

# **Novel Functional Aspects of Topoisomerase Top1 and DNA Glycosylase Thp1 in the Maintenance of Genetic and Epigenetic Stability in Yeast**

**Inauguraldissertation**

zur  
Erlangung der Würde eines Doktors der Philosophie  
vorgelegt der  
Philosophisch-Naturwissenschaftlichen Fakultät  
der Universität Basel  
von

Claudia Krawczyk  
aus Herne, Deutschland

Basel, 2014

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Primo Schär (Fakultätsverantwortlicher und Dissertationsleiter)

Dr. Serge Boiteux (Korreferent)

Basel, den 16.09.2014

Prof. Dr. Jörg Schibler

Dekan der Philosophisch-Naturwissenschaftlichen Fakultät

## Acknowledgements

First of all, I would like to thank Primo Schär for giving me the opportunity to perform my PhD studies in his laboratory; for being a great supervisor, for fruitful discussions and for his great enthusiasm.

I further thank my PhD committee, Serge Boiteux and Marc Bühler, for their input, their critical evaluation of my work and for taking their time.

Moreover, I would like to thank all current and past members of the Schär lab for the nice working atmosphere. I really enjoyed the time! Special thanks goes to Olivier for his guidance and great discussions throughout my time in this lab, for letting me become independent, for critical reading of my thesis and for delicious fruits from his garden. My gratitude goes to Annika for having a great time in- and outside of the lab, and for her never-ending motivation. I thank Faiza for scientific discussions as well as the chit-chat at the coffee machine or at lunch. Thanks to my “devil” Emina who always provided me with chocolate, but – even more important – with lots of chatting and laughing during my writing period. I would also like to thank Angelika for all the time we spend as neighbors in the lab and for the recent jogging sessions. Thanks to Stefan, Melissa, Petar, William and Sarah for the uncountable, enjoyable lunches.

I would like to thank Katrin, Mirjam and Anja sharing many coffee breaks (without coffee) and discussions.

I would like to use the opportunity to also thank all my friends in Basel and at home – thanks to Nesy and Jeanette for your deep friendship!

My deepest gratitude goes to my family. To my parents, Angelika and Peter, for their life-long support and their never-ending belief in me. To Sebastian for his love, his support, his patience and so much fun!

# Table of Contents

Abbreviations.....	3
<b>1 Summary.....</b>	<b>5</b>
<b>2 Introduction.....</b>	<b>8</b>
<b>2.1 Maintenance of Genetic and Epigenetic Information.....</b>	<b>8</b>
2.1.1 Sources of Genome Instability.....	8
2.1.2 DNA Surveillance and Repair Mechanisms.....	12
2.1.3 Regulation of the Epigenome.....	19
<b>2.2 Functions and Processing of Impeded Replication Forks.....</b>	<b>21</b>
2.2.1 Programmed Replication Fork Barriers.....	21
2.2.2 The Ribosomal Replication Fork Barrier of Budding Yeast.....	22
2.2.3 Regulation of rDNA Recombination.....	24
<b>2.3 DNA Topoisomerases in the Removal of Torsional Stress.....</b>	<b>27</b>
2.3.1 DNA Topoisomerase Functions.....	27
2.3.2 Type IB Topoisomerases.....	29
2.3.3 Repair of Irreversible Top1 Covalent Complexes.....	32
<b>2.4 Uracil as a Source of Genome Instability.....</b>	<b>33</b>
2.4.1 Uracil in DNA – Origin and Consequences.....	33
2.4.2 Uracil DNA Glycosylases .....	34
2.4.3 Current Insights into the Functional Separation of Uracil DNA Glycosylases..	39
2.4.4 <i>S. pombe</i> as a Model Organism to Study Base Excision Repair and Chromatin Regulation.....	42
<b>3 Aims of the Thesis.....</b>	<b>45</b>
<b>4 Results.....</b>	<b>46</b>
<b>4.1 Reversible Top1 Cleavage Complexes are Stabilized Strand-Specifically at the     Ribosomal Replication Fork Barrier and Contribute to Ribosomal DNA Stability ...</b>	<b>46</b>
<b>4.2 Uracil Repair Causes DNA Glycosylase-Dependent Genome Instability.....</b>	<b>48</b>
<b>4.3 Additional Results.....</b>	<b>51</b>
4.3.1 Estrogen Receptor $\beta$ Regulates Epigenetic Patterns at Specific Genomic Loci Through Interaction with Thymine DNA Glycosylase.....	51
4.3.2 Physical Interactions between Murine TDG and TET1.....	54
4.3.3 DNA Ligase 4 Function in the rDNA of Budding Yeast.....	56
<b>5 Concluding Discussion and Outlook.....</b>	<b>61</b>

**6 References..... 66**

**7 Appendix**

- I. Reversible Top1 Cleavage Complexes are Stabilized Strand-Specifically at the Ribosomal Replication Fork Barrier and Contribute to Ribosomal DNA Stability
- II. Uracil Repair Causes DNA Glycosylase-Dependent Genome Instability
- III. Estrogen Receptor  $\beta$  Regulates Epigenetic Patterns at Specific Genomic Loci Through Interaction with Thymine DNA Glycosylase

## Abbreviations

<b>2D gel</b>	two-dimensional agarose gel electrophoresis
<b>5-FU</b>	5-fluorouracil
<b>5caC</b>	5-carboxylcytosine
<b>5fC</b>	5-formylcytosine
<b>5hmC</b>	5-hydroxymethylcytosine
<b>5mC</b>	5-methylcytosine
<b>AD</b>	Gal4 activation domain
<b>AID</b>	activation induced cytosine deaminase
<b>AP site</b>	apurinic/aprimidinic site
<b>ARS</b>	autonomously replicating sequence (replication origin)
<b>ATR</b>	ATM- and Rad3-related protein
<b>BD</b>	Gal4 binding domain
<b>BER</b>	base excision repair
<b>ChIP</b>	chromatin immunoprecipitation
<b>CPT</b>	camptothecin
<b>CSR</b>	class switch recombination
<b>dHJ</b>	double Holiday Junction
<b>DMP</b>	differentially methylated position
<b>dNMP</b>	deoxynucleoside monophosphate
<b>dNTP</b>	deoxynucleoside triphosphate
<b>DSB</b>	double-strand break
<b>dUTPase</b>	deoxyuridine 5'-triphosphatase
<b>E2</b>	17 $\beta$ -estradiol
<b>ER</b>	estrogen receptor
<b>ERC</b>	extrachromosomal ribosomal DNA circle
<b>eRFB</b>	ectopic ribosomal replication fork barrier
<b>ESC</b>	embryonic stem cell
<b>H3K9meX</b>	histone 3 lysine 9 modified with X methyl groups
<b>HAT</b>	histone acetyl transferase
<b>HDAC</b>	histone deacetylase

<b>HR</b>	homologous recombination
<b>IGS</b>	intergenic spacer
<b>MBD4</b>	methyl-binding domain glycosylase 4
<b>MEF</b>	mouse embryonic fibroblast
<b>MMR</b>	mismatch repair
<b>MUG</b>	mismatch uracil glycosylase
<b>NER</b>	nucleotide excision repair
<b>NHEJ</b>	non-homologous end-joining
<b>PCNA</b>	proliferating cell nuclear antigen
<b>rDNA</b>	ribosomal DNA
<b>RENT</b>	regulator of nucleolar silencing and telophase exit complex
<b>RF</b>	replication fork
<b>RFB</b>	replication fork barrier
<b>ROS</b>	reactive oxygen species
<b>RPA</b>	replication protein A
<b>RRBS</b>	reduced representation bisulfite sequencing
<b>rRFB</b>	ribosomal replication fork barrier
<b>rRNA</b>	ribosomal RNA
<b>RTS1</b>	replication termination site 1
<b>SCE</b>	sister chromatid exchange
<b>SHM</b>	somatic hypermutation
<b>SMUG1</b>	single-strand-selective mono-functional uracil DNA glycosylase
<b>SSB</b>	single-strand break
<b>SUMO</b>	small ubiquitin-like modifier
<b>TDG</b>	thymine DNA glycosylase
<b>Tdp1</b>	tyrosyl-DNA phosphodiesterase 1
<b>TET</b>	ten-eleven translocation dioxygenase
<b>TF</b>	transcription factor
<b>Top1cc</b>	Top1 cleavage complex
<b>UDG</b>	uracil DNA glycosylase
<b>UNG</b>	uracil-N glycosylase

# 1 Summary

The genetic information stored in the DNA of all cells is crucial for normal growth of uni- and multicellular organisms. The integrity of genomes is constantly threatened by DNA damage and genetic transactions that generate torsional stress and replication stalling. Maintaining DNA stability is thus essential for life and is assured by a variety of DNA surveillance and repair mechanisms.

During DNA replication, replication fork (RF) progression is frequently disturbed by obstacles that can be either accidental or programmed, as in the case of RF barriers (RFBs) found in many organisms. The well-studied polar ribosomal RFB (rRFB) of budding yeast stalls RFs upon sequence-specific binding of Fob1. The highly repetitive structure of the ribosomal DNA (rDNA) renders it prone to homologous recombination, which may destabilize the locus. Fob1 along with a number of additional enzymes ensures rDNA homeostasis, probably by favoring genetically silent recombination outcomes and by quickly recovering the normal repeat number upon accidental copy number changes. Fob1 also mediates anchoring of the rDNA repeats to the inner nuclear membrane, which restricts the structural flexibility of the locus. This, together with the replication- and transcription-associated unwinding of the DNA helix likely generates a great amount of torsional stress in the DNA. DNA relaxation is normally achieved by the action of DNA topoisomerases that cut the DNA by forming a covalent bond with it to allow unwinding of the helix before resealing the break. Indeed, topoisomerases are important for rDNA stability and topoisomerase 1 (Top1)-dependent nicks were shown to occur close to the rRFB. It remained, however, elusive how Top1 action is regulated in the rDNA. The first aim of my thesis was the identification of factors influencing Top1 nicking activity within the rDNA and particularly at the rRFB.

I found that unusually stable Top1-DNA complexes (Top1 cleavage complexes, Top1ccs) and, hence, DNA nicks, accumulate specifically at the rRFB. This accumulation requires Fob1 and the nucleolar protein Tof2, but is independent of RF stalling, suggesting that Fob1 and Tof2 position Top1 to and stabilize the Top1cc at the rRFB. Interestingly, Top1cc stabilization by Fob1 and Tof2 does not require the rDNA context, as Fob1-dependent Top1ccs also accumulated at an ectopically located rRFB that was neither recruited to the nucleus nor to the inner nuclear membrane. We also identified Top1 nicks to account for most of the DNA double-strand breaks (DSBs) previously described to arise in wild-type cells, pointing at an *in vitro* conversion of single-stranded Top1 nicks to DSBs. On the basis of these data, we propose a model in which Top1 is recruited to the rRFB, where it nicks the DNA forming a stable Top1cc intermediate to allow for continuous relaxation, thereby contributing to the genetic stabilization of the structurally complex locus. This stabilization is achieved by protein-protein interactions with Fob1 and Tof2 or by misalignment of the break end by the structure of the



complex. In the published manuscript we thus present first evidence for locus-specific regulation of Top1 catalytic actions.

Besides regulating DNA torsion and ensuring RF integrity, genome maintenance also involves the removal of irregular DNA bases. Uracil in DNA results from cytosine deamination or uracil misincorporation during DNA replication. While the former leads to G•U mismatches and, hence, C to T transition mutations if left unrepaired, the latter is non-mutagenic, but could affect transcription factor binding. Uracil DNA glycosylases (UDGs) excise uracils from DNA, thereby initiating a base excision repair (BER) process that restores the regular DNA sequence to counter these adverse effects. Notably, as BER generates DNA single-strand breaks as intermediates, the repair of regularly occurring uracils in DNA will have an impact on the helical structure of the DNA, similar to the action of topoisomerases. In mammals, four nuclear UDGs with partially overlapping functions are expressed. While UNG2 and SMUG1 were mainly associated with “classical” uracil excision repair, TDG and MBD4 appear to have only minor roles in this process, but instead are important for epigenetic regulation of gene expression. The functional separation of the single enzymes, however, is not fully understood. Thus, the second aim of my thesis was to further separate UDG functions using *S. pombe* as a model. *S. pombe* offers a UDG system of reduced complexity as it has only two, instead of four, of these enzymes, namely an UNG2 ortholog (Ung1) and a TDG ortholog (Thp1). In addition to the presence of only two UDGs, this organism has chromatin regulation mechanisms similar to those found in mammalian cells, while having no DNA methylation.

We used a genetic approach to functionally separate the two *S. pombe* UDGs Ung1 and Thp1. By studying the loss-of-function phenotypes of the two UDGs, we found that despite a dominant uracil excision activity of Ung1 in cell-free extracts, both Ung1 and Thp1 contribute to uracil removal and mutation avoidance *in vivo*. Interestingly, Thp1 expression mediates cytotoxicity during 5-fluorouracil (5-FU) exposure and in the presence of increased amounts of genomic uracil. In line with a toxic effect of Thp1-dependent repair, Thp1 overexpression increases DNA breakage and spontaneous mutation rates. Thp1 was previously shown to have high affinity to its product abasic site (AP-site) and we reason that the long-lived AP site generated by Thp1, but not by Ung1, causes cytotoxicity. Presence of AP sites could occasionally lead to DNA breakage and, consistent with an increased lifetime of AP sites in Thp1-initiated repair, most spontaneous mitotic recombination events measured in *S. pombe* cells depend on Thp1 but not on Ung1. Thus, Ung1 and Thp1 have overlapping functions in uracil removal, however, the qualitative repair outcome appears to be different. While Ung1-mediated BER is probably fast and mostly error-free, Thp1-dependent BER appears to be slow and error-prone. The fact that both UDGs co-evolved suggests that Thp1 fulfills additional cellular functions beyond classical DNA repair. Given the epigenetic role of mammalian TDG in regulating gene expression, I explored additional Thp1 functions in gene regulation by comparing genome-wide

RNA expression profiles. While this revealed no distinct pattern of dysregulation in Thp1-deficient cells, most of the analyzed genes tended to be less expressed in the absence of Thp1. Interestingly, Thp1-deficiency also increases the variability of gene expression between replicates. We therefore conclude that Thp1-dependent processes contribute to the maintenance of a transcriptionally active chromatin. Consequently, *S. pombe* could serve as a suitable model for studying the impact of UDGs on gene expression.

Taken together, in collaboration with several colleagues, my work provided insights into a novel regulatory aspect of Top1 function in the rDNA of budding yeast and into the function of Ung1- and Thp1-dependent uracil repair in *S. pombe*.

## **2 Introduction**

### **2.1 Maintenance of Genetic and Epigenetic Information**

All genetic information required for building up unicellular as well as multicellular organisms is encoded in the deoxyribonucleic acid (DNA). The genetic alphabet consists of four nucleobases: adenine (A), thymine (T), guanine (G) and cytosine (C). An additional layer of information is added by the covalent modification of the DNA bases and by the covalent attachment of chemical groups to histones around which the DNA is wrapped for compaction. This secondary layer is called the epigenetic code or epigenetic information, as it is partially inheritable and determines the transcriptional readout of the genome. Genetic and epigenetic information are constantly threatened by a variety of exogenous and endogenous factors, and in humans the alteration of both types of information is associated with organismal aging and age-related diseases such as cancer. Therefore, the understanding of molecular events leading to such cellular transformations and of the cellular defense mechanisms is crucial for the identification of potential drug targets.

#### **2.1.1 Sources of Genome Instability**

##### **Exogenous and Endogenous DNA Lesions**

Due to its chemical nature the DNA is intrinsically instable. Indeed, the most frequent lesion found in DNA, a nucleotide that has lost its base, is of endogenous origin, resulting from the spontaneous hydrolysis of the N-glycosidic bond linking the DNA base to the sugar phosphate backbone. These so-called abasic sites, or AP sites, that arise an estimated 9000 times per day in the diploid human genome, are devoid of coding information and are thus potentially mutagenic (Wilson & Kunkel, 2000; Kim & Wilson, 2012). During DNA replication incorrect deoxynucleotides might be incorporated into the DNA generating pre-mutagenic mispairs. Although the proofreading activity of DNA polymerases is efficient, residual mispairs are found that escaped this level of repair and need to be corrected by the post-replicative mismatch repair system. Spontaneous hydrolytic deamination of cytosine, adenine and guanine to uracil, hypoxanthine and xanthine/oxanine, respectively, represent additional causes of DNA mismatch formation (Schärer, 2003; Suzuki et al., 2000). During normal aerobic metabolism, reactive oxygen species (ROS) are formed, which can oxidize DNA bases in their close proximity. For instance, 8-oxoguanine and thymine glycol can be formed through ROS-mediated oxidation and they cause mutations by miscoding or interfere with DNA replication and RNA transcription (Schärer, 2003).

In addition to these highly abundant endogenous DNA lesions, alterations in the DNA are also caused by environmental factors, such as electromagnetic radiations (e.g. UV light, ionizing radiation) and genotoxic chemicals. The energy deposited by UV or ionizing radiation is sufficient to alter or break covalent bonds, creating a variety of DNA lesions including single-strand breaks (SSBs) and double-strand breaks (DSBs). These radiations also generate ROS, which may induce various DNA base damage as well as occasional DNA-DNA or DNA-protein crosslinks. UV exposure also dimerizes DNA bases, generating adducts like cyclobutan pyrimidine dimers and other bulky lesions that disturb the double-helical structure of the DNA, are miscoding and/or polymerase blocking (Friedberg, 2006; Schärer, 2003). DNA damage is also induced by a wide range of genotoxic agents through their potential to chemically modify DNA base or to induce DNA-DNA or DNA-protein crosslinks. While their cytotoxic and mutagenic nature is undesirable in healthy individuals, many such agents are used in cancer therapy due to the fact that their cytotoxic effects are mostly replication-mediated. One example is the uracil analog 5-fluorouracil (5-FU), which induces increased incorporation of uracil and its own metabolic derivatives into DNA and RNA. Base excision repair (BER) intermediates of these lesions were shown to interfere with replication fork (RF) progression, finally causing cell death (Longley et al., 2003).

### **Replication Fork Impediments**

The DNA of a eukaryote cell is replicated once per cell cycle during S-phase. In G1 phase, the pre-initiation complex is loaded onto multiple origins of replication to “license” them for replication. Replication is started bi-directionally with two RFs moving in opposite directions along the chromosome until they merge with another RF in a termination process. The replication machinery (replisome) consists of numerous highly coordinated protein factors that ensure smooth RF progression through the chromosome. Unwinding of the DNA double helix by the helicases separates the strands, which then serve as templates for DNA synthesis by DNA polymerases in a semi-conservative mode. While DNA polymerase  $\epsilon$  is responsible for the continuous replication of the leading strand, the lagging strand is synthesized discontinuously by DNA polymerase  $\delta$ . Lagging strand synthesis is primed every 100–200 bases from RNA:DNA hybrids, giving rise to Okazaki fragments that are joined by DNA Ligase I to form an uninterrupted DNA strand. The proliferation cell nuclear antigen (PCNA) is yet another factor that forms a sliding clamp around the replicated DNA, stabilizing the replisome and hence supporting RF progression (Chagin et al., 2010; Leman & Noguchi, 2013). Stretches of single-stranded DNA are exposed during replication, especially on the lagging strand template. These are covered by the single-strand binding protein A (replication protein A,

RPA), essentially to prevent the formation of secondary DNA structures, but also to shield the exposed bases from chemical attacks.

RFs are frequently interrupted in their progression by obstacles, such as DNA helicase- or polymerase-blocking lesions. Replication arrest may also be triggered by inhibitors of the replicative polymerases, or by hydroxyurea that depletes the nucleotide pool and thus the substrates for DNA synthesis. Some DNA bound proteins also block the replisome, acting as RF barriers (RFB). A number of programmed RFBs are found in the genome of various organisms. Defined by the sequence-specific binding of proteins mediating the RFB activity, they are discussed more deeply in chapter 2.2. Another source of replication instability is the interference between DNA replication and transcription, particularly if the processes are oriented in a head-to-head collision course (Helmrich et al., 2013; Lambert & Carr, 2013). For example tRNA genes that are highly transcribed throughout the cell cycle frequently pause the RF when transcription and replication proceed in opposite direction. Similarly, highly transcribed RNA polymerase II genes impede RF progression. RF blocks have been also observed in situation where transcription and replication progress in the same direction, probably reflecting problems to replicate across DNA/RNA hybrids (R-loops), paused RNA polymerases or regions with increased topological stress (Branzei & Foiani, 2010; Lambert & Carr, 2013). Indeed, increased R-loop formation by THO/TREX deletion induced the formation of recombination intermediates during replication (Gomez-González et al., 2011). Palindromic DNA sequences have the propensity to form secondary DNA structures that can interfere with replication, especially on the leading strand (Rosche et al., 1995). Moreover, there are genomic regions particularly prone to replication stress. These are hotspots of DNA breakage and recombination and are known as chromosomal fragile sites (replication slow zones in yeast). The underlying causes are just about to be unraveled, but appear to be multifactorial. Most of these sites show slow-down or pausing of the RF, and some were also related to high transcription activities, secondary DNA structures and hypo-acetylated histones (Hashash et al, 2011; 2012; Lambert & Carr, 2013).

All these contexts interfering with RF progression require replisome stabilization (“stalled RF”) to prevent premature disassembly of the complex and to allow for resumption of DNA replication (“RF collapse”), thereby protecting from the initiation of unscheduled recombination events. To circumvent prolonged stalling, replicative repair mechanisms allowing for bypass of certain lesions are in place. These include error-prone translesion synthesis polymerases and homologous recombination (HR)-dependent error-free bypass pathways. In addition, the intra-S-phase checkpoint is important to maintain genome stability when the cell experiences increased replication stress. It blocks or slows down cell cycle progression, stabilizes the replisome to prevent RF collapse and is also implicated in the initiation of DNA repair processes and the restart of stalled RFs. A non-functional checkpoint in challenged cells was indeed shown to provoke replisome disassembly as well

as the formation of reversed RFs (Cotta-Ramusino et al., 2005; Hu et al., 2012; Lopes et al., 2001; Sogo et al., 2002; Tercero & Diffley, 2001). Single-stranded DNA stretches are formed at stalled RFs due to uncoupling of either helicase and polymerase activities or of leading and lagging strand DNA synthesis. In budding yeast, ATR (ATM- and Rad3-related protein) is thought to sense this single-stranded DNA and it cooperates with several factors to phosphorylate the essential checkpoint kinase Rad53, which in turn is required for full checkpoint activation (Cobb et al., 2004). The replisome-associated factors Mrc1 and Tof1 from budding yeast prevent uncoupling of the replisome and the DNA polymerase by inducing fork arrest, whereas Rrm3 facilitates RF progression by removing DNA-bound proteins in front of the replisome. Absence of Rrm3 leads to highly increased numbers of RFBs, which become recombination hotspots (Branzei & Foiani, 2010).

