They were compiled from a comprehensive literature search carried out by members of the FANTOM5 consortium (EpiGenes 1.3; F. Drabløs, Y. Medvedeva, A. Lennartsson, unpublished data) and supplemented with novel information from data searches. The new database named "Cell Cycle Regulation of DNA Repair and Chromatin Remodeling Genes" is found

here: http://bigr.medisin.ntnu.no/data/website_dnarep/. We have consistently used the HUGO Gene Nomenclature Committee (HGNC) approved gene names in the new database, as well as in the paper. The alternative names can be found in OMIM.

2.2 Data sets in the novel database

Seven time series gene expression data sets from five different studies were downloaded and analyzed. These include primary foreskin fibroblasts synchronized in G0-phase by serum starvation [20]; HaCaT cells, a near diploid keratinocyte cell line, synchronized at G1/S-transition by double thymidine block [22]; human cervical carcinoma HeLa cells synchronized in M by nocodazole [22] or in G1/S transition by double thymidine block (this paper, see Supplementary Materials); HeLa cells using double thymidine block and nocodazole [30]; or HeLa cells synchronized by double thymidine block, nocodazole and mitotic shake-off [25]. These studies, except [30], were carried out using gene expression microarrays to analyze multiple time points during one or several cell cycles. The single data set available on translational regulation in HeLa cells by ribosome profiling using RNA sequencing, as well as the accompanying transcriptome, were based on three time points only (G1, S and G2) [30]. These therefore do not discriminate between genes expressed in early/middle G1, and genes expressed at the G1/S transition. Note that as this study specifically measured changes in translational rates relative to transcriptional levels, some genes were identified as translationally regulated in all three phases.

2.3 Analysis of transcript profiles

Cell cycle regulated genes were identified using partial least square (PLS) regression on to a sine and cosine function. Each gene was assigned a cell cycle phase as described [22]. Transcription data from Stumpf et al. [30] were analyzed by using Cufflinks [31] followed by EBSeq in R [32] to compare the three individual time points, one from each cell cycle phase.

3. Results and Discussion

3.1 Organization of the new database "Cell Cycle Regulation of DNA Repair and Chromatin Remodeling Genes"

The screen shot shows key features of the novel database on cell cycle regulated DNA repair genes and chromatin remodeling genes (Fig. 1). The horizontal menu line displayed in the figure can be used to maneuver to any part of the database. The ARTICLE will contain the published form of the present paper. The GENE LIST lists all identified DNA repair genes and chromatin remodeling genes separately, both cell cycle regulated and not cell cycle regulated ones. By clicking on the name of individual genes, information in OMIM is available. The banner CELL CYCLE lists the cell cycle regulated DNA repair and chromatin remodeling genes, as separate lists, with reference to source of the data sets. Each gene has been assigned to the cell cycle phase in which it has the highest expression in each of the studies included. A transcriptome expression profile for each gene, based on data from our laboratory [22], foreskin fibroblasts [20], and this paper (listed as Mjelle et al., 2014 in the database; see also Supplementary Materials and Supplementary Fig. 1), can be reached by clicking on the gene name. Expression data from the other studies [25, 30] can be reached by clicking METHODS in the menu line and then the relevant study.

3.2 Results of the analyses - overview

To reduce false positives, we limited detailed transcriptome analyses [20, 22, 25, 30] to genes that were cell cycle regulated in at least two data sets. For analyses of translational regulation, the single data set available was used [30]. In sum, when analyzing the transcriptome data sets for which expression profiles is displayed in the novel database, we observe an enrichment of DNA repair genes among cell cycle regulated genes (17% compared to 8% for all genes, p=5.7e-9), whereas the number of cell cycle regulated chromatin remodeling genes is not higher than expected (11% compared to 8% for all genes, p=0.27). More detailed information, including expression profiles, is found in the novel database: http://bigr.medisin.ntnu.no/data/website_dnarep/. This page gives

information for different cell lines, cell cycle phase of highest expression (with links to expression profiles), chromosomal localization of the genes and known and presumed gene functions (links to OMIM).

It should be understood that although many genes may be upregulated several-fold in distinct cell cycle phases, a low level of expression is usually observed in other cell cycle phases. Normal primary fibroblasts have the highest number of cell cycle regulated genes (80 genes) suggesting that cell cycle regulation of several genes involved in DNA repair may be deranged in transformed cells. This could not be explained by reduced synchrony in HaCaT and HeLa cells (Supplementary Figure 1). In total, we identified 124 DNA repair genes to be cell cycle regulated in at least one data set, whereas 58 were cell cycle regulated in two or more data sets. Genes with conserved expression profiles across experiments are most frequently expressed in late G1 phase and S phase, although some have distinct expression in other cell cycle phases, e.g. TOP2A and ASF1A which are G2 phase regulated.

Although expression profiles differ significantly between cell lines, there is a reasonable overlap between the studies using transcription profiling as tool. Furthermore, the relationship between expression at the transcription level and protein level is known for some DNA repair and remodeling genes. It is reassuring to notice that transcript levels generally agree very well with protein data (when known). Some examples include G1/S genes *UNG* [10], *PCNA* [33], *POLD* and *POLA* [34], *BLM* [35], *BRCA1*[36], *FEN1* [37], *HELLS* [38], *CHAF1A* and *CHAF1B* [39], but also the G2 expressed gene *TOP2A* [40, 41]. In some cases the expression is upregulated both transcriptionally and translationally in the same cell cycle phase, e.g. MSH6, ATAD2 and POLA1, which are upregulated in G1/S or S phase. There are also several examples of genes that are either transcriptionally or translationally regulated. More detailed analyses are found below.

3.3 Translational regulation in the cell cycle

In total, 30 DNA repair genes and 11 chromatin remodeling genes were found to be translationally regulated in the cell cycle. They were mostly upregulated, although three were down-regulated, either in the G1 phase (BTG2 and Tp73) or S phase (PAPD7). Thus, translational regulation is relatively common and predominantly acts as a

mechanism for upregulation of genes in a cell cycle phase. Among the DNA repair genes, 13 were only translationally regulated (e.g. KAT5, SMC3, ERCC4 and GTF2H2). KAT5 (a histone Lys acetyltransferase, initially called TIP60), is translationally upregulated in G1 phase and contributes to DSBR. KAT5 acetylates ATM via chromatin binding and activates the ATM-pathway [42]. SMC3 (structural maintenance of chromosomes 3) is translationally upregulated in G1 and S phase. It is part of the cohesion complex that keeps sister chromatids together and it is important for DSBR by homologous recombination [43]. ERCC4 (also called XPF) is translationally upregulated in all three phases examined (G1, S, and M) and, strictly speaking, not cell cycle regulated. ERCC4 is best known for its function as a structure-specific endonuclease in complex with ERCC1 in NER, but is also involved in ICL repair and DSBR [44]. GTF2H2, a subunit of TFIIH, is translationally upregulated in G1 phase and is required both for NER and transcription.

3.4 Genes encoding base excision repair (BER) and single strand break repair (SSBR) proteins

The damage recognizing proteins in BER are DNA glycosylases that remove damaged bases that do not cause major distortion to the DNA helix structure. SSBR is generally using proteins involved in the downstream steps of BER, although this may be a simplistic view. The 11 human DNA glycosylases identified recognize both spontaneous and induced lesions. Four of the genes encoding DNA glycosylases were found to be cell cycle regulated; these are *UNG, TDG, NTHL1* and *NEIL3*. *UNG* and *NTHL1* are expressed at G1/S, NEIL3 in S/G2 and TDG in G1. In addition, UNG is translationally upregulated in G1. We found no evidence for cell cycle regulation of *MPG*, *SMUG1*, *MBD4*, *NEIL1*, *OGG1* and *MUTYH* (Fig. 2). However, evidence for S phase expression was previously reported for *NEIL1* when using prolonged serum starvation to synchronize fibroblasts [45]. The opposite expression pattern of UNG and TDG has been reported previously, both at the mRNA and protein levels [9] and our results are consistent with previous findings [9, 46, 47]. The G1/S expression of *UNG* is consistent with its defined role in immediate post-replicative removal of misincorporated dUMP [10, 48], as well as its probable role in pre-replicative correction of U:G mismatches [49]. Nuclear UNG2 (encoded by *UNG*) and NEIL1 (and NEIL3), unlike other DNA

glycosylases, are highly active on single-strand and double-strand DNA and are located in replication foci during S phase [50]. Possibly, UNG2 and NEIL1 may remove damaged bases in single-stranded DNA at the replication fork, followed by fork regression and BER by the usual mechanism for double-stranded DNA, as suggested [49, 50]. *NEIL3* was recently reported to be induced in late S phase in cells synchronized by density inhibition, but the data reveal that *NEIL3* expression continues into G2 [51], thus these data are consistent with our results. Several genes involved in downstream steps of short-patch are not cell cycle regulated, including *POLB*, encoding DNA polymerase β, *POLL*, encoding DNA polymerase λ, *APEX1*, encoding AP-endonuclease 1 and *PNKP*, encoding polynucleotide kinase phosphatase. Interestingly, PARP1 and PARP2 are S phase regulated in normal fibroblasts, consistent with a role of PARP in SSB and at least some forms of BER, but apparently not all [52]. The genes for proteins in the downstream steps of long-patch BER (*LIG1*, *PCNA*, *FEN1*, *POLD1*, *POLD3*, *POLE* and *POLE2*), are all consistently S phase or G1/S regulated. All proteins from these genes are also replication proteins. LIG1 is involved in short patch repair as well, whereas LIG3 is probably involved in nuclear short patch BER and essential for mitochondrial BER (reviewed in [6]). These results support a model where long patch BER is mainly acting in proliferating cells and that repair by long patch BER is carried out by genes induced in late G1 and early S phase, in agreement with previous results on *in vitro* BER in cell extracts [53]. The S phase expressed *POLA1* encodes DNA polymerase α/primase and has relatively low fidelity due to lack of 3' → 5' proofreading activity. It is not thought to be involved in BER but appears to be involved in repair of SSBs during replication through binding of XRCC1, presumably near the replication fork [54].

3.5 Genes encoding proteins involved in direct repair of base lesions

Generally, genes encoding proteins involved in direct repair of base alkylations are not cell cycle regulated (Fig. 3). This applies to *MGMT* required for repair of the highly mutagenic DNA lesion O^6-methylguanine, as well as human AlkB homologues *ALKBH1, ALKBH2* and *ALKBH3*. *ALKBH2* was cell cycle regulated to G1 phase in one HeLa cell strain and S in another, but not cell cycle regulated in fibroblasts and HaCaT cells. Since alkylation lesions may occur at all cell cycle stages from normal metabolites

(e.g. S-adenosylmethionine) as well as external challenges, the general absence of cell cycle regulation of these genes would be an expected outcome.

3.6 Genes encoding nucleotide excision repair (NER) proteins

NER is required for excision repair of bulky DNA adducts, e.g. from ultraviolet light. Genes for repair proteins involved in the initial steps in NER by both global genome repair (GGR) and transcription coupled repair (TCR) appear to be expressed independently of cell cycle phases, although most genes in the downstream steps are upregulated in S phase (Fig. 4). Thus, CSA (*ERCC8*) and CSB (*ERCC6*) required for TCR, are not detected as cell cycle regulated at the transcript level, neither are the XPC-RAD23B-complex in GGR or any of the other XP-genes, including *XPA*, *XPB* (*ERCC3*), *XPD* (*ERCC2*), *XPE* (*DDB1*), *XPF* (*ERCC4*) or *XPG* (*ERCC5*). However, ERCC4 is translationally upregulated in G1, S and M. Furthermore, the final steps of the pathway, which are in part shared with DNA replication and other excision repair pathways, use cell cycle regulated genes that are most frequently uregulated during G1/S phase, including *RPA1*, *PCNA*, *RFC4* and *RFC5* (subunits of RFC), *LIG1*, *POLD1* and *POLD3* (subunits of POLδ) and *POLE* and *POLE2* (encoding POLε). In addition, *POLK* (encoding POLκ) is translationally upregulated in the S phase. For gap filling, ubiquitinated PCNA recruits POLκ to the site of DNA damage, where it is found in complex with POLδ. These polymerases are responsible for synthesis of approximately half of the repair gap and POLε the rest [55]. Furthermore, UV-induced nuclear import of XPA primarily takes place in the S phase and appears to initiate NER in a p53-dependent manner [56]. The NER endonuclease XPG and the closely related yeast homolog Rad2 have a PCNA-binding domain in the C-terminal region. The Rad2-PCNA interaction was recently found to mediate arrest of cell cycle progression, suggesting a role of XPG/Rad2 other than the established endonuclease function [57].