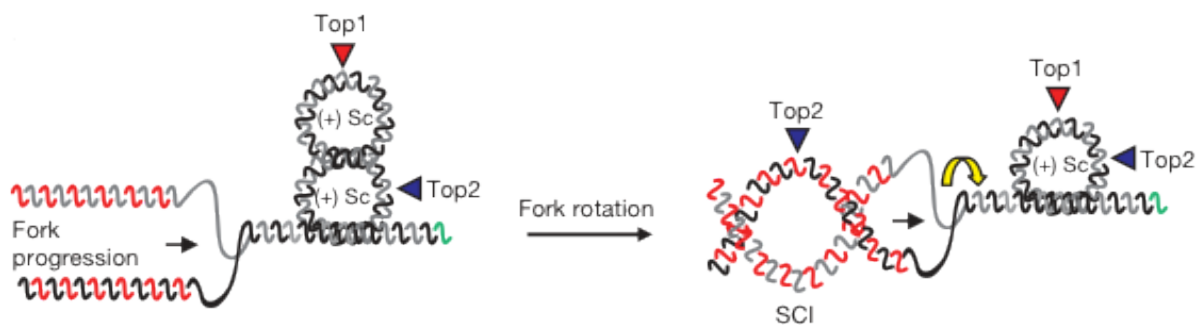
### **DNA torsional stress**

During DNA replication the unwinding of the two DNA strands by the dedicated helicases induces helical tension around the replication bubble that cannot simply diffuse, due to the rigidity of the macromolecular structure of the chromosomes. Chromosomes are divided into large domains with fixed boundaries that preclude rotation. RF movement thus leads to over-twisting ahead of the RF and to under-twisting in its wake (Figure 1), this results in a local over- and underwinding of the DNA duplex known as positive and negative supercoiling, respectively. In order to relieve the torsional stress ahead of the RF, the replisome might also move along the twist of the DNA helix, resulting in intertwined daughter strands (“precatenates”, Figure 1). To ensure continuous DNA replication, the cell expresses special nucleases called topoisomerases that break, unwind and rejoin these precatenates (chapter 2.3.1) (Postow et al., 2001; Teves & Henikoff, 2014b). In addition to the torsional stress arising through the normal RF progression, torsion-driven non-canonical DNA folding such as the formation of cruciform DNA structures might further impede replication (Branzei & Foiani, 2010; Gilbert & Allan, 2014). In addition, the reported association of expressed genes with the nuclear pore complex in yeast might impose torsional constraints (Sood & Brickner, 2014). Similarly, anchoring of the ribosomal RFB (rRFB) to the nuclear membrane prevents diffusion of DNA torsion in the highly transcribed ribosomal DNA (rDNA) of budding yeast. In line with a large demand of releasing topological stress in this genomic region, topoisomerases are required for rDNA stability (Bermejo et al., 2012).

The build-up of helical tension likewise applies to the process of transcription where RNA polymerases move along the DNA helix. Consequently, local supercoiling domains were observed at sites of active transcription that were boosted by topoisomerase inhibition. The relief of torsion is required throughout transcription and becomes particularly important when a transcribed gene is

simultaneously undergoing replication. Slowing-down and pausing of RF progression has been observed independent of the relative orientation of the two processes. In the case of converging transcription and replication machineries, this is well explained by a substantial increase of positive supercoiling in the DNA located between the replisome and the progressing RNA polymerase complex (Branzei & Foiani, 2010).

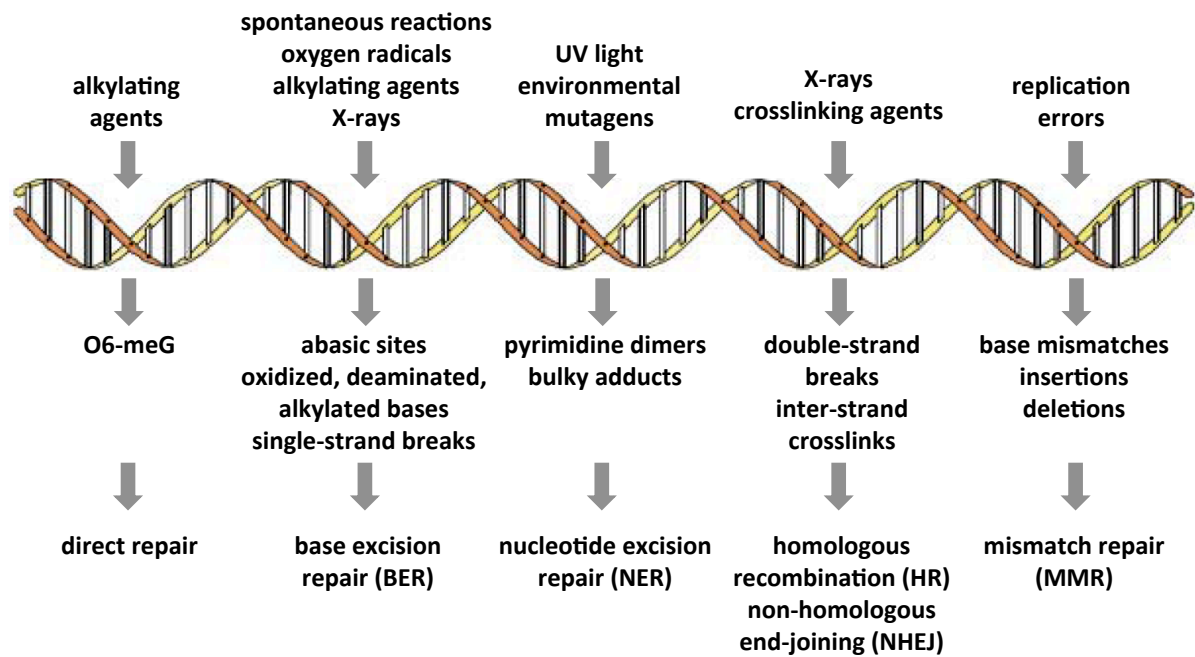
It was proposed that increased torsion of the DNA is not just a by-product of transcription and replication but might be important for the establishment of chromatin states that permit activation of gene promoters (Gilbert & Allan, 2014; Ma et al., 2013). As a proof of principle, an inducible gene promoter was shown to be activated by independent transcription of a region 1.2 kb away from the promoter (Kouzine et al., 2008). In a genome-wide approach it was further shown that under-wound chromosome regions are less compact than over-wound domains and would thus favor gene expression. Large supercoiled domains could therefore constitute an elegant mechanism to facilitate regulation of gene clusters (Naughton et al., 2013). At the nucleosome level, increased helical torsion induced *in vitro* and by topoisomerase inhibition *in vivo* triggers increased H2A/H2B histone exchange and nucleosome turnover, respectively (Sheinin et al., 2013; Teves & Henikoff, 2014a). Moreover, the same as DNA-bound proteins alter DNA topology, the local DNA topology is conversely important for binding regulatory proteins to some promoter regions (Jagelská et al., 2008; Kouzine et al., 2008).



**Figure 1. Increased DNA torsion at progressing replication forks.** The DNA helix ahead of the replication fork is overwound and forms positive supercoiling (Sc). In the wake of the fork, rotation of the replisome along the DNA helix may result in interdigitated sister chromatids (SCI), or “precatenates”. Positive supercoiling is released mainly by Top1 and partially by Top2, whereas precatenates are removed only by Top2. Figure from (Kegel et al., 2011).

### 2.1.2 DNA Surveillance and Repair Mechanisms

To prevent the genome from deleterious events caused by the different DNA lesions, a number of highly entangled DNA repair mechanisms have evolved. Figure 2 summarizes common DNA lesions and preferred pathways used for their repair.



**Figure 2. Origin and repair of DNA lesions.** Overview of endogenous and exogenous DNA damaging agents, the resulting DNA lesions and repair pathways involved in their repair. Figure adapted from (Schärer, 2003).

### Direct Reversal of DNA Lesions

A few enzymes are capable of directly reversing DNA damage, one example being the photolyase that is expressed in many prokaryotes and eukaryotes, but is absent from mammals. Upon light-mediated activation, photolyase is capable of recognizing and reversing UV-induced pyrimidine dimers. Another prominent example is the O<sup>6</sup>-methylguanine DNA methyltransferase (MGMT), also found in mammalian cells, that directly removes the alkyl group of O<sup>6</sup>-alkylguanine to restore the guanine. However, MGMT is a suicide enzyme as binding of the methyl group is irreversible and inactivates MGMT (Schärer, 2003).

### Nucleotide Excision Repair

Mammalian cells lack the photolyase, but instead utilize nucleotide excision repair (NER) to counter UV-induced DNA damage. NER acts on a broad range of bulky lesions distorting the helical structure of the DNA. Two modes of NER are distinguished: global genome repair (GGR) and transcription-coupled repair (TCR), the latter being responsible for timely repair of lesions interfering with active transcription. Both sub-pathways rely on the regulated assembly and action of approximately 30 proteins. GGR is initiated by binding of XPC-RAD23B and UV-damage DNA-binding protein (UV-DDB) to the lesion, whereas TCR is triggered by stalling of the RNA polymerase in conjunction with a set of



other NER enzymes (CSA, CSB, XAB2). Following damage recognition, the two sub-pathways prosecute a common “core” NER mechanism. While the DNA is stabilized by the single-strand binding protein RPA and a subset of Xeroderma pigmentosum complementation group proteins (XPA, XPB, XPD), the damaged DNA strand is incised 3' and 5' of the lesion by the action of the structure-specific endonuclease ERCC1-XPF, resulting in the excision of 24 to 32 nucleotides. After DNA re-synthesis by DNA polymerase  $\delta$ ,  $\epsilon$  or  $\kappa$ , the DNA is ligated by either DNA ligase III together with XRCC1 or by DNA ligase I. To facilitate NER, the chromatin was shown to be rearranged during the repair process by DNA remodeling factors (Kamileri et al., 2012; Schärer, 2013).

Defects in NER are associated with complex and heterogeneous genetic disorders. Defective GGR manifests as Xeroderma pigmentosum, which predisposes patients with a 2000-fold increased risk to skin cancer in response to sunlight. The Cockayne syndrome is caused by defects in TRC and patients suffer from pre-mature aging and from developmental and neurological defects. The pleiotropic nature of NER-related defects arise partially from the different proteins that are affected, but they might also reflect functions of NER proteins that are not directly involved in DNA repair (Kamileri et al., 2012).

### **Mismatch Repair**

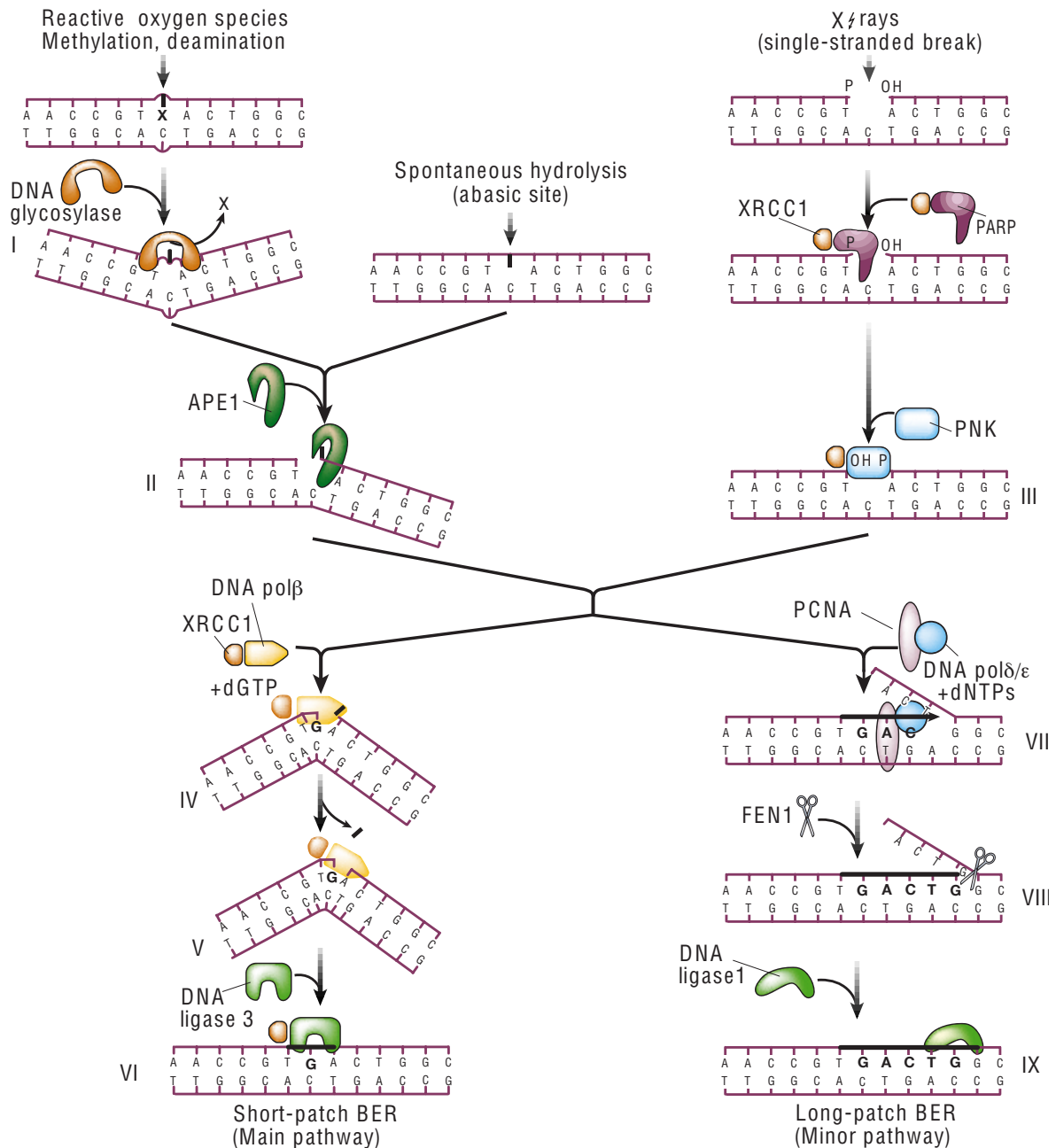
The post-replicative DNA mismatch repair (MMR) pathway removes mismatched DNA bases and small insertions/deletions that escaped the proofreading function of DNA polymerases during DNA replication. To be efficient in mutation avoidance, the MMR system must be able to specifically detect the mismatched base located in the newly replicated DNA strand. In *E. coli*, the newly synthesized DNA strand is identified by the lack of *dam* methylation directly after replication, giving rise to transiently hemi-methylated DNA. Formation of a MutS-MutL recognition complex on MMR-relevant DNA lesions activates the endonuclease activity of MutH that is specifically bound to hemi-methylated-sites and that mediates strand-specific incision during MMR. In contrast, the DNA strand recognition mechanism during MMR is not definitively resolved for many other bacteria and for eukaryotes. However, as MMR is induced by the presence of DNA nicks *in vitro*, it was speculated that the DNA interceptions between Okazaki fragments on the lagging strand could constitute a signal for MMR direction. Alternatively, interactions with replication factor C (RFC) and PCNA might assemble the nuclease PMS2 in a configuration that mediates specific incision of the newly replicated DNA strand. Mismatch-recognition is accomplished by the heterodimers of hMSH2-hMSH6 (hMutS $\alpha$ ) and hMSH2-hMSH3 (hMutS $\beta$ ). While hMutS $\alpha$  is important for recognition of base mismatches and 1 to 2 base pair insertions/deletions, hMutS $\beta$  preferentially targets larger insertion/deletions or loops. The heterodimer hMLH1-PMS2 (hMUTL $\alpha$ ) is then thought to induce replication-dependent nicking

(see above), followed by 5' to 3' excision mediated by the exonuclease EXO1. Following damage excision, the gap is filled by a replicative DNA polymerase and ligated by DNA ligase I to complete repair. As expected, deletion of MMR components highly increases spontaneous mutation rates. Defective MMR proteins characteristically cause microsatellite instability and are associated with an increased risk of hereditary nonpolyposis colorectal cancer (HNPCC) and sporadic microsatellite unstable cancers (Jiricny, 2006; G.-M. Li, 2008).

### **Base Excision Repair**

Base excision repair (BER) is dedicated to the removal of a variety of DNA base lesions resulting from oxidation, deamination, alkylation and base loss. By correcting base lesions prior to DNA replication, BER counteracts mutations (Kim & Wilson, 2012). BER includes 5 consecutive steps: (i) removal of the damaged base by a damage-specific DNA glycosylase, (ii) incision of the resulting AP site, (iii) processing of the DNA ends, (iv) gap filling by polymerases and (v) DNA ligation (Figure 3).

BER is initiated by variable damage-specific DNA glycosylases of which 11 are expressed in human cells. They use different mechanisms to scan the DNA for damage and catalyze base removal by cleaving the N-glycosidic bond that links the base to the sugar moiety of the deoxynucleotide. Mono-functional DNA glycosylases only catalyze base excision, whereas bi-functional DNA glycosylases possess an additional AP lyase activity that cleaves the DNA backbone 3' of the lesion by a  $\beta$ -elimination resulting in a 3'-unsaturated aldehyde, which can be further converted into a 3'-phosphate ( $\text{PO}_4$ ) by  $\delta$ -elimination. Both ends are refractory to polymerization and require termini processing before proceeding with BER (Jacobs & Schar, 2012; Kim & Wilson, 2012). Generally, glycosylases have a high affinity to their product AP site, making AP site dissociation a rate-limiting step of BER. Due to the mutagenic capacity of the abasic site, this slow turn-over might protect the AP site until downstream factors of BER are available (Jacobs & Schar, 2012). It was suggested that glycosylases scan the DNA for aberrant bases by non-specific DNA interactions. Crystal structures of different DNA glycosylases with their substrates revealed that upon interaction with a damaged base, the DNA helix is distorted and the damaged base is flipped into the catalytic pocket of the glycosylase. This increases the interaction surface between the base and the pocket, enabling the detection and verifications of even small base alterations. Substrate specificity is mediated by selective interactions between the substrate base and the catalytic pocket and by steric exclusion. The N-glycosidic bond is then cleaved by a nucleophilic attack on the  $\text{C}^1$  of the deoxyribose (Brooks et al., 2013; Jacobs & Schar, 2012).



**Figure 3. Core pathway of mammalian short- and long-patch base excision repair.** (I) Damaged DNA bases are excised by damage-specific DNA glycosylases to create an AP-site that might also occur by spontaneous hydrolytic deamination. (II) The AP endonuclease APE1 incises the DNA backbone 5' of the lesion. (III) Alternatively, SSBs are detected by Poly(ADP-ribose) polymerase (PARP) that recruits XRCC1. Processing of blocked ends resulting from a SSB might be achieved by polynucleotide kinase (PNK). (IV-VI) During short-patch BER, the gap is filled by DNA pol  $\beta$ , which also trims the blocking 5' end, and ligation is performed by DNA ligase III (LIGIII $\alpha$ ) together with the scaffold protein XRCC1. (VII-IX). In long-patch BER (LP-BER), a DNA polymerase synthesizes 2 to 12 nucleotides and simultaneously displaces the 5' end to form a flap that is removed by FEN1. The nick is ligated by DNA ligase I (LIGI). Figure from (Hoeijmakers, 2001).

Upon base release, the phosphodiester backbone is incised 5' of the AP site by an AP endonuclease or 3' of the AP site by the intrinsic AP lyase activity of bi-functional DNA glycosylases. The major human AP endonuclease APE1 represents 95% of the cell's AP endonuclease activity (Kim & Wilson,

2012). Depending on the enzyme used for AP site incision, the 3' and 5' ends are not suitable for direct action of a polymerase and/or DNA ligase. Those enzymes require 3'-OH and a 5'-PO<sub>3</sub> ends and non-conventional (blocking) ends, such as 5'-dRP, 5'-OH or 3'-PO<sub>4</sub>, require termini cleanup by an AP endonuclease, certain DNA polymerases ( $\beta$ ,  $\lambda$ ,  $\iota$ ) or the bi-functional polynucleotide 5'-kinase/3'-phosphatase (PNKP) (Kim & Wilson, 2012). After processing of the ends, two BER sub-pathways are distinguished. In short-patch BER (SP-BER), DNA polymerase  $\beta$  fills in the gap with the correct base before it is ligated by DNA ligase I or III in conjunction with the scaffold protein XRCC1. During long-patch BER (LP-BER), DNA polymerase  $\beta$ ,  $\delta$  or  $\epsilon$  cooperate with PCNA to replicate a stretch of 2 to 12 nucleotides while displacing the 5' DNA end. Finally, the emerging DNA flap is removed by the structure-specific nuclease FEN1 and the nick is sealed by DNA ligase I.

Most of the damaged bases ( $\approx 80\%$ ) are repaired by SP-BER (Dianov et al., 2000) and it is likely that pathway choice is determined by the DNA ends present after incision of the backbone, which in turn depends on the initiating DNA glycosylase. If a blocking 5'-end is refractory to repair, LP-BER might be chosen to avoid problems during ligation (Sung & Demple, 2006). SP-BER enzymes are also used for SSB repair (SSBR). Among the first enzymes binding to SSBs is the nick sensor poly(ADP)ribose polymerase 1 (PARP1). PARP1 recruits XRCC1, a scaffold protein without a known enzymatic function, and is thought to stabilize and protect the nicks until BER enzymes take over. PARP1 dissociates from the break upon auto-poly(ADP)ribosylation and is recycled for continuous SSB detection (Caldecott, 2008).