3.7 Genes encoding mismatch repair (MMR) proteins

MMR proteins are primarily required for correction of mismatches from replication errors. Some studies have reported cell cycle dependent expression of MMR genes while others have noted just moderate fluctuations throughout the cell cycle [12, 58, 59]. We observe that the genes encoding major components in mismatch recognition are

transcriptionally up-regulated in S or G1/S phase in normal fibroblasts, including *MLH1, MSH2* and *MSH6. MSH6* is upregulated in S phase in all cell lines, whereas the other components are less consistently upregulated. In addition, *MSH2* and *MSH6* are translationally upregulated in G1/S or S phase (Fig. 5). *MSH3* is apparently not cell cycle regulated. Generally, MSH heterodimers recognize mismatches, whereas MLH heterodimers are strand-specific endonucleases. The MSH2-MSH6 complex (MutSα) recognizes single nucleotide mismatches, whereas the MSH2-MSH3 complex (MutSβ) recognizes insertion/deletion loops. MutL homologues PMS1 and PMS2 form heterodimers with MLH1. Among these, PMS2 is translationally upregulated in G1. MSH4 and MSH5, the expression of which is normally largely limited to ovary and testis, are not cell cycle regulated in the cell lines examined. The MSH4-MSH5 complex recognizes Holliday junctions and forms a sliding clamp during meiosis, but has no apparent role in replication-associated MMR, thus there is no specific requirement for these components during S phase [60]. The gene encoding the 5'→3' excision nuclease EXO1, required for damaged strand excision in MMR is upregulated in G1/S or S phase. EXO1 has multiple functions in DNA maintenance, including MMR, replication, recombination and telomere maintenance, processes that are all associated with S phase in mammalian cells. As expected, other replication-related genes active in re-synthesis and ligation are up-regulated during S phase, including *RFC3*, *RFC*4, *RFC5*, *PCNA*, *RPA1*, *POLD1*, *POLD3* and *LIG1*. Thus, expression patterns of genes for MMR proteins collectively comply with their roles in MMR, HRR and replication.

3.8 Genes encoding Fanconi anemia (FA) pathway proteins

FA proteins have a major role in repair of inter-strand cross links (ICLs). FA is associated with increased cancer risk and cellular sensitivity to inter-strand crosslinking agents and ionizing radiation [61]. The majority of FA genes are upregulated in G1/S at the transcript level in at least two data sets (11 genes) or translationally upregulated in the S phase (3 genes). Repair processes requiring FA proteins are not limited to ICLs, although they apparently have a critical role there. The FA pathway is damage-inducible by replication fork-stalling lesions that activate ATR kinase, which in turn activates some proteins in the FA core complex. The FA pathway is complex and involves 13 FANC genes (FANC A, B, C, D1, D2, E, F, G, I, J, L, M and N) and other genes as well,

including the S phase expressed genes *BLM* (helicase), *USP1* (deubiquitinase), *BRCA1*, *RPA1, RAD1* and *RAD51*. Genes for most components in the FA core complex (*FANC A, B, E* and *L*, but not *C* and *F*) are up-regulated during S phase (Fig. 6). Although *FANCM* is not S phase expressed at transcript level, it is translationally upregulated in the S phase. Upon activation, the FA core complex ubiquitinates the I–D2 complex consisting of FANCI and FANCD2, which are apparently transcriptionally up-regulated during late S- and G2- phase. The I-D2 complex is required to promote repair by homologous recombination, which in mammalian cells takes place in the S- and G2 phase. The I–D2 complex in turn activates the downstream components BRCA2 (FANCD1), RAD51 and RAD51C, among which the critical damaged DNA-binding protein RAD51 is upregulated during G1/S- and S phase. FAN1 (Fanconi anemia associated nuclease 1) is required for ICL repair, but is not cell cycle regulated [62]. However, the FA complex may also be activating the MMR pathway through the interaction with FAN1, MLH1 and PMS2 [62], among which MLH1 is upregulated in S phase, and PMS2 translationally upregulated in G1. In conclusion most of the FA proteins and some associated proteins are S phase regulated, in agreement with the view that repair of ICLs is most efficient in the S phase.

3.9 Genes for homologous recombination repair (HRR)

In mammalian cells, HRR is restricted to the S phase and G2 phase, using the sister chromatid rather than the homologous chromosome for strand exchange [17]. Consistent with this, many of the genes involved in HRR are most highly expressed in S phase. Homologous recombination is initiated by the MRN complex consisting of MRE11A, RAD50 and NBN (NBS1). Among these, *MRE11A* is found to be induced in late S- and G2 phase whereas *NBN* is induced in G1 and G1/S (Fig. 7). Furthermore, PARP1 apparently has a role in facilitating HRR at DSBs resulting from collapsed replication forks by binding to DNA ends and inhibiting competitive binding of Ku70/80 and thereby preventing less accurate repair by NHEJ [63, 64]. We do not find PARP1 convincingly cell cycle regulated, although it is upregulated in the S phase when fibroblasts are released from serum starvation. One function of PARP1 in HRR may be recruitment of the heterodimer of BRCA1 and BARD1, each of which S phase-regulated. This recruitment is parylation-dependent [65]. In addition, SMC3 is

translationally upregulated in G1/S, but not transcriptionally regulated. SMC3 is part of the cohesion complex that keeps sister chromatids together and is important for DSBR by HR [43]. Homology search and DNA-strand invasion is performed by RAD51, which is upregulated during G1/S and S phase across three cell lines. The gene encoding RAD54, a RAD51-interacting protein required for stabilizing RAD51, is also up-regulated during S phase. RAD51-mediated strand exchange was found to be stimulated by the PSMC3IP-MND1 complex. PSMC3IP (also called HOP-2) is expressed in G1/S and was recently reported to have a regulatory role both in DSBR and recombination [66]. A recently described RAD51 interaction partner, FIGNL1 [67], interacts specifically with RAD51 and is upregulated in the S phase, similar to RAD51. Of the genes in the post-synapsis step, *BLM* [68] is found to be up-regulated in S phase in all four cell lines and *BRCA2* is possibly upregulated during G2 phase (in fibroblasts only). In conclusion, many of the genes required for HRR are upregulated in S phase, consistent with the activity of this process in the S phase.

3.10 Genes encoding proteins for non-homologous end joining (NHEJ) - a regulatory role for PTTG/securin

Generally, genes centrally involved in NHEJ were found not to be cell cycle regulated (Fig. 8). However, *PTTG1*, encoding the 202 amino acids protein PTTG1 (also called securin), consistently had the highest transcript level at the M/G1-transition in all data sets and then declines rapidly during the G1 phase. PTTG1 is a multifunctional and rather intriguing protein. It binds directly to Ku70 (also called XRCC6) and functions as a negative regulator of NHEJ [69, 70]. Thus, PTTG1 may inhibit illegitimate NHEJ until completion of replication. It was reported that PTTG1 is ubiquitinated at metaphase-anaphase transition by the E3 ligase anaphase-promoting complex (APC). It is then normally rapidly degraded in proteasomes (reviewed in [71]), thus alleviating the NHEJ-inhibition and reestablishing the capacity for NHEJ during G1 phase. Importantly, PTTG1 also controls sister chromatid separation during mitosis by binding to and inhibiting separase, which when active cleaves cohesin that holds sister chromatids together [72]. Normal levels of PTTG1 may prevent premature sister chromatid separation, thus facilitating DNA repair by HRR during the S phase and G2. However, overexpression of PTTG1 correlates strongly with aneuploidy in breast cancer [73],

possibly due to an inability of sister chromatids to separate. Overexpression of PTTG1 is associated with a wide range of malignancies, *e.g.* hematopoietic malignancies [74], colon cancer [75] and breast cancer [76]. Thus, while PTTG1 normally regulates several processes important to genomic stability, overexpression appears to promote genomic instability, stimulate proliferation and contribute to carcinogenesis. PTTG1 also functions as a transcriptional regulator of several genes involved in tumorigenesis [77]. We conclude that normally the transcript level of PTTG1 is upregulated at the M/G1-transition, but subsequently ubiquitinylated and degraded in proteasomes at metaphase-anaphase transition, allowing chromatid separation. Overexpression of PTTG1 apparently results in inability of sister chromatids to separate with increased risk of aneuploidy in cancer cells.

3.11 Chromatin assembly and remodeling in DNA repair

Chromatin remodeling complexes temporarily disrupt and remodel DNA-nucleosome interactions, thus facilitating different DNA transactions, including DNA replication, transcription and repair [78, 79]. We examined 99 genes thought to be involved in chromatin remodeling, many of which are already known to have a function in DNA repair. Among these, 39 were found to be cell cycle regulated in at least one transcriptome data set and 11 in at least two data sets. While three remodeling genes were regulated both at the transcriptional and translational level, several were only translationally regulated (CHD1, ACTL6A, ZRANB3, TP73, SMARCA1, SMARCA5 and SMARCAD1). The cell cycle regulated chromatin remodeling genes were largely G1/S-expressed genes, except TOP2A, which displayed highest expression in G2 phase.

The CAF-1 (chromatin assembly factor 1) complex consists of CHAF1A, CHAF1B and p50 and is generally required to assemble nucleosomes after DNA synthesis. We find that *CHAF1A* and *CHAF1B* are consistently most highly expressed in G1/S. CAF-1 has a role in MMR, which itself has the major normal function in the S phase. In MMR the long repair patch is generated by the S phase-regulated 5'-3' nuclease EXO1 [14]. After DNA re-synthesis in MMR, the CAF-1 complex is required to re-assemble nucleosomes, protecting DNA from excessive degradation [80, 81]. CAF-1 is also required to restore chromatin after NER [82] and HRR [7], and is thus heavily involved in DNA repair processes.

Genes for the putative ATPases/helicases ATAD2 (also called ANCCA) and HELLS (also called LSH, PASG and SMARCA6) are also consistently most highly expressed in G1/S. In addition, ATAD2 is translationally upregulated in the S phase. HELLS has a documented function in DSBR [83], and ATAD2 is likely to be involved in DNA repair. HELLS and ATAD2 are overexpressed in several cancer types and are considered as proliferation drivers. HELLS and ATAD2 are significantly coexpressed, with a Pearson correlation coefficient of 0.63 when analyzed using the database COEXPRESdb [84]. Through various interactions and diverse functions they have the potential to set up feedback loops that enhance proliferation (Fig. 9). However, whereas HELLS is overexpressed and may contribute to tumor progression in prostate cancer [85], ATAD2 is apparently not significantly expressed in prostate cancer, but overexpressed in several other common cancers [86], similar to HELLS. ATAD2 and HELLS are rather intriguing proteins that appear to have several different roles in DNA transactions. They both function as co-activators for E2F transcription factors that stimulate G1 to S phase transitions, as outlined below. ATAD2 is an AAA-protein type ATPase (ATPase Associated with diverse cellular Activities) that is thought to be involved in chromatin remodeling and regulation of transcription [87]. Recruitment of ATAD2 to chromatin requires its bromodomain that binds to histone H3K14ac [88] and histone H4K5ac [89]. Importantly, H3K14ac was recently shown to facilitate nucleotide excision repair in yeast through stabilizing binding of a chromatin remodeling complex [90]. As mentioned, ATAD2 is overexpressed in a number of human cancers, but although ATAD2 is induced by estrogen and acts as a co-regulator for estrogen and androgen receptors, overexpression is not limited to hormone-dependent tumors [86, 89, 91-94]. Furthermore, ATAD2 overexpression appears to act as a driver of proliferation in cancer cells and is associated with poor prognosis [92, 93]. Several observed functions of ATAD2 may be related to its direct interaction and function as co-activator for transcription factors that are involved in G1/S-transitions and proliferation, including E2F1, cyclin D_1, MYC, B-MYB, histone methyltransferase EZH2 and others [88, 92-95]. It is reasonable to assume that the G1/S-regulation of ATAD2 itself contributes to the specific S phase functions of the gene products mentioned.

The G1/S regulated *HELLS* gene encodes an SNF2-related ATPase/helicase and was identified as a putative helicase with highest expression in T-cells in fetal mouse

thymus [96], but is generally expressed in proliferating cells. Transcription factor FOXM1, reported to be G1/S regulated [97], is a central regulator of HELLS expression [98]. However, our laboratory (data not shown), as well as others, observed highest expression of FOXM1 in G2/M/G1 [20, 22, 25], consistent with an additional important function of FOXM1 in regulation of several G2-specific genes as well [99, 100]. HELLS promotes phosphorylation of H2AX to γH2AX and contributes to efficient repair of double-strand breaks in mouse cells [83]. HELLS contributes to *de novo* methylation of DNA by interaction with DNA methyltransferases and is involved in regulation of transcription [101-103]. It also has a role in methylation of histone H3K4 [104]. Furthermore, HELLS silences expression of the CDK4-inhibitor $p16^{INK4}$ by recruiting HDAC1 and possibly HDAC2 to the $p16^{INK4}$ gene promoter, thereby increasing cell proliferation and delaying senescence [105]. The apparent role of HELLS in progression of prostate cancer may in part be explained by its function as a co-activator for transcription factor E2F3 [85]. In contrast, HELLS deficiency causes erythroleukemia in mice [102], and deletions are associated with acute myeloblastic- and acute myelogenic leukemia in humans [106]. In sum, the S phase regulated HELLS contributes to DSBR by promoting phosphorylation of H2AX to γH2AX and has complex epigenetic functions at the DNA and histone level. In addition, it acts as co-activator for transcription factor E2F3 that increases expression of several growth promoting genes. Depending on cell type it may both prevent and promote carcinogenesis.

Genes for RAD54B and RAD54L are also most highly expressed in G1/S or S phase. RAD54L (usually called RAD54) is a much-studied member of the SNF2 family of helicases, but may not be a functional helicase [107]. RAD54L interacts with RAD51, also a G1/S DNA repair and remodeling protein, to carry out the key reactions of homology search and DNA strand invasion [107]. In this process RAD51-ssDNA stimulates RAD54L-dependent chromatin remodeling in a homology-dependent manner [108]. RAD54B is another member of the SNF2 superfamily of helicases and is involved in DNA repair by homologous recombination. In the nucleoplasm, RAD54B co-localizes with RAD51 and BRCA1, both of which are also most highly expressed in G1/S. RAD54B is not an essential DNA repair protein, but it is synthetically lethal with loss of LIG4, required for NHEJ [109]. The mechanistic function of RAD54B in recombination is not clear.