The importance of BER is underpinned by the high conservation of enzymes involved in the repair of base lesions and by the requirement of several BER enzymes for viability. While deletion of DNA glycosylases lack drastic phenotypes in embryogenesis (apart from TDG), probably reflecting overlapping functions, knocking-out of unique downstream BER factors is incompatible with embryogenesis (Cortázar et al., 2011; Cortellino et al., 2011; Robertson et al., 2009). To maintain genome stability, tight regulation of BER proteins and proper handing over of repair intermediates is essential. Regulation of BER is achieved by protein-protein interactions and by various post-translational modifications including phosphorylation, SUMOylation and ubiquitination, which can alter the protein's stability, activity, subcellular location or turnover. Due to the constantly high levels of base damage and AP sites in a cell, most BER proteins are always present and additionally adjusted upon DNA damage (Almeida & Sobol, 2007; Dianov & Hübscher, 2013).

BER activity is also an important parameter during cancer treatment as it determines the response efficiency to common anti-cancer drugs like chemotherapeutic agents and radiotherapy. Cellular down-regulation or overexpression of many BER enzymes results in enhanced sensitivities to anti-cancer drugs. Thus, inhibiting certain BER enzymes alongside with conventional cancer therapies is a promising possibility to increase therapeutic effects (Kim & Wilson, 2012).

## **DNA double-strand break repair**

DSBs are among the most dangerous DNA lesions since they can lead to chromosome alterations and cell death if left unrepaired and they lack an intact complementary strand for repair. Two repair pathways DSB repair pathways exist: homologous recombination (HR) and non-homologous end-joining (NHEJ). Quantitative utilization of both pathways varies among species with NHEJ being the major DSB repair pathway in mammals while yeast predominantly performs HR.

### *Homologous Recombination*

Repair by HR relies on the homologous sequence residing on either the sister chromatid, the homologous chromosome or elsewhere in the genome, making HR mostly error-free, though it might result in loss of heterozygosity (Hakem, 2008). The first step of HR is the 5' to 3' resection of the DNA ends by exonucleases, such as the MRN (MRE11-RAD50-NBS1) complex or CtIP, to create single-stranded DNA of up to several hundred nucleotides. These overhangs are first bound by the single-strand binding protein RPA to prevent the formation of secondary structures. In a next step, RPA is exchanged for RAD51 that forms a nucleoprotein filament (presynaptic filaments) required for both homology search and DNA strand exchange. RAD51 loading necessitates the action of various mediator proteins, most importantly BRCA2 in mammals and Rad52 in yeast. The RAD51 nucleoprotein filament then mediates homology search and invades the donor DNA duplex to form a displacement loop (D-loop). After RAD51 dissociation, the 3' end is used to prime DNA synthesis along the template, thereby extending the D-loop. From this step, several routes of HR emerge to complete DSB repair. In the DSB repair (DSBR) sub-pathway, the second DNA end invades the template, generating a double Holiday Junction (dHJ). Resolvases cleave these dHJs in a manner that gives rise to either non-crossover or crossover products. In the synthesis-dependent strand-annealing (SDSA) route, the newly synthesized DNA re-anneals to the original strand, consequently resulting in non-crossover products. The break-induced repair (BIR) route was suggested as a possibility to cope with DSBs with only one end available for resection, as it is the case at degraded telomeres and, potentially, at collapsed RFs. In this HR mode, DNA synthesis is pursued until it reaches the chromosome end, thereby generating loss of heterozygosity. In contrast to the above-mentioned sub-pathways, single-strand annealing (SSA) does not involve strand invasion and is thus independent of RAD51. SSA becomes relevant in repetitive regions, in which homology is provided by adjacent repeats. Following resection of the DNA ends, the single-stranded ends anneal at homologous sites and non-homologous ends are removed. Repair is completed by DNA polymerase-mediated gap filling and ligation. Mutation in HR genes are related to genome instability and cancer predisposition in humans. For instance, the mediators of RAD51 nucleoprotein filament formation

BRCA2 and BRCA1 are known tumor suppressor genes and mutated proteins are associated with increased risks of breast and ovarian cancers (Heyer et al., 2010; Krejci et al., 2012).

### *Non-Homologous End-Joining*

NHEJ essentially involves direct ligation of DSB ends. Though it is frequently considered to be error-prone, this depends on the nature of the broken DNA ends and on the applied sub-pathway. Besides being important during DSB repair, NHEJ is also implicated in the physiological process of V(D)J and class switch recombination (CSR) at immunoglobulin loci in B- and T-lymphocytes of the immune system. Canonical NHEJ is initiated by binding of the heterodimer KU80-KU70 to the DNA ends and subsequent recruitment of DNA PKcs. These proteins seem to stabilize the DSB by preventing resection and holding ends together. If necessary, the DSB ends are processed before ligation by DNA ligase IV along with XRCC4. Alternative NHEJ mechanisms have been suggested that are independent of the KU proteins but instead involve resection over a few base-pairs to reveal micro-homologies that are used for annealing before gap filling and ligation take place. Detailed mechanistic dissection of canonical and non-canonical pathways, however, remains difficult (Bétermier et al., 2014; Pardo et al., 2009).

### **2.1.3 Regulation of the Epigenome**

The additional layer of genetic information added by chemical modifications of histones and DNA bases, often called the “epigenetic code” or the “epigenome”, has a high impact on gene expression. The epigenome is shaped by developmental and environmental factors. During embryogenesis, cells with the same DNA sequence differentiate into various cell types without altering the genetic code. This includes gross changes in gene expression patterns accompanied by huge remodeling of the epigenome that is important to stabilize the newly established expression patterns.

A major contribution to gene regulation comes from chromatin organization. The first step of DNA compaction is the wrapping of DNA around an octamer of two H3-H4 and H2A-H2B histone protein dimers, thereby generating a nucleosome, followed by further condensation into higher order structures. For active gene transcription, the DNA needs to be accessible and, consequently, most active genes are located in the open chromatin, also called euchromatin. In contrast, the highly compacted heterochromatin is largely devoid of bulk transcription. The protruding N-terminal tails of histones are decorated by a variety of post-translational modifications including acetylation, methylation, phosphorylation and poly(ADP)ribosylation, all playing important roles in regulating nucleosomal compaction. Acetylation of lysine residues, for instance, neutralizes the positive charge

of histone proteins, thereby weakening the association between DNA and histones and decompacting the chromatin structure. Indeed, acetylated histones are generally associated with active transcription. Regulation of this mark is achieved by histone acetyl transferases (HATs) and histone deacetylases (HDACs) (Bannister & Kouzarides, 2011; Kouzarides, 2007). Another common modification, the mono-, di- and tri-methylation of basic amino acids of the histone tail alters transcriptional activity by regulating the binding of chromatin modifiers and by interacting with transcription initiation/elongation factors. Additionally, histone modifications regulate the binding of cell line-specific transcription factors (TF) that, in turn, control the accessibility of the chromatin for the transcription machinery (Greer & Shi, 2012).

DNA bases are also subject of covalent modifications, the most prominent is the methylation at the C<sup>5</sup> of cytosine. In mammals, cytosine methylation mainly occurs in CpG di-nucleotides. Regions rich in CpG di-nucleotides are called CpG islands, however, these regions exhibit low methylation levels. Interestingly, the effect of cytosine methylation on gene expression depends on its location. While 5-methylcytosines (5mC) in gene promoters or enhancers inhibits transcription, 5mCs in the gene body show no such correlation but were implicated in alternative splicing. CpG methylation additionally impacts gene expression by regulating TF binding. Cell-type specific patterns of DNA methylation are established during cell lineage commitment to determine long-time regulation of pluripotency and developmental genes. In somatic cells, the majority of CpG islands at TSSs are unmethylated. Methylation at these sites is only found at genes that require long-term silencing, such as those located on the inactive X-chromosome or those required to maintain pluripotency (Jones, 2012). Short-term regulation of gene expression is instead achieved by other means, e.g. by changing the chromatin environment by the polycomb group (PcG) or trithorax group (TrxG) proteins. While PcG proteins are involved in gene repression in part by histone 3 lysine 27 tri-methylation (H3K27me3), Trx proteins activate gene transcription by mediating histone 3 lysine 4 tri-methylation (H3K4me3) or by chromatin remodeling (Chen & Dent, 2013).

CpG methylation is established by the *de novo* DNA methyltransferases DNMT3a and DNMT3b that use S-adenosyl-L-methionine (SAM) as the methyl group donor. Maintenance of the set DNA methylation pattern is then achieved by the action of the maintenance methyltransferase DNMT1, however, DNMT3a and DNMT3b also contribute to this activity. DNMT1 preferentially recognizes hemi-methylated DNA and uses it as a template for copying the methylation pattern to the newly synthesized DNA strand (Chen & Riggs, 2011). DNA demethylation might occur passively during repeated cycles of DNA replication in the absence of DNA methylation maintenance activities, or by active demethylation mechanisms, a concept that gained high interest during the recent years. Although direct removal of 5mC was previously suggested, it is now favored that 5mCs are modified before being excised by DNA glycosylases in a BER process. Such modifications include the stepwise

oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by the ten-eleven translocation (TET) dioxygenases, the latter two being suitable substrates for the thymine DNA glycosylase (TDG) (Kohli & Zhang, 2013). Alternatively, the cytosine deaminases AID and APOBEC1 could deaminate 5mC to thymine. The resulting T•G mismatch is a suitable substrate for TDG and the methyl-binding domain glycosylase 4 (MBD4), however, this path of active DNA demethylation is under debate (Nabel et al., 2012).

## 2.2 Functions and Processing of Impeded Replication Forks

### 2.2.1 Programmed Replication Fork Barriers

Programmed replication fork barriers (RFBs) are found in most organisms. They are defined by a specific DNA sequence that mediates binding of a blocking protein. Most scheduled RFBs are of polar nature and block RFs approaching from one side while being permissive to RFs approaching from the other side. In *E. coli*, the polar RFB “Ter” forces replication through a defined direction with most genes transcribed co-directionally relative to replication. Thereby, Ter is thought to prevent head-on collisions of the RF with the transcription machinery. The blocking activity requires the binding of the Tus protein to Ter and a direct interaction between Tus and the helicase was suggested to define barrier polarity (Kaplan & Bastia, 2009).

Two programmed RFBs (replication termination site 1, RTS1; *mat1* pausing site 1, MPS1) play an essential role in mating-type switching in *S. pombe* (Arcangioli & Klar, 1991; Dalgaard & Klar, 2000). RTS1, when bound by the replication termination factor 1 (Rtf1), prevents RFs from entering the mating-type locus *mat1* and most importantly from replicating across the MSP1 (Eydmann et al., 2008). This allows the opposite fork to stall at the MPS1. At this site an imprint is formed on the lagging strand, which might be a SSB or a ribonucleotide, however, its exact nature remains elusive. In the next generation, the cell that inherited the imprint undergoes mating-type switching. This is thought to involve RF collapse and subsequent HR-dependent repair during which the information in the *mat1* locus is exchanged with the information of the opposite mating-type from a silent donor cassette (Egel, 2005; Lambert & Carr, 2013; Leman & Noguchi, 2013). RTS1 and MPS1 activities require the RF protection subunits Swi1 (Tof1 in *S. cerevisiae*, TIMELESS in *H. sapiens*) and Swi3 (Csm3 in *S. cerevisiae*, TIPIN in *H. sapiens*) (Dalgaard & Klar, 2000). Placing the RTS1 at an ectopic location resulted in increased recombination events that were further elevated by deleting the RecQ helicase *rqh1*<sup>+</sup> gene (Ahn et al., 2005; Lambert et al., 2005).

The well-studied ribosomal RFB (rRFB) found in eukaryotic genomes is located in the rDNA consisting of numerous identical repeats that cluster in one or more genomic regions. The rDNA contains the



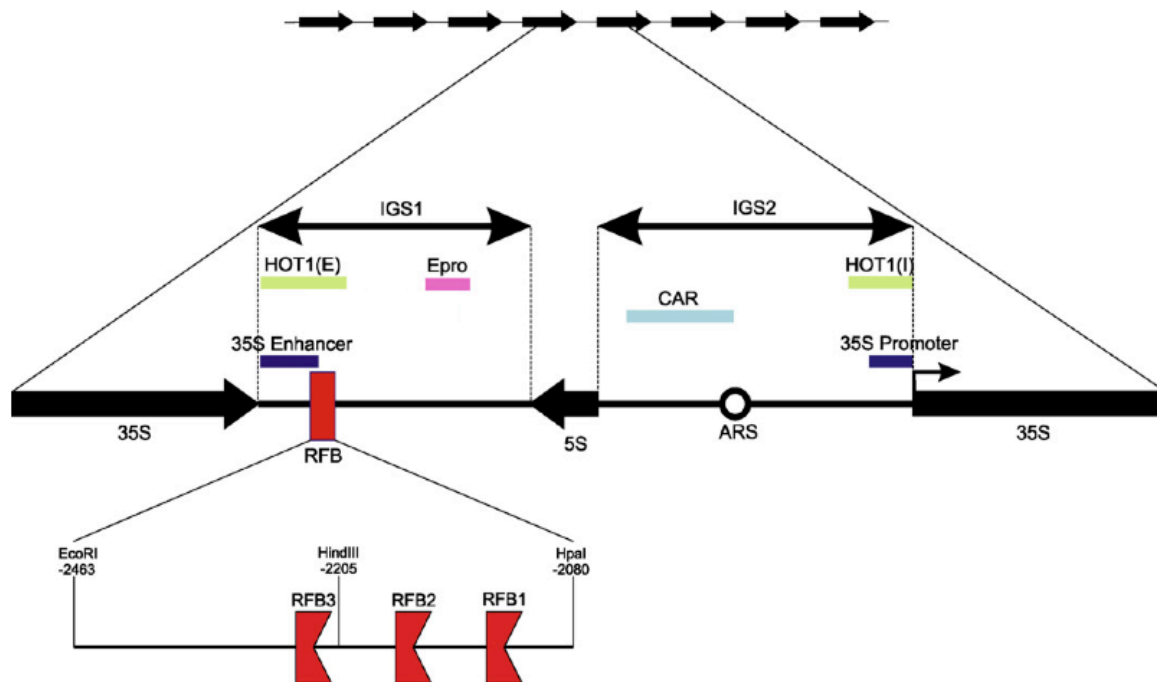
ribosomal RNA (rRNA) genes coding for the structural components of the ribosomes. These genes are highly transcribed and produce approximately 80% of all cellular RNA (Kobayashi, 2011; Mirkin & Mirkin, 2007). The rRFB is found at the 3' end of an rRNA gene and thus supposedly ensures that replication and transcription progress co-directionally to prevent head-on clashes. RF block is mediated by a sequence-specific binding of a blocking protein (Mirkin & Mirkin, 2007), but might, similarly to the situation at the Tus-Ter site in *E. coli*, involve interactions with the replicative helicase (Biswas & Bastia, 2008; Mohanty et al, 2006). The repetitive nature of the rDNA makes it highly prone to homologous recombination with the possible consequence of loss or gain of rDNA repeat units. However, maintenance mechanisms seem to counteract such events to ensure a stable copy number. Indeed, decreased copy numbers, which were induced by mutation of a DNA polymerase I subunit, were readily reversed upon reintroduction of the intact enzyme in wild-type yeast cells (Kobayashi et al., 1998).

While most characterized rRFBs are polar, the human rRFB blocks RFs approaching from either direction. The polar mouse rRFB, which serves also as the transcription termination site, is bound by the transcription termination factor 1 (TTF-1) that prevents progression of the replicative helicase in a unidirectional manner (Gerber et al., 1997; Putter & Grummt, 2002). In fission yeast, three RFBs ensure RF stalling at the 3' end of the rRNA genes. RFB1 is met first by the progressing RF and it constitutes the major barrier among the three RFBs. The blocking protein Sap1 is essential, but this may reflect additional functions in chromatin organization (de Lahondès et al., 2003; Krings & Bastia, 2005; Mejía-Ramírez et al., 2005). In contrast, stalling activity at RFB2 and RFB3 depends on Reb1 binding that, similar to the mouse TTF-1, serves not only as a replication block but also as a transcription terminator (Sánchez-Gorostiaga et al., 2004). Like at the RTS1, Reb1-mediated blocking is functional only in the presence of Swi1 and Swi3 (Krings & Bastia, 2004). Budding yeast has the most intensively studied rRFB. As I used it as a model for programmed RF stalling in eukaryotes, I will discuss it in detail in the following chapter.

## **2.2.2 The Ribosomal Replication Fork Barrier of Budding Yeast**

The rDNA in *S. cerevisiae* forms an array of 150 to 200 identical repeats of 9.1 kb length on chromosome XII (Figure 4). Each unit comprises two transcribed rRNA genes, 35S and 5S, that are transcribed by RNA polymerase I and III, respectively, separated by two intergenic spacers (IGSs). Replication is initiated bi-directionally from the origin of replication (autonomously replicating sequence, ARS) located in the IGS2. By this setting, both RFs progress co-directionally with rRNA transcription. At the unidirectional rRFB located in the IGS2 RFs approaching from the 5S are blocked from entering the 35S gene, thereby avoiding head-on collisions with the transcription machinery. It

is thought that the arrested RF is stabilized until it converges with another RF to terminate replication. The rRFB consists of three pausing sites (RFB1, RFB2, RFB3) with RFB1 as the major RFB. Together, they block more than 90% of the arriving RFs (Brewer et al., 1992; Kobayashi, 2003; Sogo et al., 2002). A prerequisite for rRFB activity is binding of the fork blocking protein Fob1 to a core rRFB sequence of about 100 base pairs (Calzada et al., 2005; Kobayashi & Horiuchi, 1996). It was suggested that the DNA is wrapped around Fob1 in a nucleosome-like structure (Kobayashi, 2003).



**Figure 4. Organization of the rDNA units in *S. cerevisiae*.** The rDNA locus of budding yeast consists of 150 to 200 identical repeats. Each unit contains the highly transcribed 35S and 5S rRNA genes separated by two intergenic spacers (IGS1, IGS2). IGS1 contains the ribosomal replication fork barrier (rRFB) mediating polar fork stalling at three distinct sites while IGS2 harbors an origin of replication (ARS). In addition, both IGS1 and IGS2 carry various regulatory regions (see text for details). CAR, cohesin associated region. Figure adapted from (Tsang & Carr, 2008).

Normally, pausing at the rRFB does not provoke replisome disassembly as indicated by the presence of various replisome components at the arrested RF. In contrast to RF stalling caused by nucleotide depletion, stabilization of the pausing RF at the rRFB does not depend on the main checkpoint kinases Mec1 (ATR in *H. sapiens*) and Rad53 (CHK2 in *H. sapiens*) (Calzada et al., 2005). This might be explained by the fact that nucleotide depletion in Rad53-deficient cells leads to the separation of helicase and polymerase action, thereby producing long stretches of single-strand DNA that serve as a signal for the S-phase checkpoint (Sogo et al., 2002). Consistently, only few nucleotides of single-stranded DNA were detected at rRFB-stalled RFs, at least in checkpoint-proficient cells (Grube et al., 2000). Though being independent of Mec1 and Rad53, pausing at the rRFB is only effective in the

presence of the checkpoint mediators Tof1 and Csm3, resembling the situation in fission yeast, both at the RFBs in the rDNA and in the mating type locus (Calzada et al., 2005). Tof1-Csm3 interacts with the Cdc45-MCM2-7 replicative DNA helicase to stabilize the arrested RF, seemingly counteracting the activity of the Rrm3 helicase required for genome-wide fork progression across protein-DNA barriers (Ivessa et al., 2003). Indeed, Rrm3 is recruited to the rRFB, possibly to transiently remove Fob1 for RF passage, however, the mechanism of RF block release is not completely elucidated (Calzada et al., 2005). Remarkably, RF pausing at tRNAs also requires Tof1, but not Mrc1, showing a similar requirement at these regions that are thought to be induced by clashes between replication and transcription machineries (Hodgson et al., 2007).

Prevention of clashes between 35S transcription and replication are mostly considered as the main rRFB function. As only less than half of the rDNA units are transcribed and only 20% of the ARSs fire per cell cycle (Brewer & Fangman, 1988; Dammann et al., 1993), occurrence of such collisions is minimized also in strains with a non-functional rRFB (*fob1Δ*). However, collisions and increased rDNA recombination were detected in Fob1-deficient strains with a reduced rDNA copy number of 10 to 20 repeats, a condition in which most units are thought to be transcriptionally active. Importantly, these events required active transcription of the 35S rRNA gene (French et al., 2003; Takeuchi, 2003). It is well possible that such collisions also occur in wild-type cells, though rarely, as the rRFB is leaky and some RFs (< 10%) can escape the block. The resulting collapsed RF could then induce recombination events, which, due to their rare occurrence, might be undetectable in classical analyses such as two-dimensional gel electrophoresis (2D gel; Tsang & Carr, 2008).

### **2.2.3 Regulation of rDNA Recombination**

The repetitive structure of the rDNA, together with the presence of an RFB, renders it prone to homologous recombination (HR). Fob1- and replication-dependent DNA breaks have been reported at the rRFB in budding yeast (Burkhalter & Sogo, 2004; Kobayashi et al., 2004; Weitao et al., 2003), yet, these breaks were later shown not to be substrate for HR-dependent DSB repair and, therefore, might not represent canonical DSBs (Fritsch et al., 2010). In line with this, RF restart at the rRFB can be accomplished without the involvement of HR (Calzada et al., 2005). HR was only detected upon deletion of Rrm3 that is required for replication through RFBs and also through the rRFB. The observed instability upon enhanced RF stalling exemplifies the need to tightly regulate RF stalling (Ivessa et al., 2000; Keil & McWilliams, 1993).

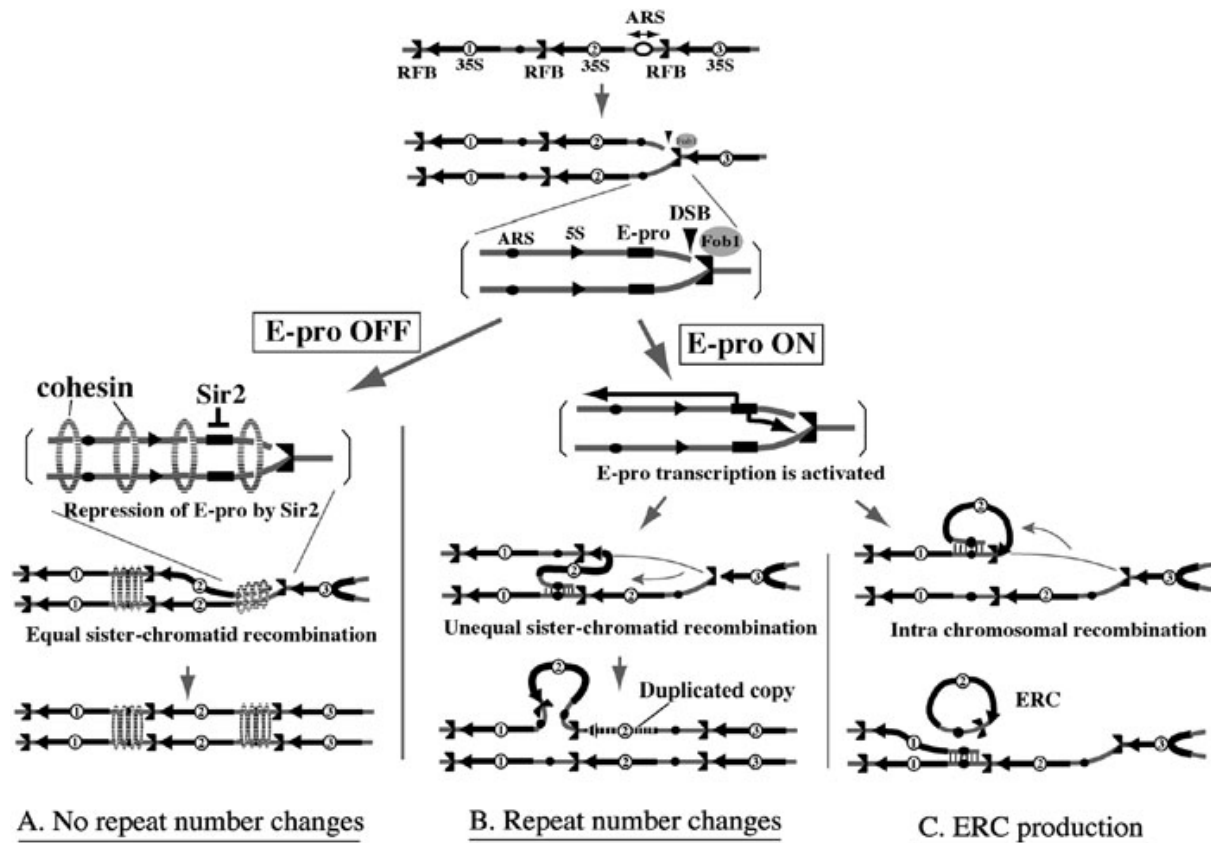
The first evidence for increased recombination in the rDNA came from the identification of an rDNA region (HOT1) that generates a recombination hotspot when placed outside the rDNA locus. HOT1 consists of two elements, one containing the enhancer of RNA polymerase I transcription (HOT1(E))

and one containing the corresponding promoter (HOT1(I); Figure 4) (Keil & Roeder, 1984; Voelkel-Meiman et al., 1987). HOT1(E) also includes the rRFB sequence, but while Fob1 is required for HOT1 activity, blocking capacity of the barrier is not (Kobayashi & Horiuchi, 1996; Ward et al., 2000). The stimulatory effect of HOT1 on recombination completely depends on RNA polymerase I transcription (Huang & Keil, 1995). Fob1-dependency can be alleviated by hyper-activation of the RNA polymerase, suggesting that Fob1 is a stimulator of RNA polymerase I-initiated transcription and, thus, recombination (Serizawa et al., 2004).

Recombination in the rDNA contributes critically to repeat homeostasis and Fob1 plays a major role in the regulation of this process. In wild-type cells, gains or losses of rDNA copies are readily equalized by unequal sister chromatid exchange (SCE), which depends on Fob1 (Johzuka & Horiuchi, 2002), however, no copy number changes have been observed in the absence of Fob1 (Kobayashi et al., 1998). HR in the rDNA also leads to the excision of single or multiple rDNA copies that form extrachromosomal ribosomal DNA circles (ERCs, Figure 5C) and are thought to contribute to yeast cell aging. As expected, Fob-deficient cells generate less ERCs and, in line with ERCs being implicated in aging, their life span is increased (Defossez et al., 1999; Sinclair & Guarente, 1997). Not surprisingly, copy number changes require classical HR genes, such as *RAD52* and *MRE11* (Kobayashi et al., 2004). As for an ectopic HOT1, RNA polymerase I is required for maintaining a stable copy number, yet, the underlying mechanism is elusive (Kobayashi et al., 1998).

Mechanistically, it was proposed that breaks within the rDNA are normally repaired by equal SCE to maintain the correct copy number (Figure 5). Instead of pairing with the equivalent unit on the sister chromatid, the broken unit might unequally anneal and recombine with an ectopic copy, which may increase or decrease the rDNA repeat number. Additionally, intra-sister chromatid recombination would explain the formation of ERCs (Figure 5C). The choice between equal or unequal SCE probably involves the regulation of cohesin association to the rDNA repeat (Kobayashi, 2005). Cohesin was shown to associate with the rDNA where it is supposed to mediate sister chromatid cohesion to hold the two homologous chromatids together (Kobayashi et al., 2004; Laloraya et al., 2000). Cohesin association with the rDNA is disrupted by active transcription from a non-coding bi-directional RNA polymerase II promoter (E-pro, Figure 4), which resides in the intergenic spacer (IGS). Silencing of E-pro by the HDAC Sir2 is thought to ensure cohesin association with the DNA (Kobayashi, 2005). By preventing cohesin dissociation from the rDNA, Sir2 thus ensures correct, parallel alignment of both sister chromatids, which in turn favors equal SCE. In line with Sir2 suppressing unequal SCE, loss of Sir2 leads to increased unequal recombination events as inferred from the incorporation of rRFB-containing plasmids into the rDNA (Benguría et al., 2003; Kaeberlein et al., 1999). When expression from E-pro was artificially suppressed, Sir2 was not longer required for rDNA stability. This clearly

supports the idea that Sir2-dependent silencing of E-pro allows for cohesin binding and, thus, for equal SCE (Kobayashi, 2005).



**Figure 5. Model for rDNA repeat homeostasis.** Fob1 binds to the rRFB and mediates the arrest of the RF that might occasionally break and give rise to DSBs. **(A)** Under normal circumstances transcription from the RNA pol II promoter E-pro is suppressed by Sir2, thereby allowing cohesin binding to the rDNA. DSBs are repaired by equal sister chromatid exchange (SCE) to preserve the wild-type copy number of repeats. **(B)** In the absence of Sir2-dependent E-pro silencing no cohesin associates with the rDNA, allowing recombination between unequal repeats of the sister chromatids. Unequal SCE leads to gains or losses of rDNA copies. **(C)** Recombination can also occur between repeats of the same chromatid, resulting in the excision of extrachromosomal ribosomal DNA circles (ERCs). Figure from (Kobayashi, 2006).

Physical interactions between Fob1 and the two RENT (regulator of nucleolar silencing and telophase exit) complex subunits Sir2 and Net1, suggest that SCE regulation does not only involve Sir2 but the whole RENT complex. Fob1 and Net1 additionally associate with the nucleolar Tof2 protein that binds the cohesin complex subunits Lrs4 and Csm1, which in turn bind cohesin and are important for sister chromatid cohesion. Consistent with the existence of such a complex, Tof2, Lrs4 and Csm1 are found at the IGS1 and are required for silencing of a reporter gene in the rDNA. Thus, Fob1-dependent recruitment of a large complex is required for cohesin-mediated sister chromatid repair. How

unequal SCE is initiated to restore the regular rDNA repeat number after accidental gains or losses, however, is not known (Huang & Moazed, 2003; Huang et al., 2006).

The same complex is seemingly involved in the anchoring of rDNA repeats to the inner nuclear membrane, thereby adding an additional complexity to the rDNA structure. Cohibin indeed interacts with the inner nuclear membrane proteins Heh1 and Nur1 (Huang et al., 2006; Mekhail et al., 2008). Abolishing this interaction by deleting Heh1, Nur1 or one of the cohibin subunits Lrs4 and Csm1 triggered rDNA instability and a shortened life span. On the contrary, an artificial gene fusion of Sir2 and Heh1 rescued the rDNA instability phenotype of Lrs4-deficient cells (Chan et al., 2011; Mekhail et al., 2008). Importantly, cohibin is crucial for silencing, whereas the inner nuclear membrane proteins are dispensable, indicating that the Fob1-RENT-Tof2-cohibin complex exerts two, at least partially, independent activities important for rDNA stability (Mekhail et al., 2008). While anchoring of the repeats to the nuclear membrane is beneficial to rDNA stability, it also imposes increased DNA torsion due to restricted mobility, suggesting the need of topoisomerases in this region. Topoisomerases have indeed an important role within the rDNA locus that is discussed in detail in chapter 2.3.2.

Additional regulators of rDNA stability are the condensins as their disruption leads to rDNA instability. It was suggested that condensins associate with sites of replication termination to which they recruit the topoisomerase Top2. Top2 would then be required to release the increased torsion resulting from the merging replication forks. In the absence of Fob1, no rRFB activity is established that prevents collisions between DNA replication and RNA transcription. In this situation, condensin becomes particularly important, probably reflecting the recruitment of Top2 for releasing torsional stress (Bhalla et al., 2002; Johzuka et al., 2006; Kobayashi, 2006; Sullivan et al., 2004; Tsang & Carr, 2008).

## **2.3 DNA Topoisomerases in the Removal of Torsional Stress**

### **2.3.1 DNA Topoisomerase Functions**

DNA topoisomerases are the main proteins to relieve torsional stress in the DNA imposed by DNA replication, transcription, repair and chromatin remodeling. They achieve this by transiently cleaving one or both DNA strands. A tyrosine residue at the active site of the topoisomerase serves as a nucleophile, attacking the phosphodiester backbone of the DNA. The resulting transesterification creates a DNA break with a covalently attached topoisomerase. After relaxation of the DNA, the break is resealed by a second transesterification, releasing the enzyme. Two types of topoisomerases are distinguished: Type I topoisomerases are monomeric and introduce a SSB through which the

second strand is passed to decrease the helical turn. In contrast, dimeric type II enzymes create a DSB to pass a second double-stranded DNA through the break. These groups are further divided into A and B subfamilies. While IA, IIA and IIB topoisomerases are attached to the 5' end of the DNA during cleavage, IB enzymes are associated with the 3' end.

Six topoisomerases are expressed in humans, two members of each type IA, IB, and IIB topoisomerases, whereas yeast cells harbor only one representative of each subfamily (Champoux, 2001; Chen et al., 2013). Type IB topoisomerases will be discussed in detail in chapter 2.3.2.

#### *Type IA topoisomerases*

Type IA topoisomerases relax negative supercoiling *in vitro*, but they fail to perform this to completion, indicating that their action is important to reach a level of torsion that is optimal for a certain functional context. The fact that activity of type IA topoisomerases depends on presence of single-stranded DNA regions serves as a good explanation for the observed interaction with RecQ helicases. Through unwinding of the DNA duplex, RecQ helicases provide topoisomerases with the required single-stranded DNA substrate. While the yeast Top3, a type IA topoisomerase, interacts with the RecQ helicase Sgs1, the human TOP3A cooperates with the BLM helicase. Type IA topoisomerases acts on synthetic DNA substrates mimicking dHJs, a proposed HR intermediate, *in vitro*. In line with a similar role *in vivo*, Top3-deficient yeast cells accumulate X-shaped DNA structures that were reduced upon Rad51 deletion. In higher eukaryotes TOP3A is embryonic lethal and this phenotype most likely reflects its importance for genome stability during development (Champoux, 2001; Chen et al., 2013).

#### *Type IIA topoisomerases*

Generally, type IIA topoisomerases are important for chromosome segregation and to counteract positive and negative DNA supercoiling. Despite exerting similar enzymatic activities, the two mammalian enzymes TOP2A and TOP2B have different biological functions, reflecting differential interactions with regulatory proteins and/or post-translational modifications (Champoux, 2001; Chen et al., 2013). TOP2A is essential for cell viability due to its role in unlinking intertwined daughter chromosomes after replication (Chen et al., 2013). TOP2B-deficient mice die at birth and TOP2B as well as the Top2 enzymes from budding yeast and the fruit fly associate with gene promoters and to the nucleosome-free region around genes, indicating a specific role of this protein in gene regulation (Lyu et al., 2006; Sperling et al., 2011; Udvardy et al., 1985; Yang et al., 2000).

### *Type IIB topoisomerases*

Type IIB topoisomerases are only found in archaea and plants. However, the SPO11 endonuclease is related to these enzymes. SPO11 is essential in meiosis where it uses the cleavage activity to induce DSBs for meiotic recombination (Chen et al., 2013; Keeney, 2008).

### **2.3.2 Type IB Topoisomerases**

The human TOP1 consists of four domains. The non-conserved N-terminus is unstructured and dispensable for catalysis, but harbors important regulatory elements, such as nuclear localization signals (NLSs) and protein interaction sites for – amongst others – several TFs, *p53* and the RecQ helicase WRN. Next to the N-terminus lies the highly conserved active core domain containing all catalytic residues except the tyrosine mediating the transesterification. The core domain is connected to the C-terminus harboring the catalytically essential tyrosine (Tyr723 in humans) by a non-conserved linker region. Together with the core domain, the C-terminus is sufficient for TOP1 relaxation activity *in vitro*. Crystal structures suggest that the protein forms a clamp-like structure that can open and close around the substrate DNA. Binding to the DNA involves multiple contacts between 14 base pairs of the phosphate backbone and the catalytic cavity of TOP1 with most interactions being established five base pairs upstream of the SSB. Some degree of sequence specificity has been observed 5' of the cleavage site [5'-(A/T)(G/C)(A/T)T-3'] with the most important feature being the 3' thymine next to the cleavage site. During catalysis, the phosphodiester backbone of the DNA is target of a nucleophilic attack by the  $O^4$  atom of the catalytic tyrosine, resulting in a covalent bond between the tyrosine and the 3' phosphate of the DNA (Top1 cleavage complex, Top1cc). Release of torsion might be accomplished by allowing the SSB end to freely rotate around the other strand before the resealing step. However, in crystal structures both ends of the SSB were attached to TOP1, suggesting a controlled movement of the SSB end (Champoux, 2001; Chen et al., 2013; Leppard & Champoux, 2005).

Functionally, IB topoisomerases relieve both positive and negative supercoiling. At least one representative is expressed in all eukaryotes and their presence is essential during development in higher eukaryotes (Chen et al., 2013). In the wake of the transcription machinery, negative torsion is formed that opens the chromatin. Occasionally, this leads to the formation of a DNA/RNA hybrid (R-loop) by invasion of the nascent RNA to the DNA duplex behind the replication fork. Consistent with TOP1's central function in relaxing torsional stress, TOP1-deficient mouse cells lead to the formation of R-loops and they accumulate stalled RFs and DNA breaks in gene-rich regions. Interestingly, the resulting slower RF progression is counteracted by firing from an increased number of origins firing



replication (Tuduri et al., 2009). In budding yeast, the type IB topoisomerase Top1 dispensable for cell viability as Top1 and Top2 have overlapping functions in the removal of torsional stress (Reid et al., 1998). It was suggested that positive supercoiling ahead of the RF is normally resolved by Top1, while Top2 is rather involved in the dissolution of precatenates and negative supercoiling in the wake of the RF (Figure 1). In line with a partially redundant role at progressing RFs, fork breakage and concomitant activation of the intra-S-checkpoint was only detected in the absence of both proteins. However, decatenation of intertwined sister chromatids after replication could constitute a non-complementable feature as Top2-deficient cells accumulate cruciform structures during S-phase and activate the M-G1 checkpoint (Bermejo et al., 2007).

### **Non-Canonical Functions of Type IB Topoisomerases**

Given their function in the regulation of torsional stress during transcription, type IB topoisomerases are expected to affect gene expression by modulating the accessibility of regulatory DNA elements for transcription factors and the RNA polymerase. It was indeed shown that type IB topoisomerase deficiencies alter gene expression at specific genes (Miao et al., 2007; Soret et al., 2003). Moreover, *top1*<sup>+</sup> deletion in *S. pombe* increases nucleosome occupancy and alters histone modifications at gene promoters, leading to reduced transcription rates. It was proposed that Top1 cooperates with the chromatin remodeler Hrp1 to maintain or induce active transcription by nucleosome eviction (Durand-Dubief et al., 2010; Gavin et al., 2001). However, in yeast, this activity might be at least partially redundant with Top2 (Pedersen et al., 2012; Sperling et al., 2011). Recently, Top1 was also implicated in the repression of transposable elements in plants. Such elements are normally silenced by non-coding RNA-directed DNA methylation and histone 3 lysine 9 di-methylation (H3K9me2), but inhibition of TOP1 $\alpha$  by camptothecin (CPT) led to de-repression of these loci. This de-repression might be readily explained by the transcriptional inactivation of silencing long non-coding RNAs or, alternatively, by affecting other epigenetic mechanisms, e.g. histone modifications (Dinh et al., 2014).

Top1 is also important for preventing transcription-associated mutagenesis (TAM), which likely involves inhibition of R-loop formation at highly transcribed genes. This is achieved by the removal of negative torsion behind the transcription machinery and by regulating splicing factors and ribonucleoprotein assembly, thereby ensuring the engagement of the nascent RNA in posttranscriptional processes (Kim & Jinks-Robertson, 2012). On the contrary, Top1 accounts for transcription-related deletions (2-5 nts) within short tandem repeats. This might reflect an increased trapping of Top1 covalent complexes (Top1ccs) on highly transcribed DNA and subsequent error-

prone repair within the repeats (Takahashi et al., 2011; Lippert et al., 2011; Kim & Jinks-Robertson, 2012).

Moreover, Top1 has an additional endonuclease activity at ribonucleotides within genomic DNA, an activity that was suggested to serve as a backup for RNase H2-dependent ribonucleotide excision repair (RER). Removal of a ribonucleotide involves an irreversible cleavage resulting in a 2', 3' cyclic phosphate at the 3' end of the SSB that requires further processing. As RNase H2 deletion also triggers Top1-dependent small deletion mutations, ribonucleotide removal in addition to trapped Top1ccs might contribute to mutagenesis (Sekiguchi & Shuman, 1997; Kim et al. 2011; Kim & Jinks-Robertson, 2012).

In addition, TOP1 was found as a regulator of a circadian protein Bmal1 and even more, TOP1 expression was regulated by circadian clock proteins, indicating an involvement of TOP1 in the circadian clock (Kuramoto et al., 2006; Onishi & Kawano, 2012; Yang et al., 2009). Type IB topoisomerases were also suggested to be important for DNA fragmentation during apoptosis. ROS are important for the initiation of apoptosis and oxidative damage could interfere with the religation of the DNA by TOP1, thereby trapping the topoisomerase on the DNA that in turn interferes with the RF, giving rise to DNA breaks (Sordet et al., 2004a; Sordet et al., 2004b; Sordet et al., 2004c).

Strikingly, functions of type IB topoisomerases in gene regulation are not restricted to their relaxation activity. Using *in vitro* transcription reconstitution it was shown that human TOP1 is capable of modulating transcription independent of its catalytic activity by being recruited to the pre-initiation complex by the general TF TFIID, where it promotes the assembly of a TFIID-TFIIA complex (Merino et al., 1993; Shykind et al., 1997).

In addition to its widely known topoisomerase function, TOP1 harbors an intrinsic kinase activity and it was suggested that proteins exerting either function adopt distinct conformations. The kinase activity of TOP1 was implicated in the process of splicing, where it phosphorylates arginine-serine-rich (SR) splicing factors. The switch between cleavage and kinase activity of TOP1 was proposed to be mediated by poly(ADP)ribose that competitively binds to the splicing factor, thereby favoring topoisomerase activity on DNA (Malanga et al., 2008; Rossi et al., 1996).

### **Function of Type IB Topoisomerases in the rDNA**

The high transcription level in the rDNA, also during DNA replication, combined with the topological constraints imposed by anchoring of the rDNA repeats to the inner nuclear membrane creates a constant need for removing torsional stress (Chan et al., 2011; Huang et al., 2006; Mekhail et al., 2008). Consistently, type IB topoisomerases are enriched in the nucleolus (Muller et al., 1985) and co-localize with RNA polymerase I at sites of active transcription (Leppard & Champoux, 2005).

Moreover, Top1 was shown to suppress mitotic recombination within the rDNA of budding yeast (Christman et al., 1988; Kim & Wang, 1989; Zhu & Schiestl, 2004), possibly by its involvement in Sir2-dependent silencing of E-pro. In line with a role in Sir2-dependent silencing, Top1 was co-purified with Tap-tagged Fob1 and a Top1-Tof2 physical interaction was reported (Huang et al., 2006; Park & Sternglanz, 1999; Smith et al., 1999). As Fob and Tof2 proteins also mediate anchoring to the inner nuclear membrane, the reported interactions also reflect the need of torsion release by Top1 at sites within the rDNA. Interestingly, Top1-related DNA nicks were observed in the promoter and enhancer region of the 35S rRNA gene and nicks in the enhancer region map near the rRFB (Vogelauer & Camilloni, 1999). These nicks are independent of both replication and transcription, but require presence of the fork blocking protein Fob1 (Di Felice et al., 2005).