Topoisomerase II encoded by *TOP2A* consistently has highest expression in the G2 phase. This is in accordance with earlier studies on TOP2A protein levels demonstrating highest levels during G2/M, although it is also expressed during the S phase [40]. Interestingly, subunits in the SWI/SNF complex, a much-studied remodeling complex, is required for binding of TOP2A to 12,000 genomic binding sites, indicating that TOP2A depends on the SWI/SNF complex to prevent entanglement of DNA at mitosis [110]. A major regulator of the *TOP2A* promoter is the transcription factor FOXM1, which itself peaks in S- and G2 phases of the cell cycle and binds to the *TOP2A* promoter and other S/G2-regulated genes [111].

3.12 DNA Polymerases

We analyzed 16 DNA polymerase genes (including subunits) in the list of DNA repair genes. Seven of these are cell cycle regulated. DNA polα1, polδ1, polδ3, polε and polε2 are all expressed during G1/S- or S phase. DNA pol α, δ and ε are members of the family B polymerases, and the S phase dependent expression confirms their role in DNA replication. PolΘ and polσ are both most highly expressed in G2 phase. PolΘ is known to function in translesion synthesis in the FA pathway; it synthesizes DNA with low fidelity and is proposed to have a role in somatic hypermutation (SHM) of immunoglobulin genes [112]. DNA pol β is found to be independent of cell cycle phase, as previously reported [113]. This is consistent with its dual roles in short patch BER pathway [6] and its role in processing of DSBs during meiosis [114].

4. Conclusions and some concluding remarks

Whereas the core mechanistic steps in different DNA repair pathways are now reasonably well understood in principle, the regulation of DNA repair processes remains elusive. Expression of DNA repair genes is regulated at several levels, including post-translational modification [3, 115, 116]. Some aspects of cell cycle regulation of DNA repair genes was reviewed some years ago and our analysis are in general agreement with this paper [117]. Regulation of DNA repair and chromatin remodeling genes at the cell cycle level is the topic of the present paper. Reassuringly, the genome-wide cell cycle transcriptome studies largely confirm studies at the protein level, when

information is available. As evident from our results and those of others, one DNA glycosylase is S phase regulated (e.g. *UNG*), another one (*TDG*) is G1-regulated, and two (SMUG1 and MBD4) are not cell cycle regulated, indicating non-redundant functions of these uracil-DNA glycosylases. Uracil-DNA glycosylase SMUG1 may serve as a general backup in genomic uracil-removal, particularly outside of S phase when UNG2 levels are lower. In addition it has an apparently unique role in removal of 5-hydroxymethyluracil from DNA. DNA glycosylases removing oxidized or alkylated bases either had insignificant cell cycle transcriptional profiles or were inconsistent between studies (NEIL3, which was detected as upregulated in late S phase, G2 and G1). However, NEIL1 was previously found to be upregulated and present in replication foci in S phase, indicating a possible role in pre-replicative repair [50].

We find that most of the FA genes are up-regulated during S phase, indicating that interstrand cross-links (ICLs) are detected and mainly repaired in S phase. The FA pathway may coordinate the activity of HR, MMR and NER proteins in ICL repair [118, 119]. Several FA proteins are known to bind to DNA containing ICLs either directly or via nonerythroid α spectrin (also called SPTAN1) [120]. This protein also associates with ICLs in telomeres in the S phase and is required for their maintenance [121]. Some repair of ICLs also takes place outside of the S phase and in non-proliferating cells. This replication-independent repair pathway is thought to be initiated by RNA polymerase stalling; repair is slow and independent of the FA proteins, but may use some NER proteins and MMR proteins [18, 122]. Proteins thought to be involved in this FA-independent ICL repair are mostly not cell cycle regulated.

DNA repair by HRR is most active during S/G2 phase when sister chromatids are available. In accordance with this, we found the highest expression of most HRR genes in the S phase. However, HMG20B (also called BRAF35) expression peaks in M/G1. HMG20B is a DNA-binding protein that interacts directly with BRCA2, which is required for HRR. BRCA2, in interaction with RAD51, is required for filament formation and strand invasion in HRR of double strand breaks [123]. Interestingly, the cell cycle pattern of HMG20B expression fits well with the observation that HMG20B and BRCA2 co-localized on mitotic chromosomes and that injection of antibodies to HMG20B caused G2-delay in the cell cycle [124]. These, and several observations not

discussed here, indicate that HMG20B is a multi-functional protein involved in DSB repair and cell cycle regulation at the G2/M transition.

We find that several genes that are critical for HRR (e.g. *RAD51, RAD54L, BRCA1* and *BLM*) are expressed during G1/S transition and S phase, but are probably mostly used later in S phase when sister chromatids increasingly are available and HRR is at its maximum [125]. Repair by NHEJ is initiated by XRCC5 and XRCC6 which are not cell cycle regulated, consistent with NHEJ being active throughout the whole cell cycle. In the MRN complex, only MRE11 is cell cycle regulated, having induced expression during late S and G2 phase. Some reports show that the MRN complex is activated by protein phosphorylation which could explain why these genes are expressed independently of the cell cycle [126-128].

Chromatin remodeling proteins have important roles in genome maintenance processes, including DNA repair and DNA replication (reviewed in [129]) and at least for some types of repair recruitment of chromatin regulators is cell cycle regulated [130]. Most likely, many mechanisms of DNA repair, DNA replication and chromatin remodeling must be integrated and co-regulated. We find that transcripts from 8 chromatin remodeling genes are most highly expressed in S or G1/S in at least two data sets. In addition, 8 chromatin remodeling genes are translationally upregulated in S or G1/S. Many of these are known, or suspected, to have roles in DSBR by HRR or in restoration of nucleosome structure after HRR. Several of these apparently have roles in restoration of the nucleosome structure after MMR and NER. A requirement for chromatin remodeling has also been documented in BER, and appears to depend on the type of nucleosomes present in proximity of the repair patch and steps in the repair process [131, 132]. In sum, DNA repair have cell cycle regulated expression more frequently than average genes, particularly in S phase. Chromatin remodeling genes, however, were not cell cycle regulated at a significantly higher level than expected.

Fig.1. Overview of the database's key features. (A) The database's main page (HOME) has a menu that provides links to background information about the database (ARTICLE, METHODS, RESEARCH GROUP) and links to the database's main sections, which are the lists of chromatin remodeling and DNA repair genes (GENE LISTS) and their cell cycle expression profiles (CELL CYCLE). (B) The "GENE LISTS" pages provide links to information about each gene's function and genomic context. Chromatin remodeling and DNA repair genes have separate pages; shown here is an excerpt of the DNA repair genes. Note that some multifunctional genes, such as HELLS or TOP2A, are present on both. (C) The "CELL CYCLE" pages show chromatin remodeling and DNA repair genes with significant cell cycle dependent regulation in at least one of the seven datasets examined. Abbreviated cell cycle phases (G1/S, S, G2, G2/M, and M/G1) show the phase in which a gene has its peak expression. The abbreviations NR and NTR indicate datasets in which the gene was not detected to be transcriptionally or translationally regulated, respectively. Chromatin remodeling and DNA repair genes have separate pages; shown here is an excerpt of the DNA repair genes, which are grouped by DNA repair pathway. Each gene links to a figure (D) that shows the gene's expression profile and the cell cycle phase distribution for the first four experiments listed in the table. (D) The expression profiles of FEN1.

Fig. 2. Cell cycle regulated genes in the BER pathway. The genes are marked with colors according to which cell cycle phase they have highest expression, as indicated in the figure. Only gene products shown to be cell cycle regulated in at least two transcriptome studies, or translationally regulated [30] are colored. Note that gene products have not been assigned to complexes, merely stage at which they are thought to be involved. Grey fill in ovals indicate that the genes are not found to be cell cycle regulated (not all shown).

Fig. 3. Genes for direct base repair are not cell cycle regulated.

Fig. 4. Cell cycle regulated genes in the NER pathway; see Fig. 2 for an explanation of the color codes. Genes initiating NER

Fig. 5. Cell cycle regulated genes in the MMR pathway; see Fig. 2 for an explanation of the color codes.

Fig. 6. Cell cycle regulated genes in the FA pathway; see Fig. 2 for an explanation of the color codes.

Fig. 7. Cell cycle regulated genes in the HRR pathway; see Fig. 2 for an explanation of the color codes.

Fig. 8. Cell cycle regulated genes in the NHEJ pathway; see Fig. 2 for an explanation of the color codes.

Fig. 9. Model for functions of HELLS – a multifunctional putative helicase and proliferation driver. (A) HELLS, has functions in development (epigenetics), as E2F co-activator in transcription to stimulate cell growth, as well as in repair of DSBs through increasing phosphorylation of H2AX to generate γH2AX. (B) Overexpressed HELLS, ATAD2 and E2Fs may set up functional loops that make them drivers of tumor cell proliferation. They are all S phase expressed and frequently overexpressed in tumors. HELLS and ATAD2 are known to interact physically with E2Fs and function as co-activators of E2F-mediated transcription for a number of genes (but not necessarily *all* genes that use E2F as transcription factor). HELLS has been shown to mediate epigenetic silencing of $P16^{INK4}$ expression, thereby causing activation of CDK4, phosphorylation of RB and release of active E2F1. In sum, HELLS, ATAD2 and E2Fs mediate expression of several genes that stimulate G1/S-transition and cell proliferation.

Supplementary Fig 1. Diagram showing cell synchronization from double thymidine synchronized HeLa cells.

Conflicts of interest

There are no conflicts of interest.

Acknowledgements

This work was supported by The Norwegian Cancer Association (Grant id.: 576160 HEK), the Research Council of Norway; grant id's: 191408 (HEK) and 205316 (PS), the Svanhild and Arne Must Fund for Medical Research, the Liaison Committee between the Central Norway Regional Health Authority and the Norwegian University of Science and Technology (G.S.) and Norwegian University of Science and Technology. RM contributed to design of the study, construction of database, analyses of data and writing the paper; SAH and PAA contributed to design and execution of laboratory work and writing the paper; GS contributed to analyses of data and writing of the paper; FD contributed to generating the list of chromatin remodeling genes and writing the paper. HEK and PS designed the study, contributed to analyzing data and preparing database, and wrote the paper.

References

[1] J.A. Daniel, A. Nussenzweig, The AID-induced DNA damage response in chromatin, Mol Cell, 50 (2013) 309-321.

[2] Y.W. Fong, C. Cattoglio, R. Tjian, The intertwined roles of transcription and repair proteins, Mol Cell, 52 (2013) 291-302.

[3] A. Sancar, L.A. Lindsey-Boltz, K. Unsal-Kacmaz, S. Linn, Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints, Annu Rev Biochem, 73 (2004) 39-85.

[4] J.H. Houtgraaf, J. Versmissen, W.J. van der Giessen, A concise review of DNA damage checkpoints and repair in mammalian cells, Cardiovasc Revasc Med, 7 (2006) 165-172.

[5] K.A. Cimprich, D. Cortez, ATR: an essential regulator of genome integrity, Nat Rev Mol Cell Biol, 9 (2008) 616-627.

[6] H.E. Krokan, M. Bjoras, Base excision repair, Cold Spring Harb Perspect Biol, 5 (2013) a012583.

[7] C. Baldeyron, G. Soria, D. Roche, A.J. Cook, G. Almouzni, HP1alpha recruitment to DNA damage by p150CAF-1 promotes homologous recombination repair, J Cell Biol, 193 (2011) 81-95.

[8] L. Hagen, B. Kavli, M.M. Sousa, K. Torseth, N.B. Liabakk, O. Sundheim, J. Pena-Diaz, M. Otterlei, O. Horning, O.N. Jensen, H.E. Krokan, G. Slupphaug, Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA, EMBO J, 27 (2008) 51-61.

[9] U. Hardeland, C. Kunz, F. Focke, M. Szadkowski, P. Schar, Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2, Nucleic acids research, 35 (2007) 3859-3867.

[10] M. Otterlei, E. Warbrick, T.A. Nagelhus, T. Haug, G. Slupphaug, M. Akbari, P.A. Aas, K. Steinsbekk, O. Bakke, H.E. Krokan, Post-replicative base excision repair in replication foci, EMBO J, 18 (1999) 3834-3844.

[11] J.A. Fischer, S. Caradonna, Analysis of nuclear uracil-DNA glycosylase (nUDG) turnover during the cell cycle, Methods Mol Biol, 761 (2011) 137-149.

[12] G. Marra, C.L. Chang, L.A. Laghi, D.P. Chauhan, D. Young, C.R. Boland, Expression of human MutS homolog 2 (hMSH2) protein in resting and proliferating cells, Oncogene, 13 (1996) 2189-2196.

[13] M. Seifert, J. Reichrath, The role of the human DNA mismatch repair gene hMSH2 in DNA repair, cell cycle control and apoptosis: implications for pathogenesis, progression and therapy of cancer, J Mol Histol, 37 (2006) 301-307.

[14] J. Jiricny, Postreplicative mismatch repair, Cold Spring Harb Perspect Biol, 5 (2013) a012633.

[15] O.D. Scharer, Nucleotide excision repair in eukaryotes, Cold Spring Harb Perspect Biol, 5 (2013) a012609.