### **2.3.3 Repair of Irreversible Top1 Covalent Complexes**

Type IB topoisomerases cleave the DNA by forming Top1 cleavage complexes (Top1cc) that are covalently attached to the DNA. Top1ccs are readily reversible under unperturbed conditions, but Top1ccs might be irreversibly trapped in their DNA-bound conformation upon misalignment of the 5' DNA termini by the presence of DNA lesions or by topoisomerase inhibitors that are used as anti-cancer drugs. One such compound is camptothecin (CPT) that inhibits or delays the religation of the Top1-mediated nick. The cytotoxicity of stabilized Top1ccs is thought to be caused by collisions between the progressing RF and Top1ccs, resulting in the formation of DNA DSBs. While this explains the preferential killing of proliferating cells, cytotoxicity is still observed in non-dividing cells through the interference of irreversibly stabilized Top1ccs with transcription (Pommier et al., 2006). To counteract the detrimental effect of accumulating irreversible Top1ccs, multiple repair pathways contribute to their repair. I will focus on the repair pathways in budding yeast as I used this model to investigate topoisomerase functions in the rDNA. Besides a likely participation of DSB repair, a main Top1cc repair pathway involves the highly conserved tyrosyl-DNA phosphodiesterase 1 (Tdp1) that hydrolyzes the phosphodiester bond between the Top1 and the 3' DNA end (Pouliot et al., 1999; Yang et al., 1996). *In vitro* studies suggest that proteolytic degradation of Top1 is necessary to allow for cleavage by Tdp1 (Interthal & Champoux, 2011). The remaining amino acid residues are then readily removed by Tdp1, resulting in a SSB with a 3'-phosphate and a 5'-OH that requires further processing. In budding yeast, the 3'OH is restored by the activity of the DNA 3'-phosphatase Tpp1 or by the two AP endonucleases Apn1 and Apn2 (Karumbati et al., 2003). Subsequent repair in mammalian cells appears to involve the SSB repair pathway, including the polynucleotide 5'-kinase/3'-phosphatase (PNKP) (Pommier et al., 2006). Besides the above-mentioned pathways, a number of nucleases have been found to counteract CPT toxicity. Epistasis analyses suggest a

contribution of the structure-specific endonucleases Rad1-Rad10 and Mus81-Mms4 to the repair of irreversible Top1ccs by removing the whole 3' DNA flap (Liu et al., 2002; Vance & Wilson, 2002). A more recent publication further implicates the nucleases Slx1-Slx4 and Mre11-Rad50 in Top1cc repair (Deng et al., 2005), however, the relative contribution of the different pathways under physiological conditions remains unclear.

## 2.4 Uracil as a Source of Genome Instability

### 2.4.1 Uracil in DNA – Origin and Consequences

Deoxynucleotides containing the bases guanine, cytosine, adenine and thymine, make up the genetic code of the DNA while in RNA thymine is replaced by uracil that lacks the C<sup>5</sup> methyl group. Normally, dUMP pools are kept low by the action of the deoxyuridine 5'-triphosphate pyrophosphatase (dUTPase) that catalyzes the conversion of dUTP to dUMP, which is in turn used for *de novo* synthesis of dTMP, catalyzed by the thymidylate synthase (Friedberg, 2006). Thymidylate synthase can be inhibited by the uracil analog 5-FU and related anti-metabolites, thereby increasing the dUMP/dTMP ratio, which ultimately favors the incorporation of dUMP into DNA. At the same time, 5-FU and its metabolites are directly incorporated into DNA and RNA, all contributing to the toxicity of the drug (Longley et al., 2003). Even in unchallenged cells, residual dUTPs escape conversion to dUMP and are occasionally incorporated into DNA, highlighting the inefficient discrimination between dUTP and dTTP of most DNA polymerases (Friedberg, 2006; Nilsen et al., 2000). A steady state levels of 400 to 600 uracils per human or murine genome have been reported (Galashevskaya et al., 2013). Given that uracil has the same coding properties as thymine, the resulting A•U pairs are not mutagenic, yet they may interfere with TF binding (Rogstad et al., 2002; Verri et al., 1990) or alter protein expression levels (Lühnsdorf et al., 2014). Interestingly, uracil in DNA is tolerated to a certain extent (Muha et al., 2012; Nilsen et al., 2000; Warner et al., 1981) although dUTPase deletion is lethal in many organisms (Dengg et al., 2006; Dubois et al., 2011; el-Hajj et al., 1988; Gadsden et al., 1993). Strikingly, deletion of the major uracil excision enzyme (Ung) alleviates the dUTPase deletion phenotype in *E. coli* and *C. elegans*, indicating that not the uracil *per se* creates the toxicity, but its excision during DNA repair (Warner et al., 1981; Dengg et al., 2006). This is also the case for a hypermorphic dUTPase mutation in *S. cerevisiae*, but not for the dUTPase null mutant, suggesting that at high uracil levels, toxicity is induced independent from Ung (Guillet et al., 2006).

An additional route leading to uracil in DNA is the spontaneous hydrolytic deamination of cytosines that occurs 70 to 200 times per day in the human genome (Visnes et al., 2009). In contrast to replication-derived uracil, unrepaired cytosine deamination generates G•U mismatches and

eventually provokes C to T transitions, a mutation commonly found in human cancers (Sousa et al., 2007). To prevent mutagenesis, the uracil has to be excised from the DNA and repair has to be completed before the site is replicated (Visnes et al., 2009). Cytosine deaminates 200-300 times faster in single-stranded DNA than in double-stranded DNA and may therefore contribute especially to mutagenesis during replication and at active genes (Krokan et al., 2002). Besides these accidental events generating mutagenic uracil in DNA, programmed cytosine deamination occurs in the context of somatic hypermutation and class-switch recombination in activated B cells, two mechanisms that are discussed in the next chapter. DNA-based uracil has additional developmental roles in *D. melanogaster*, where dUTPase-dependent uracil accumulation results in targeted cell death (Horváth et al., 2013; Muha et al., 2012), underpinning the multi-faceted responses to uracil in DNA.

## 2.4.2 Uracil DNA Glycosylases

To prevent uracil-derived mutagenesis, mechanisms for uracil removal have evolved in all organisms. An important part is played by uracil DNA glycosylases (UDGs), a family of DNA repair enzymes that initiate BER by recognizing and excising uracil from DNA. Five UDGs have been identified in mammals: the nuclear and mitochondrial form of the uracil-N glycosylase (UNG), UNG2 and UNG1, respectively, the single-strand-selective mono-functional uracil DNA glycosylase (SMUG1), the thymine DNA glycosylase (TDG) and the methyl-binding domain glycosylase 4 (MBD4, Figure 6). UNG2 and TDG will be presented in detail below, followed by a discussion about the contribution of the single UDGs to cellular functions.

	<i>E. coli</i>	<i>S. cerevisiae</i>	<i>S. pombe</i>	<i>H. sapiens</i>
<b>UDGs</b>	Ung	Ung1	Ung1	hUNG1
				hUNG2
	Mug		Thp1	hTDG
				hSMUG1
				hMBD4
<b>AP endonucleases</b>	Xth			<b>hAPE1</b>
		Apn2	<b>Apn2</b>	hAPE2
	Nfo	<b>Apn1</b>	(Apn1)	
			(Uve1p)	

**Figure 6. Uracil DNA glycosylases (UDGs) and AP endonucleases expressed in different species.** Orthologs of the different organisms are shown in each row. The main AP endonuclease is indicated in bold letters. Figure adapted from (Kanamitsu & Ikeda, 2010).

## The UNG Family of Glycosylases

UNGs are fundamental DNA repair enzymes and are highly conserved among most organisms, including viruses, highlighting their importance for genome stability (Aravind & Koonin, 2000; Olsen et al., 1989). Ung of *E. coli* and human UNG show about 58% sequence identity, demonstrating the high conservation within the UNG glycosylases family (Jacobs & Schar, 2012). It is mainly the C-terminal catalytic domain that is conserved among organism, while the N-terminal end is heterogeneous and was proposed to have regulatory functions and/or to be involved in protein-protein interactions or subcellular localization (Zharkov et al., 2010). Interestingly, *D. melanogaster* does not express any UNG enzyme and consequently tolerates uracil accumulation during vegetative growth (Horváth et al., 2013; Muha et al., 2012). In mammalian cells, nuclear and mitochondrial UNG isoforms are expressed from the *UNG* gene using alternative transcription initiation sites and alternative splicing (Helland et al., 1993; Otterlei et al., 1998). The major enzymatic activity of UNG is the excision of uracil from double-stranded and single-stranded DNA, the latter being processed more efficiently (Eftedal et al., 1993; Lindahl et al., 1977; Slupphaug et al., 1995). Despite its high preference for uracil, UNG is also active on several uracil analogs *in vitro*, such as the anti-cancer drug 5-FU (Warner & Rockstroh, 1980), 5-hydroxyuracil or isodialuric acid (Dizdaroglu et al., 1996). Human UNG2 interacts with PCNA and RPA and is thought to be the main activity for the post-replicative removal of uracils incorporated by the DNA polymerase. Consistently, UNG2 localized to the nucleus and is up-regulated in S-phase. Along this line, hUNG2 phosphorylation at Ser23 at the G1/S-phase boundary increases its interaction with RPA and replicating chromatin. Simultaneously, the catalytic turnover of hUNG2 is enhanced, probably to cope with the fast-moving RF. During S-phase, the affinity to RPA is reduced and the protein is degraded in late S/G2 phase of the cell cycle (Hagen et al., 2008; Otterlei et al., 1999).

Mechanistically, UNG is thought to scan the DNA in an open conformation using non-specific interactions until it recognizes a substrate uracil. The base is then flipped out of the DNA helix into the catalytic pocket (Friedman et al., 2009). Base-flipping could be a passive process induced by the spontaneous breathing of the DNA (Fadda & Pomès, 2011; Parker et al., 2007), but recently, a simulation of the base recognition process, starting from the reported crystal structures of human UNG, suggested that the DNA bends upon UNG binding to uracil-containing DNA, thereby facilitating base flipping (Franco et al., 2013). Once uracil is accommodated in the catalytic pocket, specific hydrogen bonds are formed between UNG1 and the base, which are important for damage recognition. Purines are sterically excluded from the narrow active site while the entry of thymine is precluded by a tyrosine residue in the catalytic pocket. In contrast, cytosine can enter, but is not correctly positioned for catalysis (Jacobs & Schar, 2012). Finally, the *N*-glycosidic bond is cleaved by a nucleophilic attack of an activated water molecule on the C<sup>1</sup> of the deoxyribose (Mol et al., 1995;

Savva & Pearl, 1995). In comparison to other DNA glycosylases, UNG has a very high turnover rate of approximately 1000 uracils per minute (Kim & Wilson, 2012).

In the adaptive immune response, UNG2 plays another essential role in somatic hypermutation (SHM) and class switch recombination (CSR) in activated B cells. SHM modulates the affinity of the antibody to the epitope by creating point mutations at the IgV locus that depend on the targeted action of the activation induced cytosine deaminase (AID). UNG2-mediated uracil excision creates AP sites, which are thought to trigger recombination or error-prone bypass during DNA replication or DNA repair, thereby inducing mutations. Similarly, rearrangement of the constant region of the antibody during CSR is thought to rely on the excision of AID-targeted uracils by UNG2 that might be further processed by AP endonucleases. Two close SSBs on opposite strands could then lead to the formation of a DSB, and, hence, induce a recombination repair process. Alternatively, AP sites may directly interfere with replication to trigger recombination events (Kavli et al., 2007; Schrader et al., 2005; Stavnezer et al., 2008; Wang, 2013). Consistent with a role of UNG in the adaptive immune response, UNG-deficient humans suffer from recurrent infections and lymphoid hyperplasia (Sousa et al., 2007) and UNG-deficient mice develop B cell lymphomas late in life (Nilsen et al., 2003). UNG2 was further implicated in the innate immune response, i.e. in the defense against retroviral infections. Similar to the situation in somatic hypermutation and class-switch recombination, this activity involves cytosine deamination, in this case mediated by the APOBEC3 family of deaminases. Again, excision of uracil followed by AP endonuclease processing or direct interference of the AP site with the RF would lead to DNA fragmentation of the synthesized viral DNA (Yang et al., 2007). Moreover, UNG2 is involved in the assembly of CENP-A, an essential histone variant required for kinetochore assembly prior to mitotic chromosome segregation. Catalytic activity is required for this function, however, the underlying mechanism is not yet resolved (Zeitlin et al., 2011; Zeitlin et al., 2005).

### **The Thymine DNA Glycosylase TDG**

TDG is the founding member of the mismatch-dependent uracil glycosylase (MUG) family that was named after the *E. coli* protein Mug (Gallinari & Jiricny, 1996; Neddermann & Jiricny, 1994; Wiebauer & Jiricny, 1989). MUG orthologs are expressed in bacteria, yeast, insects and frogs and they share a common  $\alpha/\beta$ -fold structure despite the lack of high sequence homology (Barrett et al., 1998; Cortázar et al., 2007). They harbor a conserved catalytic core flanked by non-conserved N- and C-terminal domains, the latter being involved in modulating enzymatic activities and interactions with other proteins (Hardeland et al., 2002; Steinacher & Schar, 2005). MUG proteins have a spacious catalytic cavity that accommodates a broad range of substrates, including pyrimidine derivatives and

thymine opposite of guanine, but also bulkier lesions, such as etheno cytosine (Borys-Brzywczy et al., 2005; Hardeland et al., 2003). However, despite their name, uracil mispaired with guanine appears to be the common substrate among all family members (Cortázar et al., 2007).

Crystal structures of the *E. coli* Mug and the human TDG protein revealed that for base excision, the target base is flipped into the catalytic pocket of the protein. At the same time protein residues intercalate into the DNA at the site of the lesion. Interactions with the complementary DNA strand serve as a good explanation for the high preference of the Mug family proteins for base lesions in double-stranded DNA located opposite to guanine (Baba et al., 2005; 2006; Barrett et al., 1998; 1999; Maiti et al., 2008). TDG was suggested to undergo a conformational change upon unspecific interaction with DNA. TDG would form a clamp around the DNA to scan the DNA while sliding along the helix. A striking feature of TDG and the other MUG proteins is their extremely high affinity to AP sites, which is in stark contrast to that of the UNG proteins. While the *E. coli* Mug can turn over very slowly, the mammalian TDG is completely product-inhibited (O'Neill et al., 2003; Waters & Swann, 1998; Waters et al., 1999). It was shown that covalent interaction between the C-terminus of TDG and the small ubiquitin-like modifier 1 (SUMO1) reduces the affinity to the substrate and the AP site. As a result TDG dissociates from the lesion and can initiate another round of repair (Baba et al., 2005; Hardeland et al., 2002; Steinacher & Schar, 2005; Waters et al., 1999; Waters & Swann, 1998). Product inhibition might serve as a mechanism for controlled handover of the product AP site to downstream BER factors.

### **TDG and DNA Repair**

TDG is active on a wide range of uracil derivatives, implying a main function in the repair of deaminated or oxidized cytosines (Cortázar et al., 2007). Surprisingly, no protection against DNA damaging agents and no contribution to mutation avoidance could be attributed to TDG (Cortázar et al., 2011). In mammalian cells, the contribution of TDG to the repair of deaminated cytosines appears to be rather complementary to that of UNG2 and SMUG1 (Cortázar et al., 2007).

The slow turnover of TDG might explain an interesting phenotype observed for mammalian TDG exposed to the uracil analog 5-FU. Exposure of cells to 5-FU was cytotoxic in wild-type cells but not in cells deleted for TDG. Similarly, 5-FU-dependent occurrence of DNA breaks, as well as 5-FU-delayed S-phase progression and DNA damage signaling were reduced in the absence of TDG, implying that repair by TDG is detrimental for cells facing 5-FU (Kunz et al., 2009). The tight binding of TDG to the AP site might have adverse effects under conditions of BER saturation, which is likely to occur during 5-FU treatment. Prolonged existence of AP sites could lead to accumulation of DNA breaks and finally to cell death. This non-productive engagement in DNA repair, together with the lack of general DNA



repair phenotypes, suggests that TDG is not a classical repair enzyme but has acquired additional functions.

### **TDG in the Epigenetic Regulation of Gene Expression**

During the last years, additional roles for TDG emerged in the regulation of gene expression. TDG interacts with and co-regulates several TFs and nuclear receptors including the retinoic acid receptor (RAR), retinoid X receptor (RXR), estrogen receptor  $\alpha$  (ER $\alpha$ ), the acetyltransferase CBP/p300 and the deacetylase SIRT1. Interestingly, in some of these cases, the catalytic activity of TDG is dispensable, suggesting a structural scaffold function (Chen et al., 2003; Chevray & Nathans, 1992; Goodman & Smolik, 2000; Jia et al., 2014; Madabushi et al., 2013; Missero et al., 2001; Tini et al., 2002; Um et al., 1998; Zhou et al., 2008). A main step in deciphering this role was the discovery by our and the Bellacosa group that deletion of *TDG* is embryonic lethal, a phenotype not shared with any known glycosylase (Cortázar et al., 2011; Cortellino et al., 2011). This implied a non-redundant function for TDG during embryonic development. Gene expression analyses of WT and *TDG*<sup>-/-</sup> mouse embryonic fibroblasts (MEFs) revealed differential expression between both strains. At the same time, promoters of down-regulated genes exhibit altered chromatin modification patterns with a loss of activating and a gain in repressive marks. Interestingly, these changes became apparent only in differentiating or differentiated cells and were not present in undifferentiated embryonic stem cells (ESCs), in line with a specific role during development (Cortázar et al., 2011; Cortellino et al., 2011).

In addition, TDG has been associated with the active demethylation of 5mC. Consistently, TDG locates to promoters of genes differentially expressed in TDG-proficient and deficient MEFs (Cortázar et al., 2011). Not only is TDG required for maintaining the unmethylated state of CpG islands upstream of specific genes during development, but it also mediates demethylation of genes that are upregulated upon cellular differentiation (Cortázar et al., 2011; Cortellino et al., 2011). Given the requirement for catalytically active TDG during embryogenesis, and for appropriate demethylation, this process probably involves excision of a demethylation intermediate by TDG. In line with a BER involvement, XRCC1 and APE1 associate in a TDG-dependent manner with promoters of genes that are differentially expressed in *TDG*<sup>-/-</sup> and *TDG*<sup>+/+</sup> MEFs (Cortázar et al., 2011). During the last years, different demethylation intermediates were proposed to be substrate for TDG. First, 5mC itself was discussed as a possible target, although an activity of TDG in 5mC excision could not be corroborated (Zhu et al., 2000).

Second, the cytosine deaminases AID and APOBEC1 were shown to deaminate 5mC albeit with lower activity as compared to cytosine, giving rise to T•G mispairs that constitute a suitable substrate for TDG, but also for MBD4 (Morgan et al., 2004). Supporting a deamination-related demethylation, AID

was shown to contribute to genome-wide erasure of 5mC in murine primordial germ cells (Popp et al., 2010) and to demethylation-dependent re-expression of the pluripotency genes *OCT4* and *NANOG* during somatic cell reprogramming (Bhutani et al., 2010). AID-mediated and MBD4-coupled DNA demethylation was also observed in zebrafish embryos (Rai et al., 2008).

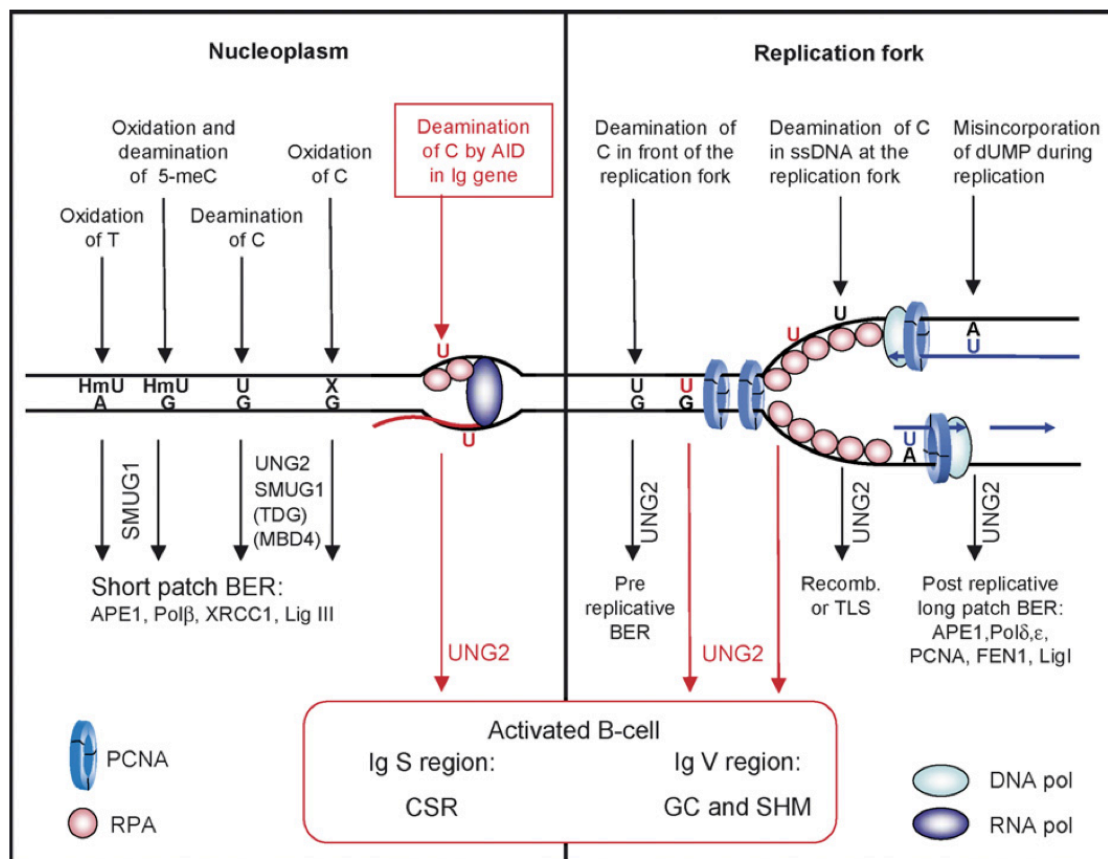
Third, experimental evidence suggests a function of the *de novo* DNA methyltransferases DNMT3a and DNMT3b in conjunction with TDG in the active elimination of DNA methylation (Métivier et al., 2008). Interaction between these proteins inhibits methylation activity of DNMT3a/b and enhances TDG's glycosylase activity (Boland & Christman, 2008; Li et al., 2007). Furthermore, DNMTa/b were shown to act as 5meC deaminases *in vitro*, thereby giving rise to the perfect TDG substrate, a G•T mismatch (Liutkeviciute et al., 2009). However, it is not yet clear whether such events can occur under physiological conditions, especially as deamination was only observed in the absence of the otherwise abundant methyl-donor S-adenosyl-L-methionine (SAM) (Kohli & Zhang, 2013).

Currently, the favored model for DNA demethylation involves the ten-eleven translocation (TET) dioxygenases, which are capable of iterative oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Kohli & Zhang, 2013). 5hmC is indeed present in a variety of undifferentiated and differentiated cells and occurs in a TET-dependent manner (Globisch et al., 2010; Ito et al., 2010; Kriaucionis & Heintz, 2009; Song et al., 2011; Tahiliani et al., 2009). 5fC and 5caC were also found in DNA of mouse ESCs, albeit at a much lower level than 5hmC (Ito et al., 2011). Although TDG lacks activity on 5hmC, it was shown to excise 5fC and 5caC *in vitro* and is so far the only glycosylase with such an activity (He et al., 2011; Maiti & Drohat, 2011). In mouse ESCs, TDG depletion caused an increase in 5caC levels while overexpression of TET and TDG in HEK293 cells diminished 5fC and 5caC levels, supporting cellular significance of TDG/TET-mediated DNA demethylation (He et al., 2011; Nabel et al., 2012). This pathway is far from being elucidated and other enzymes have been proposed to take part in the removal of DNA demethylation intermediates, such as AID and the NEIL glycosylase (Müller et al., 2014; Santos et al., 2013)

### **2.4.3 Current Insights into the Functional Separation of Uracil DNA Glycosylases**

Five UDGs are expressed in mammalian cells, UNG1 and UNG2, SMUG1, TDG and MBD4 (Figure 6). Among these, UNG2 is the major enzyme removing uracils that were misincorporated during replication (Figure 7). A preferential role for UNG2 in replicating cells is inferred from its upregulation before and degradation after S-phase, its association with PCNA and RPA and its localization to replication foci (Hagen et al., 2008; Otterlei et al., 1999). It was suggested that it does not only act in the post-replicative repair of misincorporated uracil, but also in the repair of uracil derived from cytosine deamination in unreplicated DNA. Indeed, UNG2 is also the central player in the repair of

G•U mismatches (Kavli et al., 2007). Biochemical evidence suggests that in the absence of UNG2, SMUG1 might be the predominant G•U repair activity (Nilsen et al., 2001), although SMUG1 activity appears to be species-, tissue-, and condition-dependent (Doseth et al., 2011; Kemmerich et al., 2012; Nagaria et al., 2013). While SMUG1 provides considerable G•U repair in murine cell-free extracts, this contribution is only marginal in human extracts (Doseth et al., 2011; Kemmerich et al., 2012). Mutation assays in MEFs revealed slightly increased mutation rates in *UNG2*<sup>-/-</sup> and *SMUG1*<sup>-/-</sup> single knockout strains while rates increased slightly more than in an additive manner in the *UNG2*<sup>-/-</sup> *SMUG1*<sup>-/-</sup> double knockout cells (An et al., 2005), suggesting that both enzymes are also important in living cells. Elevated mutation rates might not only reflect the removal of uracil but also the repair of other base lesions; thus, UNG2 and SMUG1 might have redundant as well as independent functions in mutation avoidance. Independent functions of these proteins might also involve a temporal or spatial separation of their action. A redundant interaction between UNG2 and SMUG1 was observed



**Figure 7. Uracil analogs in DNA and their proposed route of repair in replicated and unreplicated DNA.** UNG2 provides the main activity for the removal of uracils at the RF, i.e. misincorporated uracils and deaminated cytosines. Outside of S-phase, UNG2 and SMUG1 both contribute to the repair of deaminated and oxidized cytosines, while TDG and MBD4 are thought to fulfill only backup functions in this process. SMUG1 was proposed to have specific roles in the removal of hydroxymethyluracil (HmU) deriving from thymine oxidation or 5-methylcytosine (5mC) deamination. In addition to classical DNA repair, UNG2 is involved in the processes of class switch recombination (CSR) and somatic hypermutation (SHM) in activated B-cells (see text for details). Figure from (Kavli et al., 2007).

in the repair of gamma-irradiation induced damage (An et al., 2005), while a SMUG1-specific function appears to be the removal of 5-hydroxymethyluracil arising from thymine oxidation (Kavli et al., 2002), an activity that may also be important in RNA repair (Jobert et al., 2013). As noted before (chapter 2.4.2), UNG2 appears to have non-redundant functions in the adaptive immune system, i.e. in SHM and CSR, where SMUG1 can only partially compensate for loss of UNG2 when overexpressed (Di Noia et al., 2006). There is no evidence for TDG and MBD4 participating in this pathway although this cannot be strictly excluded. There might, however, be additional, yet unidentified, UDG-specific functions.

In contrast to SMUG1, no or only backup functions in the repair of G•U mismatches have been observed for TDG and MBD4, although this is not yet completely settled. Instead of performing classical BER, these two UDGs appear to have acquired specialized functions. TDG indeed has features untypically for a BER glycosylase. It has an extremely low enzymatic turnover and TDG-deficient cells lack typical repair phenotypes such as hypersensitivity to genotoxic agents but do show hyper-resistance to 5-FU. Instead, as thoroughly discussed in the previous chapter, TDG has a major role in the epigenetic regulation of gene expression at the levels of DNA demethylation and the regulation of histone marks. This function most likely cannot be efficiently compensated for by other glycosylases as TDG is the only known glycosylase that is essential for embryogenesis (Cortázar et al., 2011; Cortellino et al., 2011; Jacobs & Schar, 2012).

Although MBD4 with its methyl-CpG binding domain was also implicated in the regulation of gene expression, the phenotype of *MBD4* knockout is much less severe compared to that of *TDG* knockouts (Cortázar et al., 2011; Cortellino et al., 2011; Millar et al., 2002). MBD4 was associated with 5mC demethylation-mediated gene activation at specific promoters of hormone-stimulated genes (Kim et al., 2009) and overexpression of MBD4 and AID led to general DNA demethylation in zebrafish embryos (Rai et al., 2008). Moreover, MBD4 was shown to repress transcription in a HDAC-dependent manner (Kondo et al., 2005). MBD4 activity seems, however, not restricted to target-specific DNA demethylation, and additional functions related to DNA MMR, apoptosis and chromosome stability have been reported (Abdel-Rahman et al., 2008; Cortellino et al., 2003; Sansom et al., 2004; Sansom et al., 2003). Finally, MBD4 has classical BER functions. It is capable of removing the deamination products of cytosine and 5mC, i.e. uracil and thymine, respectively, mispaired with guanine, and it contributes to mutation avoidance at CpG sites (Hendrich et al., 1999; Millar et al., 2002). However, as it is the case for TDG, MBD4-mediated repair is rather toxic under saturating conditions induced by 5-FU exposure (Cortellino et al., 2003).

With respect to a potential temporal separation of function, I would like to point out the cell cycle-dependent regulation of TDG and UNG2. While UNG2 expression peaks in S-phase, TDG is eliminated

at the beginning of S-phase via ubiquitination and subsequent proteasomal degradation and re-expression is only observed in G2 phase (Hagen et al., 2008; Hardeland et al., 2007). Importantly, incomplete degradation of TDG during S-phase interferes with S-phase progression and cell proliferation (Hardeland et al., 2007). In line with this, TDG degradation during S-phase requires interaction between TDG and PCNA (Shibata et al., 2014; Slenn et al., 2014). This cell cycle-dependent regulation might reflect the need of highly efficient UNG2 at the moving replication fork, while TDG with its low turnover could be beneficial outside of the replication context. This might also be true for the *E. coli* ortholog of TDG, Mug, inactivation of which causes a mutator phenotype specifically in stationary phase cells (Mokkapati et al., 2001). Although TDG might be important for removal of G•U mismatches outside of S-phase, the lack of mutation phenotype in the *TDG*<sup>-/-</sup> knockout indicates that UNG2 can compensate for this function. Taken together, UNG2 and SMUG1 are important for the bulk of uracil removal with possible minor contributions by TDG and MBD4. The latter two appear to have evolved additional functions in the epigenetic regulation of gene expression. Further research will be required to more clearly identify and separate the function of the different UDGs.

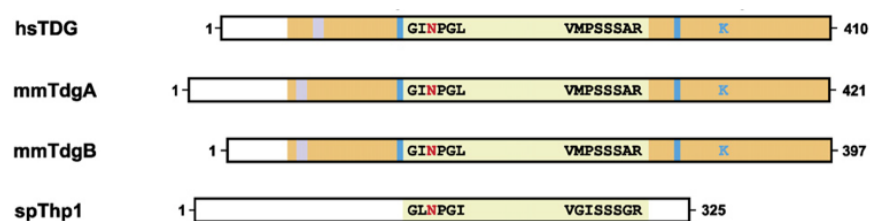
#### **2.4.4 *S. pombe* as a model organism to study Base Excision Repair and Chromatin Regulation**

Yeasts are widely used for investigating DNA repair mechanisms and finding potential human drug targets (Kanamitsu & Ikeda, 2010; Kelley et al., 2003). They share many proteins and mechanisms with multicellular organisms, yet they have relatively small genomes, which can easily be manipulated for genetic studies. We decided to use the fission yeast model to investigate the functional separation of UDGs because, in contrast to budding yeast, *S. pombe* expresses orthologs of both UNG (Ung1) and TDG (Thp1) while it lacks SMUG1 and MBD4 orthologs. Therefore, it constitutes a simplified model to analyze the relative contribution of Ung1 and Thp1 to DNA repair and gene regulation. Another advantage is the presence of chromatin regulatory mechanisms that are similar to those in humans. For instance, heterochromatin formation in mammals involves methylation of histone 3 lysine 9 (H3K9) and subsequent binding of heterochromatin protein 1 (HP1). This mechanism is conserved in *S. pombe* that expresses the HP1 ortholog Swi6 (Olsson & Bjerling, 2011). Moreover, *S. pombe* is proficient in RNA interference (RNAi)-mediated heterochromatin assembly (Bühler, 2009). The presence of epigenetic mechanisms, even if they are not fully conserved with those in multicellular organisms, is clearly advantageous for the investigation of potential Thp1 functions in gene regulation. The lack of CpG methylation in *S. pombe* (Antequera et

al., 1984; Capuano et al., 2014; Wilkinson et al., 1995) helps to address potential DNA methylation-independent mechanism implicated by specific phenotypic features of the *TDG*<sup>-/-</sup> knockout mouse (Chen et al., 2003; Cortázar et al., 2011; Goodman & Smolik, 2000; Madabushi et al., 2013; Sjolund et al., 2013; Um et al., 1998; J. Zhou et al., 2008). In this respect, fission yeast is a valuable model to study UDG functions, complementing the available mammalian cell lines or mouse models.

Although many BER enzymes are conserved between yeast and mammals, there are prominent differences in enzyme functionality and repertoire (Kelley et al. 2003; Boiteux & Jinks-Robertson, 2013). This includes altered substrate glycosylases and different enzymes accomplishing AP site cleavage. Actually, the bi-functional DNA glycosylase Nth1 is the main fission yeast enzyme performing AP site incision by its intrinsic lyase activity, whereas the major AP endonuclease Apn2 is mostly responsible for the subsequent conversion of the blocking 3'-end to a processible 3'-OH (Kanamitsu & Ikeda, 2010). Interestingly, due to a nonsense mutation in the laboratory fission yeast strains, Apn1 has no or only little function in AP site incision. This mutation is not found in other, naturally occurring *S. pombe* strains, suggesting that Apn1 contributes to AP site incision activity in these strains (Laerdahl et al., 2011).

Ung1 is highly conserved between fission yeast and human showing 51% sequence identity. Its predominant localization to the nucleus (Elder et al., 2003) suggests a functional similarity to the nuclear human UNG2. So far biochemical assays proved Ung1 activity on uracil, which could be inhibited by the classical UNG inhibitor Ugi. Interestingly, overexpression of the glycosylase induced elevated mutation rates, checkpoint dependent cell cycle delay, cell death and AP site accumulation. The observed phenotypes did not relate to catalytic activity of Ung1, but rather reflect its affinity to AP site (Elder et al., 2003), a feature known for many glycosylases (Jacobs & Schar, 2012). However, no obvious phenotype was yet reported for Ung1-deficient cells. In *S. cerevisiae*, Ung1 deletion led to a 20-fold increase in C to T transition, probably reflecting the lack of alternative uracil glycosylases (Impellizzeri, 1991).



**Figure 8. Comparison of TDG homologs in human, mouse and fission yeast.** Schematic alignment of TDG homologs. The highly conserved catalytic core is shown yellow. The active site motif GINPGL is slightly different in *S. pombe* (GLNPGI) without altering the proposed catalytic asparagine (N). The second motif contributes to the sequence selectivity and is less conserved. The N-terminal and C-terminal parts are less conserved than the catalytic domain. K, sumoylation site; blue bar, SUMO interaction motif; violet bar, cryptic AT-hook. Figure adapted from (Cortázar et al., 2007).

Thp1 was attributed to the family of MUG glycosylases, thus presenting a homolog to human TDG (Figure 8). There is moderate amino acid sequence conservation between the human and fission yeast TDG/Thp1 with the highest similarity in the catalytic core domain, which forms the typical  $\alpha/\beta$ -fold structure of MUG proteins. A slight variation in the active site motif (GLNPGI) is present in Thp1 as compared to TDG (GINPGL), however, the essential catalytic asparagine of TDG remains unchanged. This asparagine is required to initiate the hydrophilic attack on the N-glycosidic bond between the target base and the corresponding sugar. An additional, less conserved sequence motif contributing to sequence-selectivity, is not well conserved in fission yeast. Similarly, the N- and C-terminal parts are less conserved than the catalytic core and vary between species in terms of sequence and size. The SUMOylation site responsible for releasing TDG from AP sites appears absent in Thp1. Functionally, Thp1 is active on a broad range of substrate, including uracil, 5-FU, 5-hydroxyuracil and deaminated purines, i.e. xanthine, oxanine, hypoxanthine. Unlike its human counterpart, Thp1 has no activity towards T•G mismatches, one possible explanation being the lack of 5mC in the yeast genome, whose deamination is a main source of this mispair. Strikingly, contrary to TDG, Thp1 is not mismatch-specific and also excises its substrates from single-stranded DNA, although the catalytic efficiency is highest for substrates base-paired with guanine. It is also important to note that TDG and Thp1 both are product-inhibited and thus not released from their product AP site (Dong et al., 2008; Hardeland et al., 2003; Hardeland et al., 2000).

### 3 Aims of the Thesis

The purpose of this PhD thesis was to establish and use refined yeast models to address specific functional and mechanistic aspects of topoisomerase 1 (Top1) and uracil DNA glycosylase (UDG) functions in the maintenance of DNA structure and the encoded (epi-)genetic information.

The ribosomal DNA (rDNA) in budding yeast is constantly transcribed at high levels, also during S-phase. Unwinding of the DNA duplex during DNA replication and transcription creates DNA torsion that requires dissociation along the chromosome or relaxation by topoisomerases. Topological constraints imposed by the chromosomal structure itself, but also by the reported anchoring of the rDNA to the inner nuclear membrane by Fob1, would prevent dissociation of torsion and thus lead to an accumulation of torsional stress in the rDNA, unless relieved by topoisomerases. Site-specific Top1-dependent nicks were observed near the ribosomal replication fork barrier (rRFB), suggesting a specific function of this protein at this block site. However, the mechanism underlying the specific occurrence of DNA nicks at these sites as well as the implicated coordination of the topoisomerase remained elusive.

**A first aim of my thesis was the identification of factors coordinating Top1 activity in the rDNA and to study their role in rDNA maintenance. Also, I wanted to test the hypothesis that Top1 activity is related to the previously observed non-canonical DSBs.**

UDGs are important for genome maintenance due to their classical role in BER-mediated uracil excision. In addition, our and other groups found additional roles for TDG in the epigenetic regulation of gene expression, particularly during cell differentiation. This epigenetic TDG function involves regulation of chromatin at the level of DNA methylation and histone modifications. Due to the expression of four nuclear glycosylases with undefined redundant functions in mammals, determining the contribution of individual enzymes to classical uracil excision repair and separating their specialized biological functions, such as in epigenetic gene regulation, is difficult.

**The second aim of my thesis was to dissect the biological functions of the two UDGs expressed in *S. pombe* and to assess specifically a potential role of the TDG homolog Thp1 in gene regulation. Pursuing from genetic experiments performed by a former collaborator, I characterized novel aspects of UDG functions in the repair of damaged bases and focused on the identification of potential epigenetic functions of Thp1. This allowed me to evaluate the use of *S. pombe* as a complementary model organism for the study of UDGs in epigenetic regulation of gene expression.**



## 4 Results

### 4.1 Reversible Top1 Cleavage Complexes are Stabilized Strand-Specifically at the Ribosomal Replication Fork Barrier and Contribute to Ribosomal DNA Stability

(Published manuscript, see Appendix I)

The DNA has to be faithfully replicated prior to each cell division. During this process the moving replication fork (RF) encounters numerous obstacles that interfere with its progression. Eventually, RFs stall and then require stabilization to allow for removal of the block and/or for resumption of replication. Alternatively, the RF merges with another fork to terminate replication. In addition to accidental fork blocks, programmed RFBs mediate RF pausing at defined places. Such programmed blockage is found at the 3' end of a highly transcribed ribosomal RNA (rRNA) gene in several eukaryotes. The best-studied example is the ribosomal replication fork barrier (rRFB) of *S. cerevisiae*, the function of which requires sequence-specific binding of Fob1. Fob1 does not only mediate fork stalling, but is also crucial for regulating rDNA recombination (Tsang & Carr, 2008) and for anchoring of the rDNA repeats to the inner nuclear membrane (Chan et al., 2011; Huang et al., 2006; Mekhail et al., 2008). The repetitive structure of the rDNA makes it prone to unequal homologous recombination, which is why the control of recombination is crucial for rDNA stability (Kobayashi, 2006). Previously, DNA DSBs have been shown to occur at the rRFB (Burkhalter & Sogo, 2004; Kobayashi et al., 2004; Weitao et al., 2003), yet, they seem to be of non-canonical nature as they are not processed by the classical DSB repair pathways (homologous recombination, non-homologous end-joining) (Fritsch et al., 2010).

Topoisomerases are important for the relaxation of DNA torsional stress and for rDNA stability (Christman et al., 1988; Kim & Wang, 1989; Zhu & Schiestl, 2004). Top1-dependent DNA nicks were mapped to the rRFB region of budding yeast (Vogelauer & Camilloni, 1999), but the mechanism of their positioned occurrence remained elusive. Therefore, we sought to investigate the interaction of Top1 with the rRFB, taking advantage of the feature of topoisomerases to form usually short-lived covalent reaction intermediates with the DNA, known as topoisomerase cleavage complexes (Top1ccs). I mapped Top1cc enrichment across the rDNA unit using chromatin immunoprecipitation (ChIP) without crosslinking the DNA (Takahashi et al., 2011). Under these conditions, only DNA-bound Top1ccs are detected, but no free Top1. We found a high Fob1-dependent enrichment of Top1ccs at the rRFB region, indicative of a Fob1 requirement for Top1cc formation. This enrichment did not require the rDNA context as Top1ccs also accumulated at an ectopically located rRFB (eRFB).

Despite being sufficient for Fob1 recruitment and RFB activity, we found the eRFB sequence insufficient to promote anchoring to the inner nuclear membrane as inferred from its subcellular localization. Thus, we could show that Top1cc formation requires Fob1, but neither the rDNA context, nor nucleolar localization, nor perinuclear anchoring.

I was further interested in whether Top1ccs only accumulate at stalled RFs. I synchronized cells and found similar Top1cc enrichment in S- and G1-phase cells. Moreover, we introduced the unidirectional eRFB in different orientations to compare Top1cc enrichment at blocking and permissive eRFBs. We found that Top1ccs accumulate at eRFBs independent of their orientation, further suggesting that Top1 is recruited and/or stabilized by Fob1 and not by stalled RFs.

To exclude the possibility of Top1ccs being damaged and incapable of religating the nick, I compared Top1cc enrichment of wild-type cells with that of cells depleted for factors involved in the repair of irreversible Top1ccs, i.e. *rad1Δtdp1Δ* and *mus81Δtdp1Δ*. Compromising the repair of irreversible Top1cc complexes did not alter Top1cc levels at the eRFB, pointing at a stabilization of reversible Top1ccs.

To further elucidate the mechanism of Top1cc stabilization, we performed standard ChIP experiments (crosslinked) in cells expressing Fob1-9myc and found that Fob1 association to the rRFB is reduced in Top1-deficient cells, indicating that Top1 and Fob1 stabilize each other at the rRFB. Furthermore, deletion of the nucleolar protein Tof2 that interacts with both Fob1 and Top1 (Huang et al., 2006; Park & Sternglanz, 1999) reduced Fob1 association to the eRFB while it completely abolished Top1cc enrichment. These observations pointed at the formation of a common complex at the rRFB, consisting of at least Fob1 and Tof2, that is important for Top1cc formation.

Finally, we showed that most of the previously identified rRFB-associated DSBs are in fact Top1-dependent. Most likely these breaks reflect DNA melting between the single-stranded Top1 nicks at the rRFB and stalled RFs under the applied experimental conditions. Such melting would give rise to molecules that appear as DSBs on Southern blots. Whereas the previously mapped Top1 nicks at the rRFB are formed throughout the cell cycle, the DSBs were only observed in S-phase. This is in agreement with Top1 nicks being persistently present, as *in vitro* conversion to a DSB can only occur when a second “DNA break” is present at the RF.

Taken together, we could show that Top1ccs are stabilized in a complex containing Top1, Fob1 and Tof2 at rRFBs independent of the ribosomal context and rDNA anchoring to the inner nuclear membrane.

#### Contribution:

I designed and conducted most of the experiments together with Olivier Fritsch. I constructed Top1-13myc strains and performed all non-crosslinked Top1cc ChIPs presented in this study. I performed

CPT sensitivity assays to confirm published hypersensitivities in *rad1Δtdp1Δ* and *mus81Δtdp1Δ* strains. Further I conducted 2D gel and 1-dimensional (1D) gel electrophoresis and quantified DSB levels at the eRFB. I also used pulse-field gel electrophoresis (PFGE) to determine the rDNA repeat stability in the different strains. Finally, I prepared figures and wrote the manuscript together with Olivier Fritsch.