[16] K.K. Chiruvella, Z. Liang, T.E. Wilson, Repair of double-strand breaks by end joining, Cold Spring Harb Perspect Biol, 5 (2013) a012757.

[17] M. Jasin, R. Rothstein, Repair of strand breaks by homologous recombination, Cold Spring Harb Perspect Biol, 5 (2013) a012740.

[18] A.J. Deans, S.C. West, DNA interstrand crosslink repair and cancer, Nat Rev Cancer, 11 (2011) 467-480.

[19] U. Jandt, O. Platas Barradas, R. Portner, A.P. Zeng, Mammalian cell culture synchronization under physiological conditions and population dynamic simulation, Appl Microbiol Biotechnol, (2014).

[20] Z. Bar-Joseph, Z. Siegfried, M. Brandeis, B. Brors, Y. Lu, R. Eils, B.D. Dynlacht, I. Simon, Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells, Proc Natl Acad Sci U S A, 105 (2008) 955-960.

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

[22] J. Pena-Diaz, S.A. Hegre, E. Anderssen, P.A. Aas, R. Mjelle, G.D. Gilfillan, R. Lyle, F. Drablos, H.E. Krokan, P. Saetrom, Transcription profiling during the cell cycle shows that a subset of Polycomb-targeted genes is upregulated during DNA replication, Nucleic acids research, 41 (2013) 2846-2856.

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

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

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

[26] L.J. Jensen, T.S. Jensen, U. de Lichtenberg, S. Brunak, P. Bork, Co-evolution of transcriptional and post-translational cell-cycle regulation, Nature, 443 (2006) 594-597.

[27] C.M. van der Meijden, D.S. Lapointe, M.X. Luong, D. Peric-Hupkes, B. Cho, J.L. Stein, A.J. van Wijnen, G.S. Stein, Gene profiling of cell cycle progression through S-phase reveals sequential expression of genes required for DNA replication and nucleosome assembly, Cancer Res, 62 (2002) 3233-3243.

[28] R.D. Wood, M. Mitchell, T. Lindahl, Human DNA repair genes, 2005, Mutat Res, 577 (2005) 275-283.

[29] R.D. Wood, M. Mitchell, J. Sgouros, T. Lindahl, Human DNA repair genes, Science, 291 (2001) 1284-1289.

[30] C.R. Stumpf, M.V. Moreno, A.B. Olshen, B.S. Taylor, D. Ruggero, The translational landscape of the mammalian cell cycle, Mol Cell, 52 (2013) 574-582.

[31] C. Trapnell, D.G. Hendrickson, M. Sauvageau, L. Goff, J.L. Rinn, L. Pachter, Differential analysis of gene regulation at transcript resolution with RNA-seq, Nature biotechnology, 31 (2013) 46-53.

[32] N. Leng, J.A. Dawson, J.A. Thomson, V. Ruotti, A.I. Rissman, B.M. Smits, J.D. Haag, M.N. Gould, R.M. Stewart, C. Kendziorski, EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments, Bioinformatics, 29 (2013) 1035-1043.

[33] R. Bravo, Synthesis of the nuclear protein cyclin (PCNA) and its relationship with DNA replication, Experimental cell research, 163 (1986) 287-293.

[34] C.M. Simbulan-Rosenthal, D.S. Rosenthal, A.H. Boulares, R.J. Hickey, L.H. Malkas, J.M. Coll, M.E. Smulson, Regulation of the expression or recruitment of components of the DNA synthesome by poly(ADP-ribose) polymerase, Biochemistry, 37 (1998) 9363-9370.

[35] J. Wang, J. Chen, Z. Gong, TopBP1 controls BLM protein level to maintain genome stability, Mol Cell, 52 (2013) 667-678.

[36] S.D. Dimitrov, D. Lu, N. Naetar, Y. Hu, S. Pathania, C. Kanellopoulou, D.M. Livingston, Physiological modulation of endogenous BRCA1 p220 abundance suppresses DNA damage during the cell cycle, Genes Dev, 27 (2013) 2274-2291.

[37] Z. Guo, J. Kanjanapangka, N. Liu, S. Liu, C. Liu, Z. Wu, Y. Wang, T. Loh, C. Kowolik, J. Jamsen, M. Zhou, K. Truong, Y. Chen, L. Zheng, B. Shen, Sequential posttranslational modifications program FEN1 degradation during cell-cycle progression, Mol Cell, 47 (2012) 444-456.

[38] T.M. Geiman, K. Muegge, Lsh, an SNF2/helicase family member, is required for proliferation of mature T lymphocytes, Proc Natl Acad Sci U S A, 97 (2000) 4772-4777.

[39] S.E. Polo, S.E. Theocharis, J. Klijanienko, A. Savignoni, B. Asselain, P. Vielh, G. Almouzni, Chromatin assembly factor-1, a marker of clinical value to distinguish quiescent from proliferating cells, Cancer Res, 64 (2004) 2371-2381.

[40] M.M. Heck, W.N. Hittelman, W.C. Earnshaw, Differential expression of DNA topoisomerases I and II during the eukaryotic cell cycle, Proc Natl Acad Sci U S A, 85 (1988) 1086-1090.

[41] S. Yamanaka, N.R. Campbell, F. An, S.C. Kuo, J.J. Potter, E. Mezey, A. Maitra, F.M. Selaru, Coordinated effects of microRNA-494 induce G(2)/M arrest in human cholangiocarcinoma, Cell Cycle, 11 (2012) 2729-2738.

[42] A. Kaidi, S.P. Jackson, KAT5 tyrosine phosphorylation couples chromatin sensing to ATM signalling, Nature, 498 (2013) 70-74.

[43] L. Strom, H.B. Lindroos, K. Shirahige, C. Sjogren, Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair, Mol Cell, 16 (2004) 1003-1015.

[44] S.Q. Gregg, A.R. Robinson, L.J. Niedernhofer, Physiological consequences of defects in ERCC1-XPF DNA repair endonuclease, DNA Repair (Amst), 10 (2011) 781-791.

[45] T.K. Hazra, T. Izumi, I. Boldogh, B. Imhoff, Y.W. Kow, P. Jaruga, M. Dizdaroglu, S. Mitra, Identification and characterization of a human DNA glycosylase for repair of modified bases in oxidatively damaged DNA, Proc Natl Acad Sci U S A, 99 (2002) 3523-3528.

[46] T.A. Nagelhus, G. Slupphaug, T. Lindmo, H.E. Krokan, Cell cycle regulation and subcellular localization of the major human uracil-DNA glycosylase, Experimental cell research, 220 (1995) 292-297.

[47] S.J. Muller, S. Caradonna, Cell cycle regulation of a human cyclin-like gene encoding uracil-DNA glycosylase, J Biol Chem, 268 (1993) 1310-1319.

[48] H. Nilsen, I. Rosewell, P. Robins, C.F. Skjelbred, S. Andersen, G. Slupphaug, G. Daly, H.E. Krokan, T. Lindahl, D.E. Barnes, Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication, Mol Cell, 5 (2000) 1059-1065.

[49] B. Kavli, O. Sundheim, M. Akbari, M. Otterlei, H. Nilsen, F. Skorpen, P.A. Aas, L. Hagen, H.E. Krokan, G. Slupphaug, hUNG2 is the major repair enzyme for removal of uracil from U:A matches, U:G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup, J Biol Chem, 277 (2002) 39926-39936.

[50] M.L. Hegde, P.M. Hegde, L.J. Bellot, S.M. Mandal, T.K. Hazra, G.M. Li, I. Boldogh, A.E. Tomkinson, S. Mitra, Prereplicative repair of oxidized bases in the human genome is mediated by NEIL1 DNA glycosylase together with replication proteins, Proc Natl Acad Sci U S A, 110 (2013) E3090-3099.

[51] G. Neurauter, S. Scholl-Burgi, A. Haara, S. Geisler, P. Mayersbach, H. Schennach, D. Fuchs, Simultaneous measurement of phenylalanine and tyrosine by high performance liquid chromatography (HPLC) with fluorescence detection, Clin Biochem, 46 (2013) 1848-1851.

[52] C.E. Strom, F. Johansson, M. Uhlen, C.A. Szigyarto, K. Erixon, T. Helleday, Poly (ADP-ribose) polymerase (PARP) is not involved in base excision repair but PARP inhibition traps a single-strand intermediate, Nucleic acids research, 39 (2011) 3166-3175.

[53] M. Akbari, J. Pena-Diaz, S. Andersen, N.B. Liabakk, M. Otterlei, H.E. Krokan, Extracts of proliferating and non-proliferating human cells display different base excision pathways and repair fidelity, DNA Repair (Amst), 8 (2009) 834-843.

[54] N. Levy, M. Oehlmann, F. Delalande, H.P. Nasheuer, A. Van Dorsselaer, V. Schreiber, G. de Murcia, J. Menissier-de Murcia, D. Maiorano, A. Bresson, XRCC1 interacts with the p58 subunit of DNA Pol alpha-primase and may coordinate DNA repair and replication during S phase, Nucleic acids research, 37 (2009) 3177-3188.

[55] T. Ogi, S. Limsirichaikul, R.M. Overmeer, M. Volker, K. Takenaka, R. Cloney, Y. Nakazawa, A. Niimi, Y. Miki, N.G. Jaspers, L.H. Mullenders, S. Yamashita, M.I. Fousteri, A.R. Lehmann, Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells, Mol Cell, 37 (2010) 714-727.

[56] Z. Li, P.R. Musich, M.A. Serrano, Z. Dong, Y. Zou, XPA-mediated regulation of global nucleotide excision repair by ATR Is p53-dependent and occurs primarily in S-phase, PLoS One, 6 (2011) e28326.

[57] S.L. Yu, M.S. Kang, H.Y. Kim, C.M. Gorospe, T.S. Kim, S.K. Lee, The PCNA binding domain of Rad2p plays a role in mutagenesis by modulating the cell cycle in response to DNA damage, DNA Repair (Amst), 16 (2014) 1-10.

[58] H. Klingler, C. Hemmerle, F. Bannwart, R. Haider, M.S. Cattaruzza, G. Marra, Expression of the hMSH6 mismatch-repair protein in colon cancer and HeLa cells, Swiss Med Wkly, 132 (2002) 57-63.

[59] M. Meyers, M. Theodosiou, S. Acharya, E. Odegaard, T. Wilson, J.E. Lewis, T.W. Davis, C. Wilson-Van Patten, R. Fishel, D.A. Boothman, Cell cycle regulation of the human DNA mismatch repair genes hMSH2, hMLH1, and hPMS2, Cancer Res, 57 (1997) 206-208.

[60] T. Snowden, S. Acharya, C. Butz, M. Berardini, R. Fishel, hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes, Mol Cell, 15 (2004) 437-451.

[61] Y. Kee, A.D. D'Andrea, Molecular pathogenesis and clinical management of Fanconi anemia, J Clin Invest, 122 (2012) 3799-3806.

[62] A. Smogorzewska, R. Desetty, T.T. Saito, M. Schlabach, F.P. Lach, M.E. Sowa, A.B. Clark, T.A. Kunkel, J.W. Harper, M.P. Colaiacovo, S.J. Elledge, A genetic screen identifies FAN1, a Fanconi anemia-associated nuclease necessary for DNA interstrand crosslink repair, Mol Cell, 39 (2010) 36-47.

[63] C. Beck, I. Robert, B. Reina-San-Martin, V. Schreiber, F. Dantzer, Poly(ADP-ribose) polymerases in double-strand break repair: Focus on PARP1, PARP2 and PARP3, Experimental cell research, (2014).

[64] S.B. De Lorenzo, A.G. Patel, R.M. Hurley, S.H. Kaufmann, The Elephant and the Blind Men: Making Sense of PARP Inhibitors in Homologous Recombination Deficient Tumor Cells, Frontiers in oncology, 3 (2013) 228.

[65] M. Li, X. Yu, Function of BRCA1 in the DNA damage response is mediated by ADP-ribosylation, Cancer cell, 23 (2013) 693-704.

[66] R.J. Pezza, O.N. Voloshin, A.A. Volodin, K.A. Boateng, M.A. Bellani, A.V. Mazin, R.D. Camerini-Otero, The dual role of HOP2 in mammalian meiotic homologous recombination, Nucleic acids research, 42 (2014) 2346-2357.

[67] J. Yuan, J. Chen, FIGNL1-containing protein complex is required for efficient homologous recombination repair, Proc Natl Acad Sci U S A, 110 (2013) 10640-10645.

[68] K.A. Manthei, J.L. Keck, The BLM dissolvasome in DNA replication and repair, Cell Mol Life Sci, 70 (2013) 4067-4084.

[69] D.S. Kim, J.A. Franklyn, V.E. Smith, A.L. Stratford, H.N. Pemberton, A. Warfield, J.C. Watkinson, T. Ishmail, M.J. Wakelam, C.J. McCabe, Securin induces genetic instability in colorectal cancer by inhibiting double-stranded DNA repair activity, Carcinogenesis, 28 (2007) 749-759.

[70] F. Romero, M.C. Multon, F. Ramos-Morales, A. Dominguez, J.A. Bernal, J.A. Pintor-Toro, M. Tortolero, Human securin, hPTTG, is associated with Ku heterodimer, the regulatory subunit of the DNA-dependent protein kinase, Nucleic acids research, 29 (2001) 1300-1307.

[71] F. Uhlmann, Chromosome cohesion and separation: from men and molecules, Curr Biol, 13 (2003) R104-114.

[72] I. Waizenegger, J.F. Gimenez-Abian, D. Wernic, J.M. Peters, Regulation of human separase by securin binding and autocleavage, Curr Biol, 12 (2002) 1368-1378.

[73] H. Karra, R. Pitkanen, M. Nykanen, K. Talvinen, T. Kuopio, M. Soderstrom, P. Kronqvist, Securin predicts aneuploidy and survival in breast cancer, Histopathology, 60 (2012) 586-596.

[74] A. Dominguez, F. Ramos-Morales, F. Romero, R.M. Rios, F. Dreyfus, M. Tortolero, J.A. Pintor-Toro, hpttg, a human homologue of rat pttg, is overexpressed in hematopoietic neoplasms. Evidence for a transcriptional activation function of hPTTG, Oncogene, 17 (1998) 2187-2193.

[75] A.P. Heaney, R. Singson, C.J. McCabe, V. Nelson, M. Nakashima, S. Melmed, Expression of pituitary-tumour transforming gene in colorectal tumours, Lancet, 355 (2000) 716-719.

[76] C. Solbach, M. Roller, C. Fellbaum, M. Nicoletti, M. Kaufmann, PTTG mRNA expression in primary breast cancer: a prognostic marker for lymph node invasion and tumor recurrence, Breast, 13 (2004) 80-81.

[77] Y. Tong, T. Eigler, Transcriptional targets for pituitary tumor-transforming gene-1, J Mol Endocrinol, 43 (2009) 179-185.

[78] J.E. Mermoud, S.P. Rowbotham, P.D. Varga-Weisz, Keeping chromatin quiet: how nucleosome remodeling restores heterochromatin after replication, Cell Cycle, 10 (2011) 4017-4025.

[79] C.L. Peterson, G. Almouzni, Nucleosome dynamics as modular systems that integrate DNA damage and repair, Cold Spring Harb Perspect Biol, 5 (2013).

[80] L.Y. Kadyrova, E.R. Blanko, F.A. Kadyrov, CAF-I-dependent control of degradation of the discontinuous strands during mismatch repair, Proc Natl Acad Sci U S A, 108 (2011) 2753-2758.

[81] B. Schopf, S. Bregenhorn, J.P. Quivy, F.A. Kadyrov, G. Almouzni, J. Jiricny, Interplay between mismatch repair and chromatin assembly, Proc Natl Acad Sci U S A, 109 (2012) 1895-1900.

[82] Q. Zhu, G. Wani, H.H. Arab, M.A. El-Mahdy, A. Ray, A.A. Wani, Chromatin restoration following nucleotide excision repair involves the incorporation of ubiquitinated H2A at damaged genomic sites, DNA Repair (Amst), 8 (2009) 262-273.

[83] J. Burrage, A. Termanis, A. Geissner, K. Myant, K. Gordon, I. Stancheva, The SNF2 family ATPase LSH promotes phosphorylation of H2AX and efficient repair of DNA double-strand breaks in mammalian cells, J Cell Sci, 125 (2012) 5524-5534.

[84] T. Obayashi, Y. Okamura, S. Ito, S. Tadaka, I.N. Motoike, K. Kinoshita, COXPRESdb: a database of comparative gene coexpression networks of eleven species for mammals, Nucleic acids research, 41 (2013) D1014-1020.

[85] B. von Eyss, J. Maaskola, S. Memczak, K. Mollmann, A. Schuetz, C. Loddenkemper, M.D. Tanh, A. Otto, K. Muegge, U. Heinemann, N. Rajewsky, U. Ziebold, The SNF2-like helicase HELLS mediates E2F3-dependent transcription and cellular transformation, EMBO J, 31 (2012) 972-985.

[86] M. Ciro, E. Prosperini, M. Quarto, U. Grazini, J. Walfridsson, F. McBlane, P. Nucifero, G. Pacchiana, M. Capra, J. Christensen, K. Helin, ATAD2 is a novel cofactor for MYC, overexpressed and amplified in aggressive tumors, Cancer Res, 69 (2009) 8491-8498.

[87] E.Y. Hsia, M.L. Goodson, J.X. Zou, M.L. Privalsky, H.W. Chen, Nuclear receptor coregulators as a new paradigm for therapeutic targeting, Adv Drug Deliv Rev, 62 (2010) 1227-1237.

[88] A.S. Revenko, E.V. Kalashnikova, A.T. Gemo, J.X. Zou, H.W. Chen, Chromatin loading of E2F-MLL complex by cancer-associated coregulator ANCCA via reading a specific histone mark, Mol Cell Biol, 30 (2010) 5260-5272.

[89] C. Caron, C. Lestrat, S. Marsal, E. Escoffier, S. Curtet, V. Virolle, P. Barbry, A. Debernardi, C. Brambilla, E. Brambilla, S. Rousseaux, S. Khochbin, Functional characterization of ATAD2 as a new cancer/testis factor and a predictor of poor prognosis in breast and lung cancers, Oncogene, 29 (2010) 5171-5181.

[90] M.R. Duan, M.J. Smerdon, Histone H3 Lysine 14 (H3K14) Acetylation Facilitates DNA Repair in a Positioned Nucleosome by Stabilizing the Binding of the Chromatin Remodeler RSC (Remodels Structure of Chromatin), J Biol Chem, 289 (2014) 8353-8363.

[91] F. Boussouar, M. Jamshidikia, Y. Morozumi, S. Rousseaux, S. Khochbin, Malignant genome reprogramming by ATAD2, Biochim Biophys Acta, 1829 (2013) 1010-1014.

[92] R. Fouret, J. Laffaire, P. Hofman, M. Beau-Faller, J. Mazieres, P. Validire, P. Girard, S. Camilleri-Broet, F. Vaylet, F. Leroy-Ladurie, J.C. Soria, P. Fouret, A comparative and integrative approach identifies ATPase family, AAA domain containing 2 as a likely driver of cell proliferation in lung adenocarcinoma, Clin Cancer Res, 18 (2012) 5606-5616.

[93] E.V. Kalashnikova, A.S. Revenko, A.T. Gemo, N.P. Andrews, C.G. Tepper, J.X. Zou, R.D. Cardiff, A.D. Borowsky, H.W. Chen, ANCCA/ATAD2 overexpression identifies breast cancer patients with poor prognosis, acting to drive proliferation and survival of triple-negative cells through control of B-Myb and EZH2, Cancer Res, 70 (2010) 9402-9412.

[94] M.B. Raeder, E. Birkeland, J. Trovik, C. Krakstad, S. Shehata, S. Schumacher, T.I. Zack, A. Krohn, H.M. Werner, S.E. Moody, E. Wik, I.M. Stefansson, F. Holst, A.M. Oyan, P. Tamayo, J.P. Mesirov, K.H. Kalland, L.A. Akslen, R. Simon, R. Beroukhim, H.B. Salvesen, Integrated genomic analysis of the 8q24 amplification in endometrial cancers identifies ATAD2 as essential to MYC-dependent cancers, PLoS One, 8 (2013) e54873.

[95] J.X. Zou, A.S. Revenko, L.B. Li, A.T. Gemo, H.W. Chen, ANCCA, an estrogen-regulated AAA+ ATPase coactivator for ERalpha, is required for coregulator occupancy and chromatin modification, Proc Natl Acad Sci U S A, 104 (2007) 18067-18072.

[96] C.D. Jarvis, T. Geiman, M.P. Vila-Storm, O. Osipovich, U. Akella, S. Candeias, I. Nathan, S.K. Durum, K. Muegge, A novel putative helicase produced in early murine lymphocytes, Gene, 169 (1996) 203-207.

[97] Y. Liu, Z. Gong, L. Sun, X. Li, FOXM1 and androgen receptor co-regulate CDC6 gene transcription and DNA replication in prostate cancer cells, Biochim Biophys Acta, 1839 (2014) 297-305.

[98] A. Waseem, M. Ali, E.W. Odell, F. Fortune, M.T. Teh, Downstream targets of FOXM1: CEP55 and HELLS are cancer progression markers of head and neck squamous cell carcinoma, Oral Oncol, 46 (2010) 536-542.

[99] J. Laoukili, M.R. Kooistra, A. Bras, J. Kauw, R.M. Kerkhoven, A. Morrison, H. Clevers, R.H. Medema, FoxM1 is required for execution of the mitotic programme and chromosome stability, Nat Cell Biol, 7 (2005) 126-136.

[100] M.T. Teh, E. Gemenetzidis, D. Patel, R. Tariq, A. Nadir, A.W. Bahta, A. Waseem, I.L. Hutchison, FOXM1 induces a global methylation signature that mimics the cancer epigenome in head and neck squamous cell carcinoma, PLoS One, 7 (2012) e34329.

[101] H. Zhu, T.M. Geiman, S. Xi, Q. Jiang, A. Schmidtmann, T. Chen, E. Li, K. Muegge, Lsh is involved in de novo methylation of DNA, EMBO J, 25 (2006) 335-345.

[102] T. Fan, A. Schmidtmann, S. Xi, V. Briones, H. Zhu, H.C. Suh, J. Gooya, J.R. Keller, H. Xu, J. Roayaei, M. Anver, S. Ruscetti, K. Muegge, DNA hypomethylation caused by Lsh deletion promotes erythroleukemia development, Epigenetics, 3 (2008) 134-142.

[103] K. Myant, I. Stancheva, LSH cooperates with DNA methyltransferases to repress transcription, Mol Cell Biol, 28 (2008) 215-226.

[104] Q. Yan, J. Huang, T. Fan, H. Zhu, K. Muegge, Lsh, a modulator of CpG methylation, is crucial for normal histone methylation, EMBO J, 22 (2003) 5154-5162.

[105] R. Zhou, L. Han, G. Li, T. Tong, Senescence delay and repression of p16INK4a by Lsh via recruitment of histone deacetylases in human diploid fibroblasts, Nucleic acids research, 37 (2009) 5183-5196.

[106] D.W. Lee, K. Zhang, Z.Q. Ning, E.H. Raabe, S. Tintner, R. Wieland, B.J. Wilkins, J.M. Kim, R.I. Blough, R.J. Arceci, Proliferation-associated SNF2-like gene (PASG): a SNF2 family member altered in leukemia, Cancer Res, 60 (2000) 3612-3622.

[107] S.J. Ceballos, W.D. Heyer, Functions of the Snf2/Swi2 family Rad54 motor protein in homologous recombination, Biochim Biophys Acta, 1809 (2011) 509-523.

[108] Z. Zhang, H.Y. Fan, J.A. Goldman, R.E. Kingston, Homology-driven chromatin remodeling by human RAD54, Nat Struct Mol Biol, 14 (2007) 397-405.

[109] S. Oh, Y. Wang, J. Zimbric, E.A. Hendrickson, Human LIGIV is synthetically lethal with the loss of Rad54B-dependent recombination and is required for certain chromosome fusion events induced by telomere dysfunction, Nucleic acids research, 41 (2013) 1734-1749.

[110] E.C. Dykhuizen, D.C. Hargreaves, E.L. Miller, K. Cui, A. Korshunov, M. Kool, S. Pfister, Y.J. Cho, K. Zhao, G.R. Crabtree, BAF complexes facilitate decatenation of DNA by topoisomerase IIalpha, Nature, 497 (2013) 624-627.

[111] G.D. Grant, L. Brooks, 3rd, X. Zhang, J.M. Mahoney, V. Martyanov, T.A. Wood, G. Sherlock, C. Cheng, M.L. Whitfield, Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors, Mol Biol Cell, 24 (2013) 3634-3650.

[112] K. Masuda, R. Ouchida, A. Takeuchi, T. Saito, H. Koseki, K. Kawamura, M. Tagawa, T. Tokuhisa, T. Azuma, O.W. J, DNA polymerase theta contributes to the generation of C/G mutations during somatic hypermutation of Ig genes, Proc Natl Acad Sci U S A, 102 (2005) 13986-13991.

[113] B.Z. Zmudzka, A. Fornace, J. Collins, S.H. Wilson, Characterization of DNA polymerase beta mRNA: cell-cycle and growth response in cultured human cells, Nucleic acids research, 16 (1988) 9587-9596.

[114] D. Kidane, A.S. Jonason, T.S. Gorton, I. Mihaylov, J. Pan, S. Keeney, D.G. de Rooij, T. Ashley, A. Keh, Y. Liu, U. Banerjee, D. Zelterman, J.B. Sweasy, DNA polymerase beta is critical for mouse meiotic synapsis, EMBO J, 29 (2010) 410-423.

[115] G.L. Dianov, U. Hubscher, Mammalian base excision repair: the forgotten archangel, Nucleic acids research, 41 (2013) 3483-3490.

[116] Y. Zhao, J.R. Brickner, M.C. Majid, N. Mosammaparast, Crosstalk between ubiquitin and other post-translational modifications on chromatin during double-strand break repair, Trends Cell Biol, (2014).

[117] D. Branzei, M. Foiani, Regulation of DNA repair throughout the cell cycle, Nat Rev Mol Cell Biol, 9 (2008) 297-308.

[118] L.C. Wang, J. Gautier, The Fanconi anemia pathway and ICL repair: implications for cancer therapy, Crit Rev Biochem Mol Biol, 45 (2010) 424-439.