## 4.2 Uracil Repair Causes DNA Glycosylase-Dependent Genome Instability

(Manuscript in preparation, see Appendix II)

Uracil in DNA derives from spontaneous cytosine deamination or from misincorporated dUMPs during replication, the former being mutagenic if left unrepaired. Misincorporated uracils are not mutagenic, but they may interfere with transcription factor (TF) binding (Brégeon et al., 2003; Rogstad et al., 2002; Verri et al., 1990). To counteract the adverse effects of uracil in DNA, uracil DNA glycosylases (UDGs) are in place to recognize and excise uracil for the replacement with a canonical base in a DNA repair process (base excision repair, BER). Five UDGs are expressed in human cells (UNG1, UNG2, SMUG1, TDG, MBD4). UNG2 appears to be the predominant enzyme for the removal of both misincorporated uracil and deaminated cytosines, although SMUG1 clearly contributes to the latter activity (Nilsen et al., 2001). Remarkably, epigenetic roles rather than the maintenance of DNA sequence integrity have been ascribed to TDG and MBD4 (Sjolund et al., 2013). Unlike mammalian cells, *S. pombe* expresses only two UDGs, Ung1 and Thp1, thus offering a reduced complexity for a functional dissection of these UDGs. Importantly, certain epigenetic mechanisms of gene regulation are conserved in *S. pombe* (Bühler, 2009; Olsson & Bjerling, 2011; Millar & Grunstein, 2006), while DNA methylation is absent (Antequera et al., 1984; Capuano et al., 2014; Wilkinson et al., 1995). This provides an opportunity to investigate putative epigenetic features of Thp1, as implicated by TDG knockout in mouse (Chen et al., 2003; Cortázar et al., 2011; Goodman & Smolik, 2000; Madabushi et al., 2013; Sjolund et al., 2013; Um et al., 1998; Zhou et al., 2008), i.e. DNA methylation-independent mechanisms of gene regulation.

To address the functions of the two *S. pombe* UDGs, we first compared the uracil excision capacity of cell-free extracts prepared from wild-type, *ung1Δ*, *thp1Δ* and *thp1Δ ung1Δ* cells. Using base removal assays, we detected uracil excision activity from U•G and U•A mismatches only in the presence of Ung1. In the absence of Ung1, Thp1 was not able to constitute detectable uracil excision unless overexpressed, suggesting that Ung1 is the predominant UDG for uracil removal also in *S. pombe*.

Contradictory to the biochemical evidence, however, uracil accumulated in *S. pombe* cells only when both UDGs were absent. To further analyze the biological function of Ung1 and Thp1, we determined spontaneous mutation rates in fluctuation tests addressing mutation at two independent genomic loci. Mutation rates were synergistically increased in the *thp1Δ ung1Δ* double mutant while either single mutant showed no effect on the mutation rate. Given the accumulation of uracils and mutations in the genome of UDG-deficient cells, we concluded that Thp1 and Ung1 compensate for each other in uracil removal and mutation avoidance *in vivo* despite the undetectable Thp1 activity on uracil *in vitro*. Determination of mutation spectra further revealed a specific synergistic increase of C to T transitions, consistent with deaminated cytosines and, hence, G•U mismatches being the responsible substrate. While these results pointed at redundant functions of Thp1 and Ung1 in uracil removal, we observed increased transversion rates solely in Thp1-deficient cells suggesting additional separate functions for Thp1, in line with the broad substrate spectrum observed for this protein (Hardeland et al., 2003).

Mammalian TDG was shown to sensitize cells to treatment with the uracil analog 5-FU (Kunz et al., 2009). We tested the impact of Thp1 on 5-FU sensitivity in *S. pombe* and found that Thp1 expression is detrimental during 5-FU treatment, similar to the situation in mammals. While absence of Thp1 increased cellular resistance towards 5-FU, Thp1 overexpression greatly enhanced 5-FU cytotoxicity. Moreover, Thp1 overexpression induced DNA fragmentation and increased spontaneous mutation rates (data not shown). Together, these findings are in line with Ung1 and Thp1 having additional, non-overlapping functions associated with a qualitatively different outcome of the repair process. The low turnover of Thp1 and the observed product inhibition suggests a slower hand-over of AP sites to downstream BER enzymes by Thp1 than by Ung1 (Hardeland et al., 2003). In case of BER saturation, labile AP sites would thus accumulate and could confer cytotoxicity particularly during Thp1-dependent repair. To test this possibility, we expressed human AID in *S. pombe* to trigger deamination of cytosine to uracil. AID expression in *S. pombe* induced cell death that was most pronounced in wild-type and *ung1Δ* cells, but significantly decreased in *thp1Δ* strains. This result is consistent with the concept that Thp1-mediated uracil repair is slower or less productive than that initiated by Ung1. Having shown that Thp1 expression has adverse effects during 5-FU treatment or upon AID expression, we reasoned that the low turnover of Thp1 generates AP sites with an increased lifetime. AP sites might cause occasional DNA breakage that in turn could target recombination-dependent repair. We thus determined spontaneous mitotic recombination rates in wild-type, *ung1Δ*, *thp1Δ* and *thp1Δ ung1Δ* cells and found decreased rates only in Thp1-deficient cells, i.e. in *thp1Δ* and *thp1Δ ung1Δ* mutant strains. In fact, 60% of the detected spontaneous recombination events relied on the expression of Thp1, but not Ung. We further tested the contribution of Thp1 to damage-induced recombination rates and exposed wild-type and *thp1Δ* cells

to a non-lethal X-ray dose. Strikingly, we only detected an X-ray-induced increase in recombination rates in the wild-type strain, but not in the Thp1-deficient strain. These results further indicate that Ung1- and Thp1-dependent DNA repair have qualitatively different repair outcomes. In contrast to the fast and mostly error-free repair mediated by Ung1, Thp1-dependent repair appears to be slow and non-productive. We assume that this reflects the slow dissociation from its product AP site, thereby preventing further repair. As the X-ray dose applied for the recombination assay presumably generates predominantly oxidative damage, Thp1 might also binds to AP sites deriving from hydrolytic DNA deamination or the base excision by other DNA glycosylases. However, the side-by-side evolution of Ung1 and Thp1 suggests additional cellular functions for Thp1 beyond classical DNA repair.

During the last years, human TDG turned out to be important in the epigenetic regulation of gene expression (Sjolund et al., 2013). To test for a similar function of the fission yeast protein, we compared genome-wide RNA expression profiles of wild-type and Thp1-deficient cells, yet, we could not detect pronounced patterns of dysregulation in the absence of Thp1. Interestingly, however, most of the analyzed mRNAs tended to be less expressed in *thp1Δ* cells (92% of all analyzed mRNAs). Principal component analysis of the mRNAs further revealed that deletion of *thp1<sup>+</sup>* increases the variability between the analyzed replicates. The decreased variability was also reflected by the overall increased standard deviations (SDs) of the transcript levels in Thp1-deficient compared to wild-type cells. Given that each triplicate consists of a pool of RNAs from three independent strains (isolated from a cross), the variation between single strains might be even higher, however, this remains to be addressed. Therefore, Thp1 might have additional roles in maintaining an transcriptionally active chromatin status. However, further investigation is required to confirm an epigenetic function of Thp1.

Taken together, we were able to show that despite their overlapping function in uracil excision and mutation avoidance, the different repair kinetics of Ung1 and Thp1 appear to result in different repair outcomes. The high fidelity Ung1 enzymes contributes to classical DNA repair by efficiently inducing BER of DNA lesions. In contrast, the high affinity of Thp1 to AP sites seems to prevent or delay proper repair of the AP site, thus explaining the mutagenic and recombinogenic repair outcome by Thp1. In addition, we presented evidence for a function of Thp1 in gene regulation.

#### Contribution:

I assembled and evaluated data from the biochemical, genetic and pulsed-field gel-electrophoresis (PFGE) experiments (base release assay, genomic uracil incorporation, mutation rate, 5-FU- and AID-sensitivity) that were performed by Marc Bentele. For the genome-wide expression analysis, I performed genetic crosses and RNA isolation together with Olivier Fritsch. After the initial

bioinformatic analysis, I further analyzed the transcriptome data and prepared all figures except for the principal component analyses. I planned Thp1-13myc ChIP sequencing experiment, constructed and checked the Thp1-13myc-tagged strain for its uracil excision activity and tested different ChIP conditions. I confirmed pull-down of Thp1-13myc and ensured that the ChIP conditions are suitable to reproduce ChIP results of a published strain. I coordinated next generation sequencing and bioinformatic analysis of the data and explored possible candidate genes. I wrote the manuscript and adjusted figures of the uracil excision part.

## **4.3 Additional Results**

### **4.3.1 Estrogen Receptor $\beta$ Regulates Epigenetic Patterns at Specific Genomic Loci Through Interaction with Thymine DNA Glycosylase**

(Manuscript in Preparation, see Appendix III)

The two estrogen receptors (ERs) ER $\alpha$  and ER $\beta$  belong to the family of nuclear receptors. The two transcription factors (TFs) are important for normal development, reproducibility and for the functionality of the immune, cardiovascular, musculoskeletal and central nervous systems. The ERs are activated by a ligand, the main endogenous molecule being 17 $\beta$ -estradiol (E2). Additionally, environmental chemicals such as polycyclic aromatic hydrocarbons, phthalates and pesticides are known to act as ER ligands. Upon ligand binding, the activated ERs bind to estrogen response elements (EREs) in the genome to regulate expression of target genes (Heldring et al., 2007). This regulation involves several co-activators, e.g. chromatin remodelers (Métivier et al., 2003). Interestingly, nuclear receptors, including retinoic acid receptor (RAR), retinoid X receptor (RXR), ER $\alpha$  and ER $\beta$ , have been implicated in the epigenetic regulation of gene expression (Benbrook et al., 2014; Magnani & Lupien, 2014; Rüegg et al., 2011; Tammimies et al., 2012, Kangaspeska et al., 2008; Kim et al., 2009; Marques et al., 2013; Métivier et al., 2008; Thomassin et al., 2001). Recently, ER $\beta$  was shown to modulate 5mC methylation of a single CpG site in the glucose transporter 4 (*GLUT4*) gene, suggesting a function in regulation of DNA demethylation at this site (Rüegg et al., 2011).

DNA methylation is important for the regulation of gene expression and it undergoes gross remodeling during cellular differentiation to suppress pluripotency and activate cell-specific genes. During the last years, the TET proteins turned out as a major player in the process of active DNA demethylation and TDG plays an important role in this pathway. By excising TET-mediated 5mC oxidation products, the TET-TDG pathway supposedly maintains certain gene promoters in an unmethylated state and mediates demethylation of others in the course of cellular differentiation

(Cortázar et al., 2011; Cortellino et al., 2011). How demethylation activities are targeted to specific genes remains, however, unclear. In this study, we wanted to investigate the function of ER $\beta$  in the regulation of DNA methylation and to address a potential cooperation of ER $\beta$  with TDG.

First, we analyzed the effect of ER $\beta$  on DNA methylation on a genome-wide level. We used mouse embryonic fibroblasts (MEFs) deriving from ER $\beta^{+/+}$  (wild-type) and ER $\beta^{-/-}$  ( $\beta$ erko) mice (Rüegg et al., 2011) and analyzed their genome-wide methylation status using reduced representation bisulfite sequencing (RRBS) (Meissner et al., 2008). By considering only CpGs that are unmethylated (less than 20% of methylated reads) in one cell line and methylated (more than 80% of methylated reads) in the other, 8071 differentially methylated positions (DMPs) were identified. Of these DMPs, 6016 were hypo- and 2055 hypermethylated in  $\beta$ erko MEFs and both subsets were associated with developmental genes. The observed methylation differences were not restricted to single CpGs, but spreaded across multiple sites. Methylation correlated inversely with expression of the nearby genes. By comparing our data with publicly available datasets of genome-wide chromatin marks we found that hypomethylated DMPs significantly overlap with characteristic marks of repressed or poised genes, indicating that associated genes are poorly expressed in wild-type cells while they are more active in  $\beta$ erko MEFs. We tried to complement the methylation phenotype by re-introducing ER $\beta$  into  $\beta$ erko MEFs, however, we could restore wild-type methylation only in a subset of hypermethylated DMPs, but not in the remaining hyper- or hypomethylated genes.

Next, we were interested how DMP-associated genes are expressed during cell differentiation. Therefore, we analyzed expression of selected genes in wild-type mouse embryonic stem cells (ESCs) and MEFs. Interestingly, regions of hypo- and hypermethylated DMPs have decreased and increased transcription, respectively, in differentiated MEFs as compared to ESCs, indicating that ER $\beta$  is involved in the regulation of methylation during differentiation. We also addressed physical association of ER $\beta$  to DMP regions in chromatin immunoprecipitation (ChIP) experiments. While ER $\beta$  was enriched at all analyzed sites in ESCs, the situation in MEFs was less clear. Thus, ER $\beta$  appears to have important regulatory functions in undifferentiated ESCs, during differentiation and probably also in differentiated MEFs.

As ER $\alpha$  interacts with TDG (Chen et al., 2003), we checked the possibility of ER $\beta$  also interacting with TDG. Indeed, we observed interactions between ER $\beta$  and TDG using different methods. In far Western blot analysis, recombinant TDG specifically bound to immobilized ER $\beta$ . For yeast-two hybrid assays, we fused ER $\beta$  to the Gal4 DNA binding domain (BD) and TDG to the Gal4 activation domain (AD). Cells co-expressing ER $\beta$ -BD and TDG-AD constructs induced expression of the reporter gene as revealed by growth on selective media. This specific interaction was detectable in two yeast strains harboring different reporter genes. We further tried to identify the ER $\beta$  domain responsible for interaction. Yet, no detectable reporter gene expression was observed between TDG and either

tested ER $\beta$  domain, suggesting that the full-length ER $\beta$  is required to establish optimal interactions with TDG.

Given the physical interaction between TDG and ER $\beta$ , we analyzed the effect of TDG on the expression of ER $\beta$ -induced gene expression. In the presence of ER $\beta$ , expression of the Luciferase reporter gene was increased. TDG further enhanced Luciferase transcription irrespective of the presence or absence of the ligand. In line with a co-regulation of transcription, TDG was enriched at ER $\beta$ -regulated genes and TDG association was lost in  $\beta$ erko cells at two genes tested.

TDG is thought to process mainly 5caC and 5fC during active demethylation, consistent with an accumulation of both demethylation intermediates in *TDG* knockout ESCs (Kohli & Zhang, 2013; Shen et al., 2013; Song et al., 2013). Assuming that ER $\beta$  recruits TDG to regulated genes, 5caC and 5fC would accumulate at DMPs in the absence of TDG. We compared our RRBS data with published genome-wide 5fC localization data (Song et al., 2013) and found that 32% and 46% of hyper- and hypomethylated DMPs, respectively, overlap with 5fC in *TDG*<sup>-/-</sup> cells.

Taking this data together, we propose a dual model: Regions that are hypermethylated upon ER $\beta$  deletion are normally bound by ER $\beta$  that in turn recruits the TET-TDG DNA demethylation system. This stimulates gene expression and, additionally, TDG would counteract DNA hypermethylation by removing demethylation intermediates. On the other hand, hypomethylated genes are suppressed upon differentiation. The remodeling of DNA methylation during this process requires TDG and its absence leads to 5fC and 5caC accumulation. As these marks were suggested to be poorly recognized by the maintenance methyltransferase Dnmt1 (Wu & Zhang, 2011), this results in passive demethylation during differentiation. Together with an ER $\beta$ -dependent TDG recruitment, this explains well the observed hypomethylation in ER $\beta$ -deficient MEFs.

#### Contribution:

I planned, adapted and conducted the yeast two-hybrid assays using the Y187 yeast strain to confirm the interaction between ER $\beta$  and TDG and to test for interactions between different ER $\beta$  domains and TDG. I prepared the figures and contributed to writing of the yeast-two hybrid relevant parts of the manuscript.



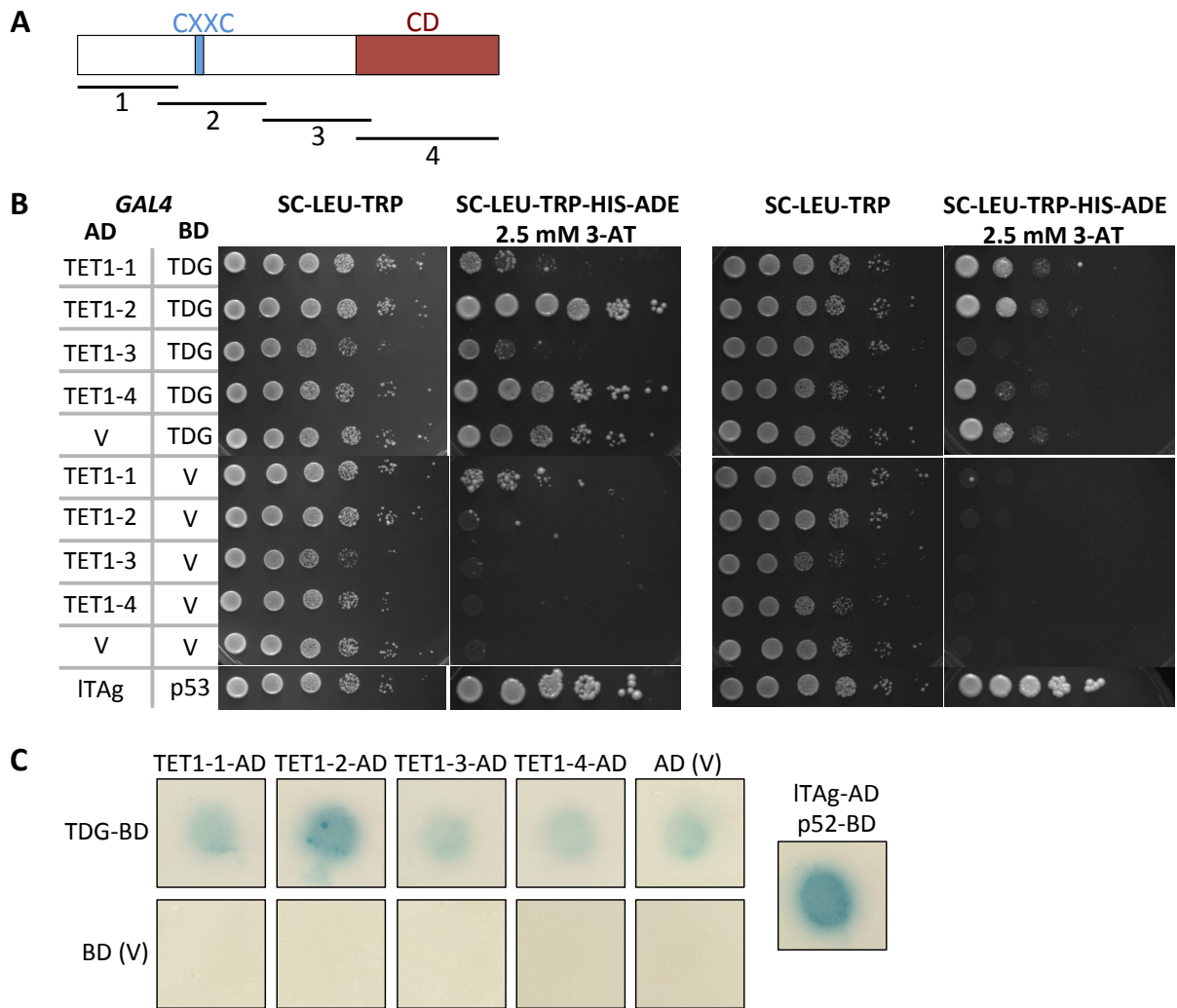
### 4.3.2 Physical Interactions between Murine TDG and TET1

(unpublished data)

The mammalian TDG is thought to take part in active DNA demethylation during embryonic development. Indeed, TDG is required at certain gene promoters to maintain an unmethylated state or for demethylation during cell differentiation at others (Cortázar et al., 2011; Cortellino et al., 2011). TDG was suggested to cooperate with the TET proteins for active demethylation. TET enzymes are capable of stepwise oxidation of 5mC to 5hmC, 5fC and 5caC *in vitro* (Kohli & Zhang, 2013) and detection of this oxidation products *in vivo* indicates a biological relevance for TETs in active DNA demethylation (Ito et al., 2011; Tahiliani et al., 2009). Interestingly, TDG is the only DNA glycosylase with 5fC and 5caC excision activity (He et al., 2011; Maiti & Drohat, 2011), suggesting it acts downstream of the TET proteins in active DNA demethylation. Consistently, the inactivation of TDG leads to an accumulation of the 5fC and 5caC demethylation intermediates in genomic DNA of embryonic stem cells (ESCs) (Kohli & Zhang, 2013; Shen et al., 2013; Song et al., 2013).

Based on this model, Alain Weber from our group tested for a potential interaction between murine TDG and TET1 by pull-down assays. He used TDG and TET1 carrying a C-terminal GST and HIS tag, respectively. When he immobilized the TDG-GST on glutathione beads and applied free TET1-HIS protein, he could enrich TDG in the presence of TET1. Interactions were also seen when TET1-HA was used as the bait protein that was immobilized by nickel-nitrilotriacetic acid (Ni-NTA) and TDG-GST was added as the prey protein (Alain Weber, unpublished data). During the establishment of TET1 purification, he further collected evidence that expression of recombinant TET1 is more stable when co-expressed with TDG (Alain Weber, unpublished data).

To corroborate an interaction between TDG and TET1, I performed yeast-two hybrid assays in the *S. cerevisiae* strain AH109 (Clontech) that harbors two Gal4-inducible reporter genes *HIS3* and *ADE2*. One protein is fused to the Gal4 activation domain (AD) while the other is fused to the *Gal4* binding domain (BD). Interaction of the two co-expressed proteins brings the Gal4 BD and AD in close proximity, thereby activating expression of the reporter genes as scored by growth on selective medium lacking histidine and adenine. Because of the huge molecular size of TET1, but also to be able to map the domains responsible for interaction, I engineered four overlapping domains of TET1 (1–4, Figure 9A), with TET1-2 containing the CXXC zinc-binding domain and TET1-4 consisting of the catalytic domain, and tested their interaction with full-length TDG. As expression of TDG fused to the AD results in the formation of large slow-growing cells independent of TET1 expression, we pursued the yeast-two hybrid experiments with TDG fused to the BD only.