[119] S.A. Williams, J.B. Wilson, A.P. Clark, A. Mitson-Salazar, A. Tomashevski, S. Ananth, P.M. Glazer, O.J. Semmes, A.E. Bale, N.J. Jones, G.M. Kupfer, Functional and physical interaction between the mismatch repair and FA-BRCA pathways, Hum Mol Genet, 20 (2011) 4395-4410.

[120] L.W. McMahon, J. Sangerman, S.R. Goodman, K. Kumaresan, M.W. Lambert, Human alpha spectrin II and the FANCA, FANCC, and FANCG proteins bind to DNA containing psoralen interstrand cross-links, Biochemistry, 40 (2001) 7025-7034.

[121] P. Zhang, U. Herbig, F. Coffman, M.W. Lambert, Non-erythroid alpha spectrin prevents telomere dysfunction after DNA interstrand cross-link damage, Nucleic acids research, 41 (2013) 5321-5340.

[122] H.L. Williams, M.E. Gottesman, J. Gautier, The differences between ICL repair during and outside of S phase, Trends Biochem Sci, 38 (2013) 386-393.

[123] T. Thorslund, S.C. West, BRCA2: a universal recombinase regulator, Oncogene, 26 (2007) 7720-7730.

[124] L.Y. Marmorstein, A.V. Kinev, G.K. Chan, D.A. Bochar, H. Beniya, J.A. Epstein, T.J. Yen, R. Shiekhattar, A human BRCA2 complex containing a structural DNA binding component influences cell cycle progression, Cell, 104 (2001) 247-257.

[125] M. Shrivastav, L.P. De Haro, J.A. Nickoloff, Regulation of DNA double-strand break repair pathway choice, Cell Res, 18 (2008) 134-147.

[126] T. Furuta, H. Takemura, Z.Y. Liao, G.J. Aune, C. Redon, O.A. Sedelnikova, D.R. Pilch, E.P. Rogakou, A. Celeste, H.T. Chen, A. Nussenzweig, M.I. Aladjem, W.M. Bonner, Y. Pommier, Phosphorylation of histone H2AX and activation of Mre11, Rad50, and Nbs1 in response to replication-dependent DNA double-strand breaks induced by mammalian DNA topoisomerase I cleavage complexes, J Biol Chem, 278 (2003) 20303-20312.

[127] J.H. Lee, T.T. Paull, Direct activation of the ATM protein kinase by the Mre11/Rad50/Nbs1 complex, Science, 304 (2004) 93-96.

[128] L.S. Symington, J. Gautier, Double-strand break end resection and repair pathway choice, Annu Rev Genet, 45 (2011) 247-271.

[129] M. Papamichos-Chronakis, C.L. Peterson, Chromatin and the genome integrity network, Nat Rev Genet, 14 (2013) 62-75.

[130] G. Bennett, M. Papamichos-Chronakis, C.L. Peterson, DNA repair choice defines a common pathway for recruitment of chromatin regulators, Nat Commun, 4 (2013) 2084.

[131] H. Menoni, D. Gasparutto, A. Hamiche, J. Cadet, S. Dimitrov, P. Bouvet, D. Angelov, ATP-dependent chromatin remodeling is required for base excision repair in conventional but not in variant H2A.Bbd nucleosomes, Mol Cell Biol, 27 (2007) 5949-5956.

[132] S. Nakanishi, R. Prasad, S.H. Wilson, M. Smerdon, Different structural states in oligonucleosomes are required for early versus late steps of base excision repair, Nucleic acids research, 35 (2007) 4313-4321.

Mjelle et al., Figure 1

A

Cell cycle regulation of human DNA repair and chromatin remodeling genes

Robin Mjelle, Siv Anita Hegre, Per Arne Aas, Finn Drabløs, Geir Slupphaug, Pål Sætrom, Hans E. Krokan

HOME | ARTICLE | GENE LISTS | CELL CYCLE | METHODS | RESEARCH GROUP

Chromatin remodeling | DNA repair | Chromatin remodeling | DNA repair

Welcome to the webpage on DNA repair and chromatin remodeling genes and their cell cycle regulation

B

This table provides an updated list of DNA repair genes. The table also contains chromatin remodeling genes with putative DNA repair function. Each gene is linked to OMIM for functional information and to NCBI and UCSC genome browser for detailed information on the gene.

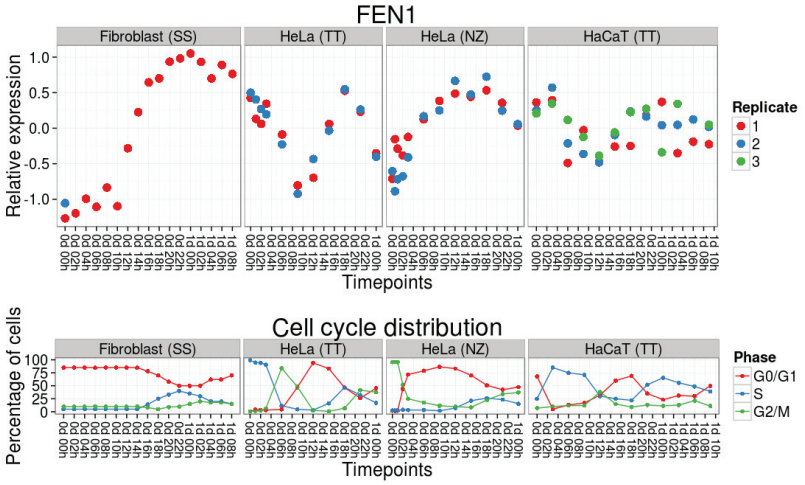
Gene Symbol (Link to function in OMIM)	RefSeq ID (Link to NCBI)	Chromosome Location (Link to UCSC Genome Browser)
ACTB	NM_001101.3	7p22.1
ACTL6A	NM_004301.3	3q26.33
ACTL6B	NM_016188.4	7q22.1
ACTR3B	NM_001040135.2	7q36.1
ACTR5	NM_024855.3	20q11.23
ACTR6	NM_022496.4	12q23.1
ALKBH1	NM_006020.2	14q24.3
ALKBH2	NM_001001655.2	12q24.11
ALKBH3	NM_139178.3	11p11.2
ANKRD17	NM_032217.3	4q13.3
APEX1	NM_001244249.1	14q11.2
APEX2	NM_001244249.1	Xp11.2
APLF	NM_173545.2	2p13.3
APTX	NM_001195248.1	9p13.3

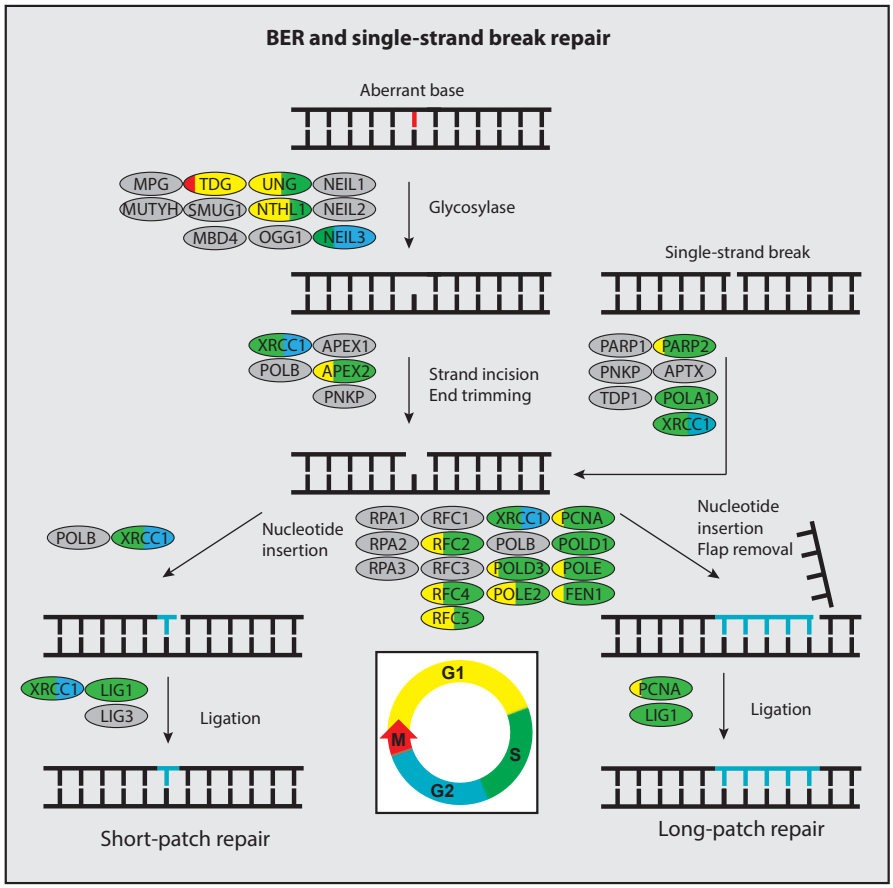
C

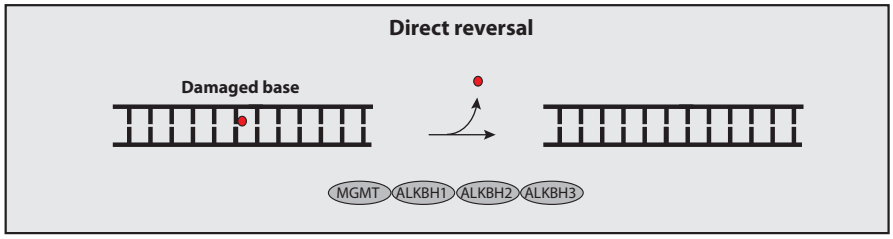
The table lists cell cycle regulated DNA repair genes sorted by pathway. Five transcriptome data sets, in which the genes that are periodically expressed in at least one data set, and one ribosomal profiling (translationally regulated) data set are included. Periodically expressed genes are indicated with cell cycle phase in which they show the highest expression level. Gene sets in which a given gene is not significantly cell cycle regulated is indicated as NR (not transcriptionally regulated in the cell cycle) or NTR (not translationally regulated in the cell cycle) in transcriptome and ribosomal profiling data sets, respectively. Four experiments are supplemented with expression profiles which can be found by clicking the gene name. Synchronization methods are abbreviated as: SS=serum starvation, TT= double thymidine block, NZ=nocodazole, MS=mitotic shake-off

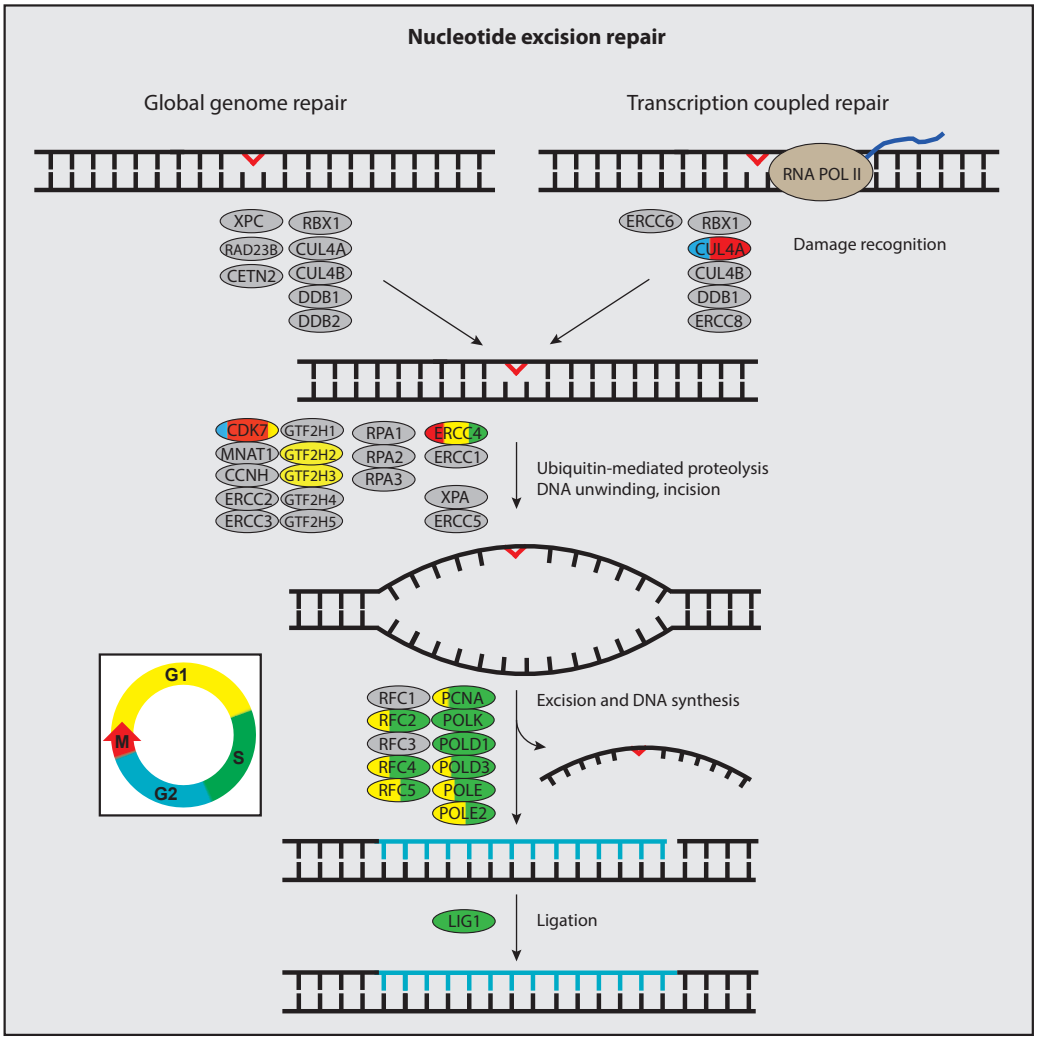
Experiment	Bar-Joseph Z et al. (2008)	Mjelle et al. (2014)	Peña-Diaz J et al. (2013)	Peña-Diaz J et al. (2013)	Whitfield ML et al. (2002)	Stumpf et al. (2013)	Stumpf et al. (2013)
Cell type	Fibroblast	HeLa	HeLa	HaCaT	HeLa	HeLa	HeLa
Synchronization	SS	TT	NZ	TT	TT/NZ/MS	TT and NZ	TT and NZ
Method	Microarray	Microarray	Microarray	Microarray	Microarray	RNA-Seq	Ribosomal profiling
Base Excision Repair Pathway							
APEX2	NR	G1/S	G1/S	NR	NR	S	NTR
APTX	NR	NR	M/G1	NR	NR	NR	NTR
CHAF1A	G1/S	NR	G1/S	G1/S	G1/S	NR	NTR
CHAF1B	G1/S	G1/S	G1/S	G1/S	G1/S	S	NTR
CSNK1D	NR	G1/S	NR	M/G1	NR	S	NTR
CSNK1E	NR	G2/M	S	NR	NR	NR	NTR
ERCC6L	S	G2	NR	G2	NR	NR	NTR
FEN1	G1/S	G1/S	S	G1/S	S	S	NTR
HMGB1	S	NR	NR	NR	NR	NR	NTR
HMGB2	S	NR	G2	G2	NR	S,G1 and M	NTR

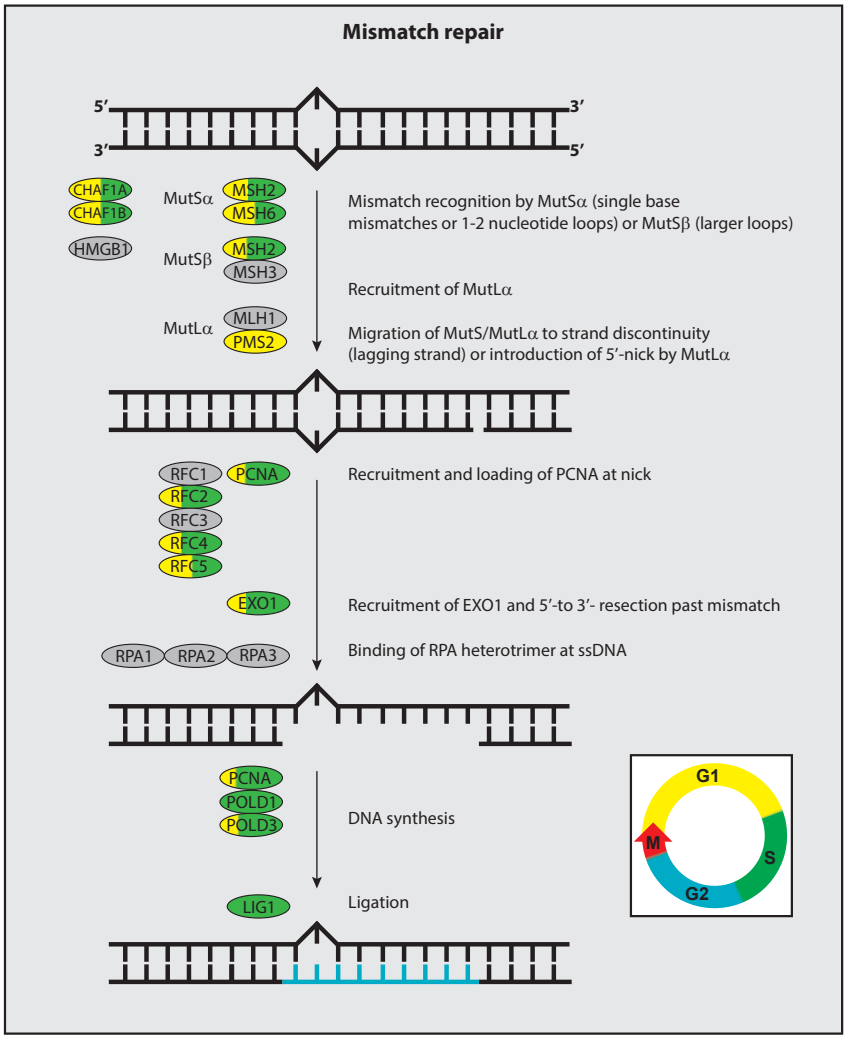
D

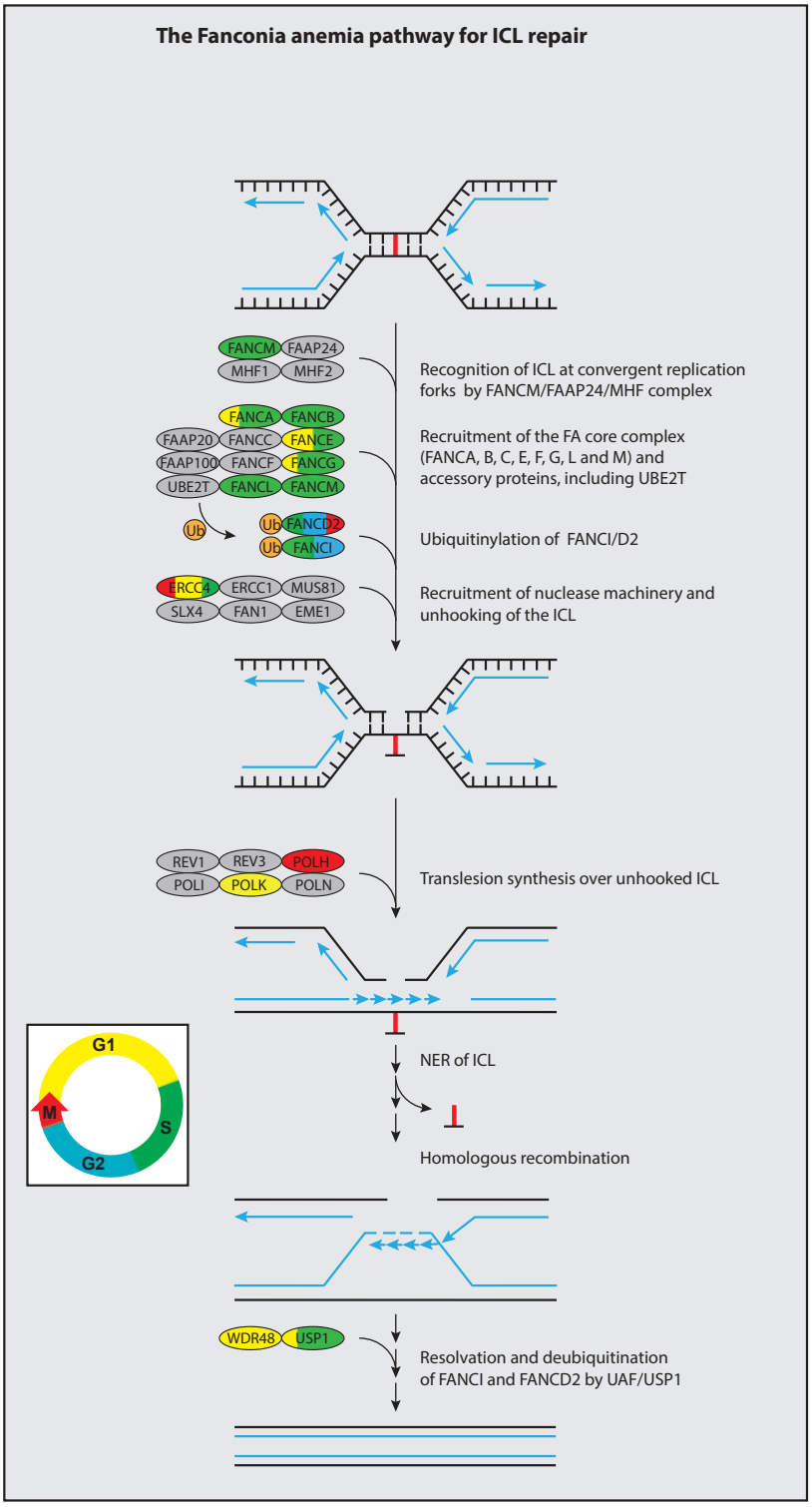


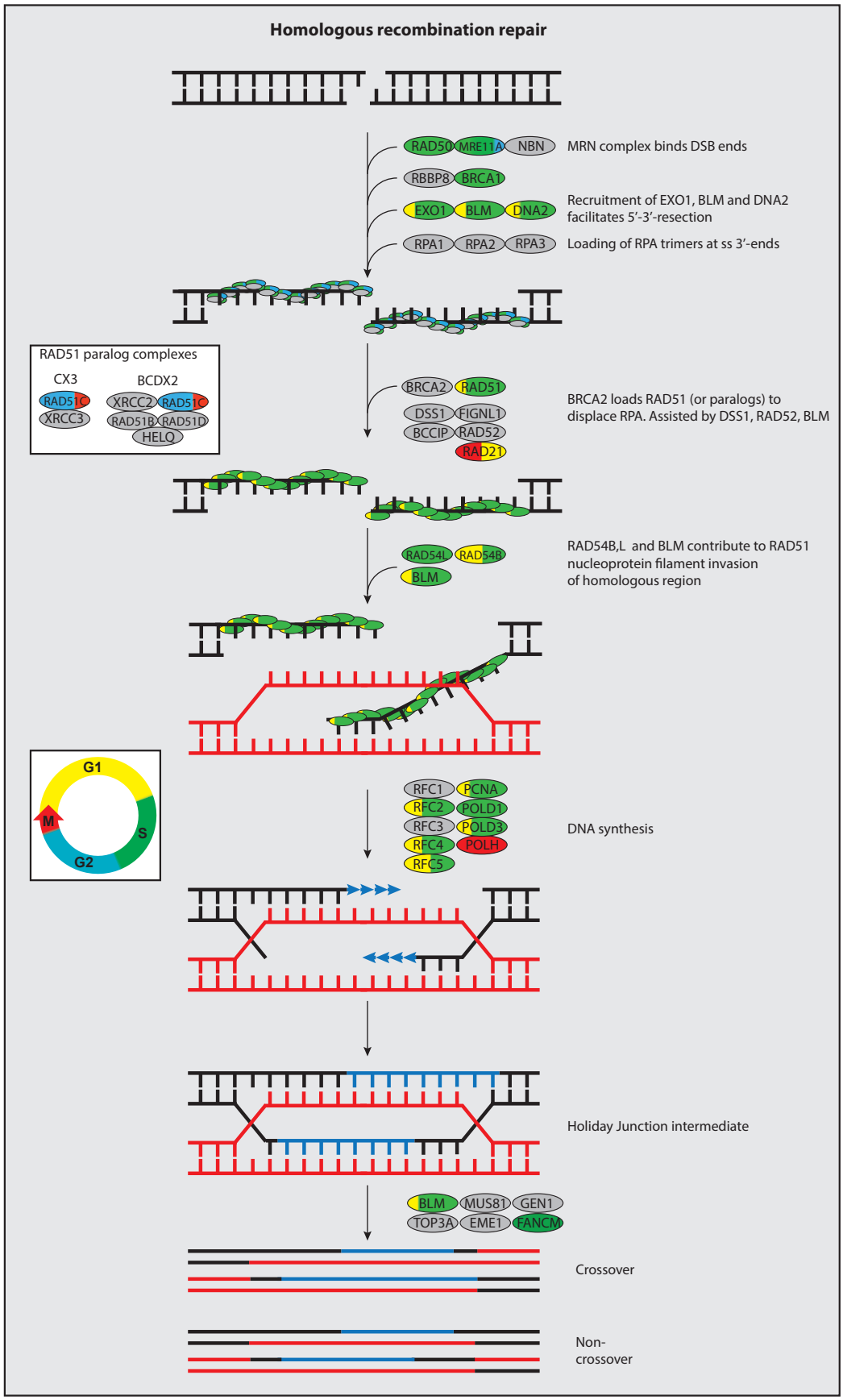

BER and single-strand break repair
Aberrant base
MPG
TDG
UNG
NEIL1
MUTYH
SMUG1
NTHL1
NEIL2
MBD4
OGG1
NEIL3
Glycosylase
Single-strand break
XRCC1
APEX1
POLB
APEX2
PNKP
Strand incision
End trimming
PARP1
PARP2
PNKP
APTX
TDP1
POLA1
XRCC1
POLB
XRCC1
Nucleotide insertion
RPA1
RPA2
RPA3
RFC1
RFC2
RFC3
RFC4
RFC5
XRCC1
POLB
PCNA
POLD1
POLD3
POLE
POLE2
FEN1
Nucleotide insertion
Flap removal
XRCC1
LIG1
LIG3
Ligation
G1
S
G2
M
PCNA
LIG1
Ligation
Short-patch repair
Long-patch repair

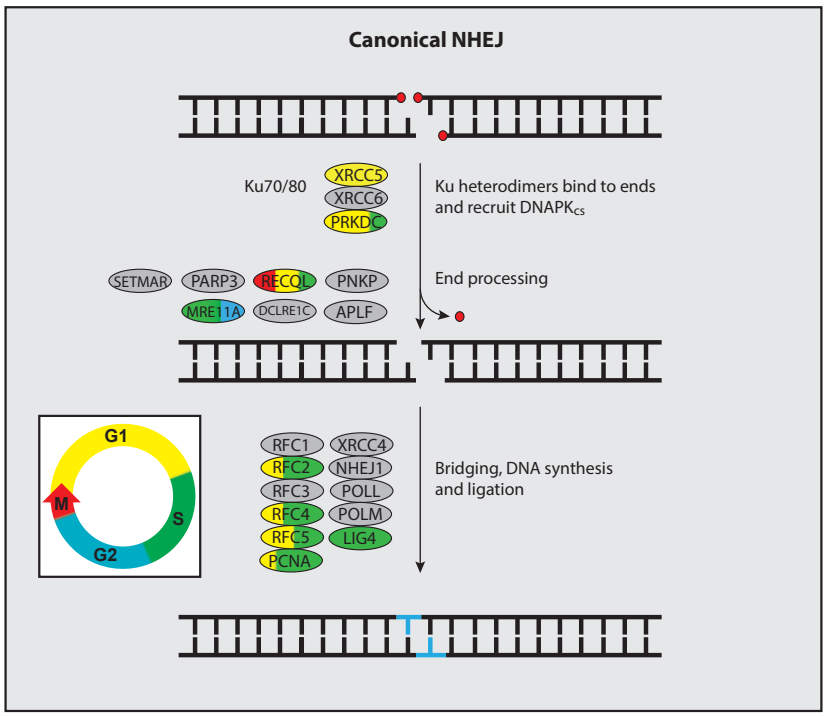

Direct reversal
Damaged base
MGMT
ALKBH1
ALKBH2
ALKBH3

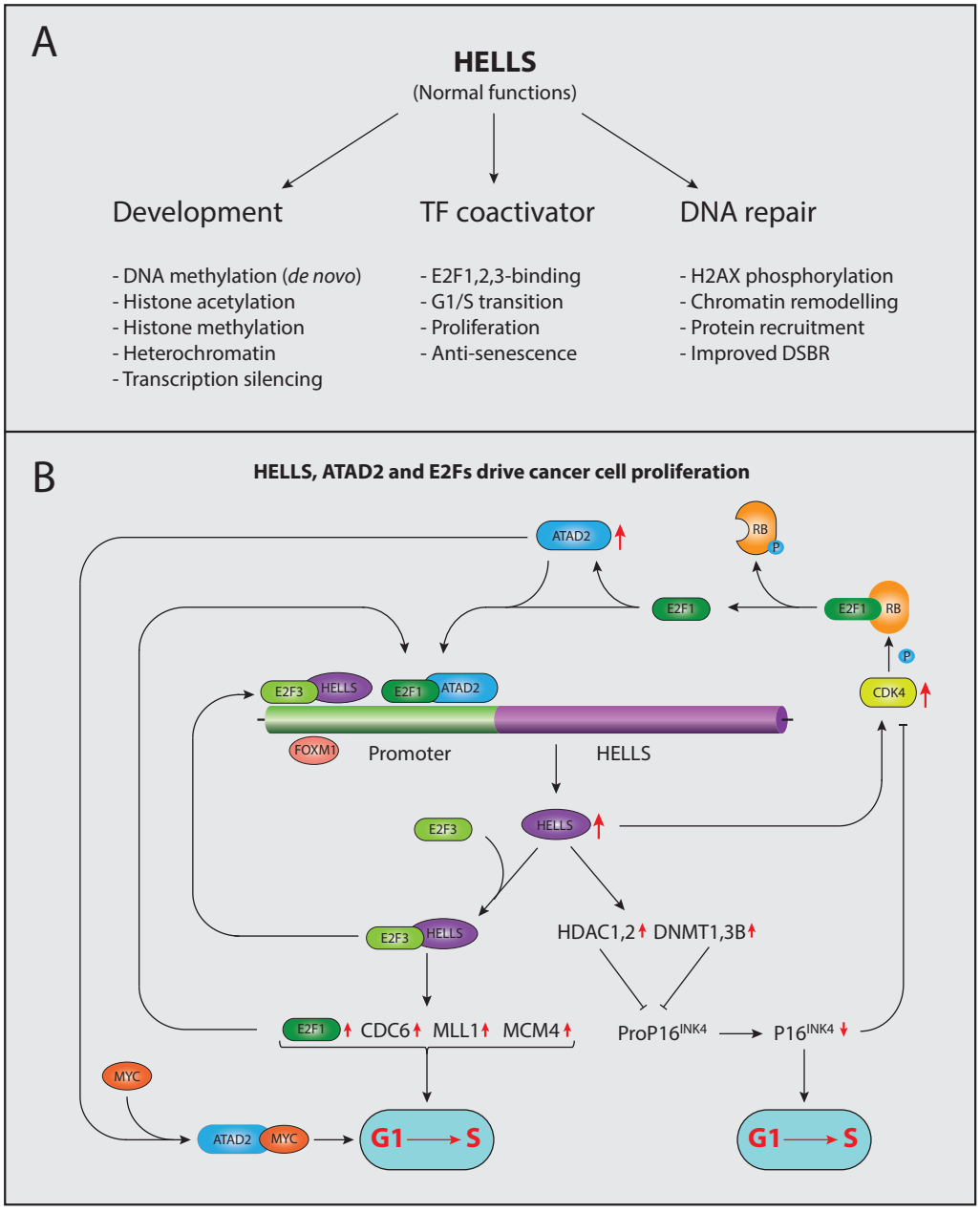

Nucleotide excision repair
Global genome repair
Transcription coupled repair
RNA POL II
XPC
RBX1
RAD23B
CUL4A
CETN2
CUL4B
DDB1
DDB2
ERCC6
RBX1
CUL4A
CUL4B
DDB1
ERCC8
Damage recognition
CDK7
GTF2H1
MNAT1
GTF2H2
CCNH
GTF2H3
ERCC2
GTF2H4
ERCC3
GTF2H5
RPA1
RPA2
RPA3
ERCC4
ERCC1
XPA
ERCC5
Ubiquitin-mediated proteolysis
DNA unwinding, incision
G1
S
G2
M
RFC1
RFC2
RFC3
RFC4
RFC5
PCNA
POLK
POLD1
POLD3
POLE
POLE2
Excision and DNA synthesis
LIG1
Ligation


Mismatch repair
5′
3′
3′
5′
CHAF1A
CHAF1B
MutSα
MSH2
MSH6
HMGB1
MutSβ
MSH2
MSH3
Mismatch recognition by MutSα (single base mismatches or 1-2 nucleotide loops) or MutSβ (larger loops)
Recruitment of MutLα
MutLα
MLH1
PMS2
Migration of MutS/MutLα to strand discontinuity (lagging strand) or introduction of 5′-nick by MutLα
RFC1
RFC2
RFC3
RFC4
RFC5
PCNA
Recruitment and loading of PCNA at nick
EXO1
Recruitment of EXO1 and 5′-to 3′- resection past mismatch
RPA1
RPA2
RPA3
Binding of RPA heterotrimer at ssDNA
PCNA
POLD1
POLD3
DNA synthesis
LIG1
Ligation
G1
S
G2
M

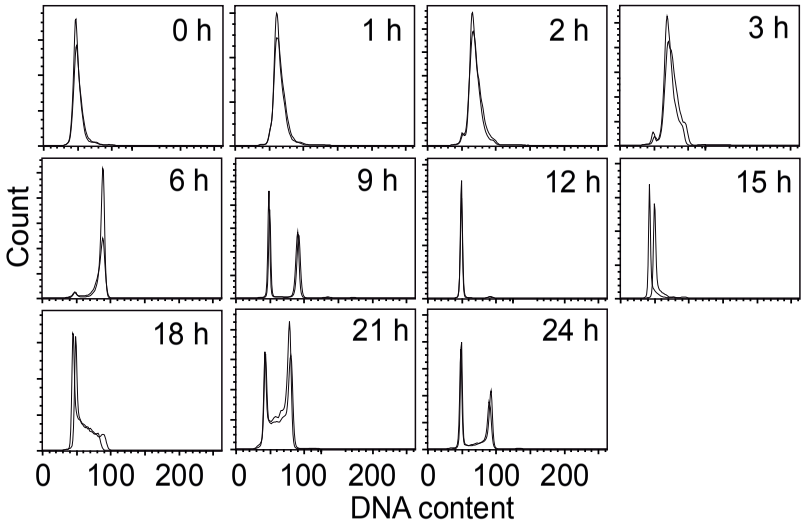

The Fanconia anemia pathway for ICL repair
FANCM
FAAP24
MHF1
MHF2
Recognition of ICL at convergent replication forks by FANCM/FAAP24/MHF complex
FANCA
FANCB
FAAP20
FANCC
FANCE
FAAP100
FANCF
FANCG
UBE2T
FANCL
FANCM
Recruitment of the FA core complex (FANCA, B, C, E, F, G, L and M) and accessory proteins, including UBE2T
Ub
Ub
FANCD2
Ub
FANCI
Ubiquitinylation of FANCI/D2
ERCC4
ERCC1
MUS81
SLX4
FAN1
EME1
Recruitment of nuclease machinery and unhooking of the ICL
REV1
REV3
POLH
POLI
POLK
POLN
Translesion synthesis over unhooked ICL
G1
S
G2
M
NER of ICL
Homologous recombination
WDR48
USP1
Resolvation and deubiquitination of FANCI and FANCD2 by UAF/USP1

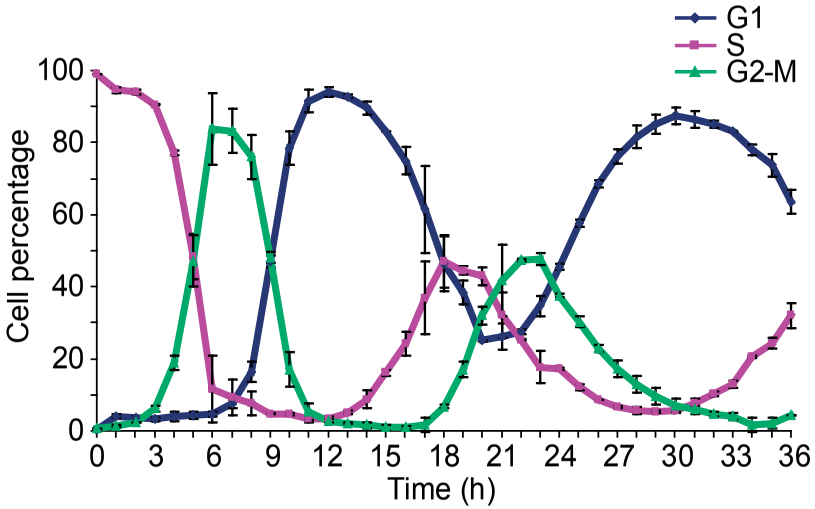

Homologous recombination repair
RAD50
MRE11A
NBN
MRN complex binds DSB ends
RBBP8
BRCA1
EXO1
BLM
DNA2
Recruitment of EXO1, BLM and DNA2 facilitates 5'-3'-resection
RPA1
RPA2
RPA3
Loading of RPA trimers at ss 3'-ends
RAD51 paralog complexes
CX3
RAD51C
XRCC3
BCDX2
XRCC2
RAD51C
RAD51B
RAD51D
HELQ
BRCA2
RAD51
DSS1
FIGNL1
BCCIP
RAD52
RAD21
BRCA2 loads RAD51 (or paralogs) to displace RPA. Assisted by DSS1, RAD52, BLM
RAD54L
RAD54B
BLM
RAD54B,L and BLM contribute to RAD51 nucleoprotein filament invasion of homologous region
G1
S
G2
M
RFC1
RFC2
RFC3
RFC4
RFC5
PCNA
POLD1
POLD3
POLH
DNA synthesis
Holiday Junction intermediate
BLM
TOP3A
MUS81
EME1
GEN1
FANCM
Crossover
Non-crossover


Canonical NHEJ
Ku70/80
XRCC5
XRCC6
PRKDC
Ku heterodimers bind to ends
and recruit DNAPKcs
SETMAR
PARP3
RECQL
PNKP
MRE11A
DCLRE1C
APLF
End processing
G1
S
G2
M
RFC1
RFC2
RFC3
RFC4
RFC5
PCNA
XRCC4
NHEJ1
POLL
POLM
LIG4
Bridging, DNA synthesis
and ligation


A
HELLS
(Normal functions)
Development
- DNA methylation (de novo)
- Histone acetylation
- Histone methylation
- Heterochromatin
- Transcription silencing
TF coactivator
- E2F1,2,3-binding
- G1/S transition
- Proliferation
- Anti-senescence
DNA repair
- H2AX phosphorylation
- Chromatin remodelling
- Protein recruitment
- Improved DSBR
B
HELLS, ATAD2 and E2Fs drive cancer cell proliferation
ATAD2
RB
P
E2F1
E2F1
RB
P
CDK4
E2F3
HELLS
E2F1
ATAD2
FOXM1
Promoter
HELLS
E2F3
HELLS
E2F3
HELLS
HDAC1,2
DNMT1,3B
E2F1
CDC6
MLL1
MCM4
ProP16INK4
P16INK4
MYC
ATAD2
MYC
G1 S
G1 S

Supplementary Figure 1





Paper III

Is not included due to copyright

Paper IV

Is not included due to copyright