**Figure 9. Yeast-two hybrid assay to detect interactions between full-length TDG and TET1 domains. (A)** Scheme of TET1 domains cloned into the Gal4 activation domain (AD): TET1-1 (bp 1-1473), TET1-2 (bp 1189 – 2793), TET1-3 (bp 2663 – 4208), TET1-4 (bp 4098 – 6171). CXXC and catalytic domain (CD) are indicated. **(B)** Yeast-two hybrid assays using HIS3 and ADE3 reporter genes. Serial dilution of strains co-expressing TET1 domains fused to the AD and the TDG fused to the Gal4 binding domain (BD) or the respective vector controls (V). The large T antigen (ITAg) and p53 fused to the AD and BD, respectively serve as a positive control. Two replicates are shown. **(C)** Yeast-two hybrid assay using the reporter gene  $\beta$ -galactosidase. Colony lift assays were performed on cells transformed with constructs as in (B) and incubated with X-gal for 17 h.

I performed drop tests from two independent transformations of the AH109 strain (Figure 9B) as well as one pilot experiment using another yeast strain (Y187) that harbors  $\beta$ -galactosidase as the reporter gene (Figure 9C). In this strain, appearance of blue color indicates an interaction between tested proteins. While auto-activation was consistently observed in cells expressing TDG fused to the AD, most of the cells expressing the BD vector control lacked any signal. Generally, signals were variable and we occasionally observed less growth of cells co-expressing TDG-BD and the TET1-1, TET1-3 or TET1-4 domains fused to the AD. However, we saw a stronger interaction signal for cells co-expressing TDG-BD and TET1-2-AD as compared to the corresponding vector controls. This indicates that TET1 interaction with TDG might be mediated by residues within the TET1-2, e.g. the CXXC domain. Noteworthy, it was previously observed that TDG being expressed in yeast-two hybrid

assays without an interaction partner shows stronger residual growth on selective media than when co-expressed with another non-interacting protein. Hence, the residual growth of the cells expressing TDG-BD only, is likely to be an artifact. Thus, together with the evidence from pull-down experiments, these data provide solid evidence for an interaction of TET1 with TDG and that this interaction might involve the CXXC domain or other residues within the second TET1 domain (TET1-2).

### Material and Methods:

The Matchmaker™ yeast-two hybrid system (Clontech) was used. I inserted a synthetic fragment coding for a FLAG tag (between the BD and the insertion site of the protein of interest) into the BD expression plasmid pAS2.1 BD using NcoI and PstI restriction sites (pAS2.1 BD FLAG). The coding sequences of TDG and TET1 domains were cloned into the BD (pAS2.1 BD FLAG) and the AD (pACT2 AD) of the Gal4 protein. Strains AH109 (*MAT $\alpha$* , *trp1-901*, *leu2-3, 112*, *ura3-52*, *his3-200*, *gal4 $\Delta$* , *gal80 $\Delta$* , *LYS2::GAL1<sub>UAS</sub>-GAL1<sub>TATA</sub>-HIS3*, *MEL1*, *GAL2<sub>UAS</sub>-GAL2<sub>TATA</sub>-ADE2*, *URA3::MEL1<sub>UAS</sub>-MEL1<sub>TATA</sub>-lacZ*) and Y187 (*MAT $\alpha$* , *ura3-52*, *his3-200*, *ade2-101*, *trp1-901*, *leu2-3, 112*, *gal4 $\Delta$* , *met<sup>-</sup>*, *gal80 $\Delta$* , *URA3::GAL1<sub>UAS</sub>-GAL1<sub>TATA</sub>-lacZ*) were co-transformed with 50-500 ng of bait and prey plasmids according to the Clontech manual. For AH109, interactions were assessed by spotting serial dilutions of cells on selective medium (SC-LEU-TRP-ADE-HIS) supplemented with 2.5 mM 3AT (3-Amino-1,2,4-triazole), a competitive inhibitor of the *HIS3* gene product. Cells were incubated for 6 to 7 days at 30°C.  $\beta$ -Galactosidase activity was assayed using the Y187 strain performing colony lift assays as described in the Clontech manual. Briefly,  $10^6$  cells were dropped on SC medium selecting for the plasmids (SC-LEU-TRP) and grown for 24 h at 30°C. Cells were transferred to filter paper (Filtrak, 80 g/m<sup>2</sup>) before snap-freezing in liquid nitrogen and subsequent thawing for cells lysis. The filter with the lysed cells was soaked with 2 ml of Z buffer (100 mM Na phosphate buffer pH 7.0, 10 mM KCl, 1 mM MgSO<sub>4</sub>, 33  $\mu$ M  $\beta$ -mercaptoethanol, 817  $\mu$ M X-Gal) and incubated at 30°C for up to 17 h.

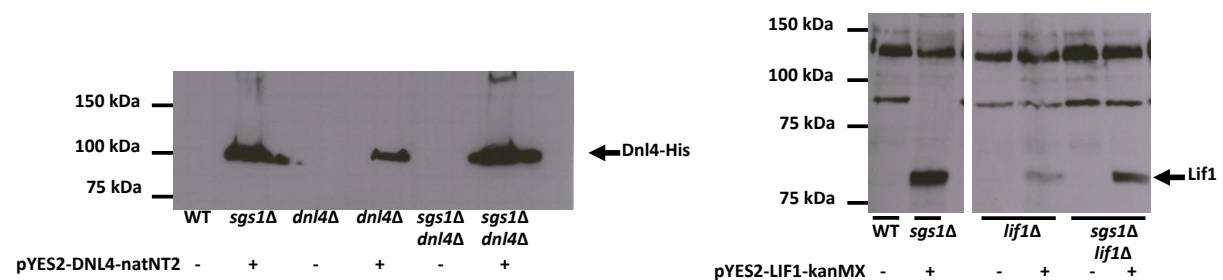
### 4.3.3 DNA Ligase 4 Function in the rDNA of Budding Yeast

(unpublished data)

Non-homologous end-joining (NHEJ) is a key DNA repair pathway re-joining DSB ends. As DSBs and DNA recombination events were detected in the ribosomal DNA (rDNA) of budding yeast (Burkhalter & Sogo, 2004; Kobayashi et al., 2004; Weitao et al., 2003), our group previously tested the influence of the DNA ligase 4 (Dnl4), an essential NHEJ factor, on the integrity of the rDNA (Fritsch et al., 2010).

Surprisingly, in the absence of the RecQ helicase Sgs1, a condition of increased DNA instability, Dnl4 appeared to contribute to DSB formation as inferred from the decreased level of breaks in the *sgs1Δdnl4Δ* double mutant when compared to either single mutant, with *dnl4Δ* cells showing wild-type levels. Moreover, the formation of extrachromosomal ribosomal DNA circles (ERCs) – circular DNA molecules consisting of single or multiple rDNA repeats that are excised from the rDNA in a homologous recombination (HR)-dependent process (see chapter 2.2.3 for detailed information) – was increased upon *DNL4* deletion in the *sgs1Δ*, but not in the wild-type background. This contradicted the idea that increased DSB levels at the ribosomal replication fork barrier (rRFB) lead to elevated ribosomal recombination and ERC formation. Moreover, deletion of the Dnl4 cofactor *LIF1* exhibited increased ERC levels in the wild-type, but not in the *sgs1Δ* background, opposite to the phenotype of the Dnl4 deficiency. These conflicting data suggested that Dnl4 might perform unknown functions at the rRFB, not involving the NHEJ pathway.

To address this exciting possibility, I first wanted to test whether re-expression of Dnl4 and/or Lif1 could restore the observed wild-type phenotypes. To that end I performed complementation analyses of ERC formation. Using existing *DNL4* and *LIF1* expression plasmids, I exchanged the marker cassette to make it suitable for our strains. I transformed WT, *sgs1Δ*, *dnl4Δ* and *sgs1Δ dnl4Δ* strains with the *DNL4* expression plasmid and WT, *sgs1Δ*, *lif1Δ* and *sgs1Δ lif1Δ* strains with the *LIF1* expression vector. To prevent massive overexpression, I grew cells in YPD, without inducing expression from the *nmt1* promoter and I detected Dnl4 and Lif1 by Western blot analysis (Figure 10).

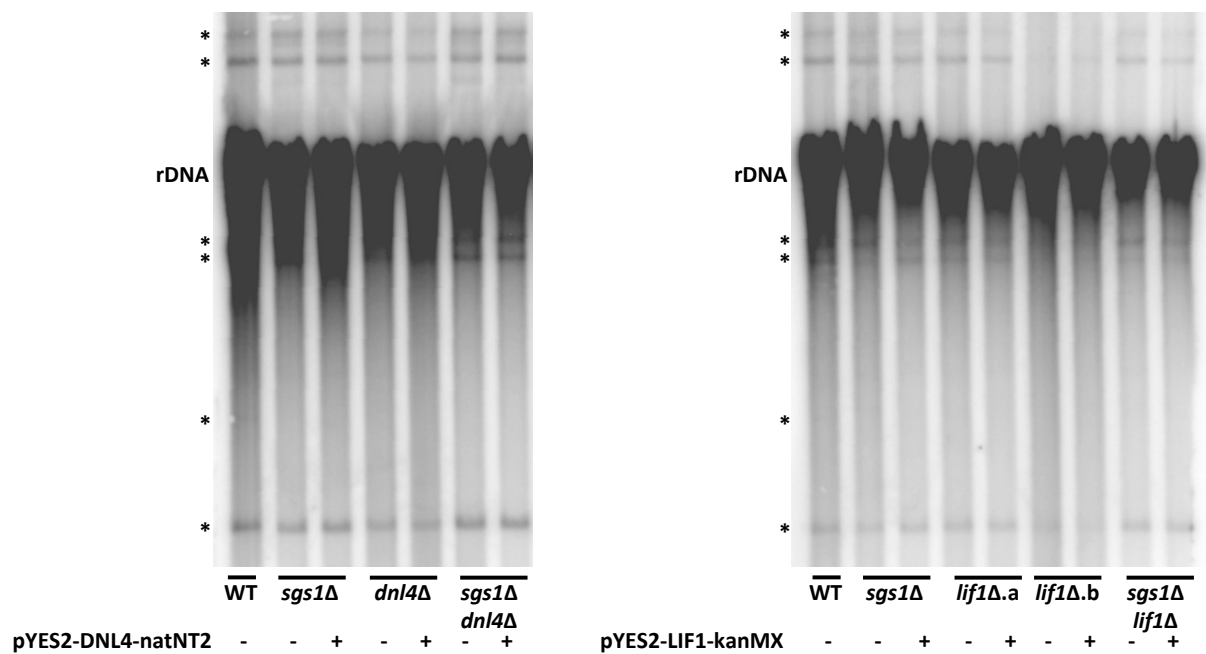


**Figure 10. Expression of Dnl4 and Lif1.** Western blot analysis of protein extracts of strains transformed with pYES2-DNL4-natNT2 or pYES2-LIF1-kanMX that were prepared from YPD cultures (basal protein expression under non-inducing conditions). Dnl4 was detected using an anti-His antibody (Covance), Lif1 by a polyclonal rabbit anti-Lif1 antibody. The bands of Dnl4 and Lif1 expression are indicated.

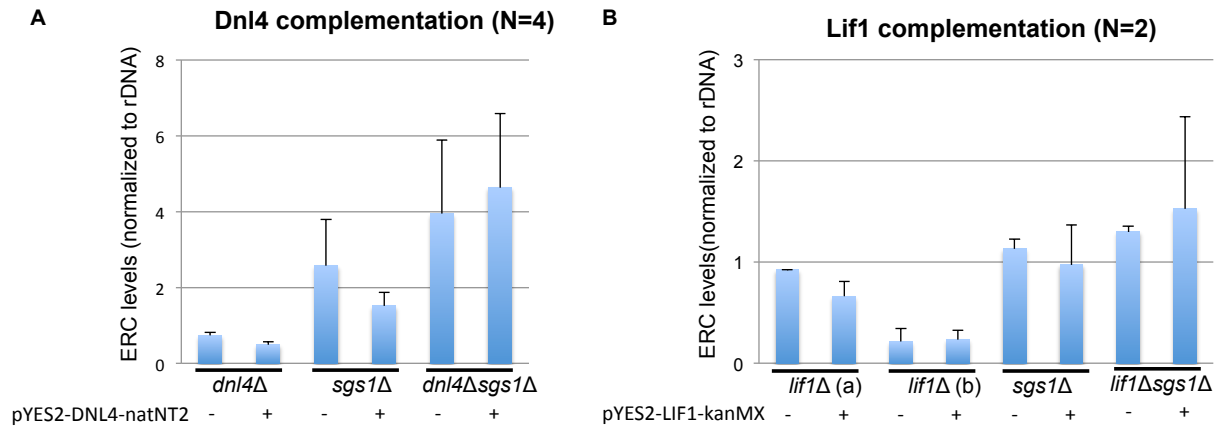
I isolated DNA from cells transformed with either of the expression vectors and subjected it to gel electrophoresis and subsequent Southern blotting (Figure 11). All detectable ERC species were quantified and normalized to the bulk of rDNA (Figure 12). For the Dnl4-expressing strains, ERC levels of four experiments were normalized to those of wild-type cells (Figure 12A). As shown previously (Fritsch et al., 2010), ERC levels of *dnl4Δ* cells are similar to wild-type levels. In contrast, ERC levels

were increased in *sgs1Δ* and even further in the *sgs1Δ dnl4Δ* double mutant cells. Although the tendencies were comparable to that of the published ERC levels, the total increase was lower in my experiments. Unfortunately, I was not able to obtain a clear complementation by re-introducing Dnl4. While in *sgs1Δ* cells ERC occurrence was reduced upon introduction of Dnl4, it was unchanged in the complemented *sgs1Δ dnl4Δ* double mutant. This might be explained by the non-endogenous expression of Dnl4. Although the cells were grown under non-inducing conditions with only residual expression levels, these levels are likely to be different compared to those in wild-type cells. Given the inability to complement the wild-type ERC levels by re-expression of Dnl4, protein expression levels might be crucial for the Dnl4 function in the rDNA.

As for ERC analysis in Lif1-complemented cells, I was not able to reproduce the increase levels previously observed in *lif1Δ*, *sgs1Δ* and *lif1Δ sgs1Δ* cells. ERC levels in two different *lif1Δ* strains were similar to or lower than those in wild-type in two experimental replicas (Figure 12B). Moreover, re-expression of Lif1 did not alter the ERC levels in any strain.



**Figure 11. Southern Blots for ERC detection.** DNA of cells expressing *DNL4* or *LIF1* was extracted and separated on 0.8% agarose gels. An rDNA-specific DNA probe was used to detect the ERCs and the rDNA. Asterisks indicate the ERC species included in the quantification.



**Figure 12. ERC quantification.** Quantitation of ERC signals from Figure 11. **(A-B)** The total signal from ERC species was normalized to the bulk of rDNA and to wild-type ERC levels. Cells were transformed with the *DNL4* (A) or *LIF1* (B) expression plasmids as indicated. SD is shown for four (*Dnl4*) and two (*Lif1*) repeats.

Taken together the lack of reproducibility in the case of *Lif1* complementation and the inability to complement the *Dnl4* phenotype, I concluded that measurement of ERCs is not suitable for analyzing *Dnl4* and *Lif1* functions at the rDNA, at least with the experimental setup used for these experiments. In addition, chromatin immunoprecipitation (ChIP) experiments with chromatin from strains expressing *DNL4-9myc* or *LIF1-13myc*, did not show a robust enrichment of *Dnl4* nor *Lif1* at the rRFB (data not shown). However, due to lack of a genomic region that could serve as a positive control for the functionality of these ChIP experiments, it remained elusive whether the applied conditions were unsuitable for *Dnl4* and *Lif1* or whether there was no enrichment at the rRFB. Because of these uncertainties, we decided not to further pursue this project.

## Material and Methods

### Cloning of DNL4 and LIF1 expression vectors

For a marker exchange, the kanMX cassette of the pYM45 vector was PCR-amplified using primers carrying overhangs with XbaI restriction sites. The obtained fragment was digested with XbaI and cloned in NheI digested *DNL4* and *LIF1* expression plasmids pYes2-DNL4-URA3 and pYes2-LIF1-URA3 (Schär laboratory), respectively. The resulting plasmids pYES2-DNL4-kanMX and pYES2-LIF1-kanMX were confirmed by restriction digests. For a second marker switch, a BglII/SacI fragment containing the natNT2 cassette was cut out from pYM17 and inserted into eponymous sites of pYES2-DNL4-kanMX, resulting in the pYES-DNL4-natNT2 vector.

### Yeast transformation

50 ml of YPD were inoculated with 500  $\mu$ l of a 5 ml overnight culture and incubated for four to five hours (30°C, 230 rpm). After centrifugation (1930 g, 3 min), cells were washed in 20 ml of water,

spun down, resuspended in 400  $\mu$ l of TEL (100 mM lithium acetate, 10 mM Tris-HCl pH7.5, 1 mM EDTA) and chilled on ice. 100 ng DNA was mixed with 40  $\mu$ g single-stranded DNA, 100  $\mu$ l of the cell suspension and 700  $\mu$ l Li-PEG (100 mM lithium acetate, 10 mM Tris-HCl pH7.5, 1 mM EDTA, 40% PEG 4000). The suspension was incubated at 30°C for 20 minutes, followed by a 30-minutes incubation at 42°C. Cells were washed twice with water, resuspended in 100  $\mu$ l of water and plated for selection.

#### Western Blot analysis

For the preparation of cell extracts, cells were grown in 30 ml YPD and harvested at a concentration of  $5 \times 10^6$  cells/ml. Cells were incubated in 3 ml of 0.1 M EDTA-KOH pH8.0 and 10 mM DTT for 10 min at 30°C before spheroplasts were produced in 3 ml of 20 mM potassium phosphate pH7.1, 1.1 M sorbitol and 0.33 mg/ml zymolyase at 30°C. Cells were washed twice in 1 ml of ice-cold wash buffer (5 mM Tris-HCl pH7.4, 20 mM EDTA-KOH pH8.0, 1 M sorbitol, 1% aprotinin A, 0.5 mM PMSF). Spheroplasts were then resuspended in 150  $\mu$ l elution buffer (50 mM HEPES-KOH pH7.5, 100 mM KCl, 2.5 MgCl<sub>2</sub>, 1% aprotinin A, 1 mM PMSF, 2  $\mu$ g/ml antipain, 300  $\mu$ g/ml benzamidin, 0.5  $\mu$ g/ml leupeptin, 1  $\mu$ g/ml pepstatin A, 20  $\mu$ g/ml TPCK, 10  $\mu$ g/ml TLCK) before they were lysed by the addition of 150  $\mu$ l of elution buffer containing 0.5% Triton X-100. After centrifugation (11000 rpm, 10 min), the pellet was resuspended in 30  $\mu$ l of elution buffer containing 0.25% Triton X-100.

The extracts were boiled for 5 min in 1x SDS loading buffer (2% SDS, 62.5 mM Tris-HCl pH6.8, 4 mM EDTA, 10% glycerol, 2.5%  $\beta$ -mercaptoethanol) and 20  $\mu$ l were loaded on an SDS-PAGE gel for separation. Proteins were transferred to nitrocellulose membranes (Whatman) overnight. For detection of Dnl4, the membrane was blocked three times in TEN (20 mM Tris-HCl pH 8.0, 1 mM EDTA, 140 mM NaCl) containing 5% milk (TEN5M) for 15 min. Incubation with the primary antibody (anti-HIS, Covance, 1:1000) in TEN5M for 2 h was followed by two washes in TEN5M for 15 min and 2 h of incubation with the secondary antibody (anti-mouse-HRP, GE Healthcare, 1:5000) in TEN5M. Finally, the membrane was washed five times in TEN for 5 min, developed and exposed to X-ray films. Detection of Lif1 was performed in the same way, except for the usage of TEN containing 7.5% milk and 0.2% TWEEN-20. The primary antibody (anti-Lif1, from rabbit) was diluted 1:1000 and incubated at 4°C while the secondary antibody (anti-rabbit-HRP, GE Healthcare) was diluted 1:5000.

#### ERC analysis

Cells were grown in YPD to late logarithmic phase and genomic DNA was isolated using Qiagen genomic tips. 2  $\mu$ g DNA were loaded on a 0.8% agarose gel and run in 1x TAE (65 V, 17 h, RT) before being stained with ethidium bromide. Southern blot analysis was performed as published (Fritsch et al., 2010). Signal quantitation was performed using the ImageQuant TL 1D program (GE Healthcare)

and manual background subtraction was applied. All ERC fractions (asterisks in Figure 11) were pooled and normalized to the bulk of rDNA and to the wild-type level.

## 5 Concluding Discussion and Outlook

In the regular relaxation of torsional stress in DNA, Top1 transiently cleaves one strand of the DNA, forming a covalent complex to allow for DNA relaxation, before it reseals the nick. We showed that at the rRFB, in contrast to the normally transient nature of the covalent complex, Top1 covalent cleavage complexes (Top1ccs) are highly enriched, suggesting they are stabilized in the DNA-bound conformation by preventing or delaying religation of the nick. This stabilization might be achieved by protein-protein interactions between Top1, Fob1 and Tof2, as we showed the latter two to be essential for Top1cc accumulation and interactions between the three proteins were reported before (Huang et al., 2006; Park & Sternglanz, 1999). Such interactions could indeed induce a conformational change of Top1 to temporarily reduce its ligation capacity. A proof of principle may be given in a recent report about the regulation of the catalytic activity of DNA gyrase, the type IIA topoisomerase of *E. coli*. Co-crystallization of DNA gyrase with the inhibitory protein YacG revealed that interaction between both proteins induces remodeling of the DNA gyrase binding pocket, thereby preventing DNA binding and relaxation (Vos et al., 2014). Since we provided evidence that Top1cc levels at the ribosomal replication fork barrier (rRFB) are not affected by compromising the repair of pathological Top1ccs, we consider rRFB Top1ccs as intermediates of a physiological process. It is tempting to propose that Fob1, as a main regulator of ribosomal DNA (rDNA) structure, directly recruits a ribosomal complex containing at least Fob1, Tof2 and Top1. In line with this idea, we also detected Fob1-dependent Top1cc enrichment at ectopically located rRFBs that were neither recruited to the nucleolus nor to the inner nuclear membrane. Interestingly, we detected a partial reduction of Fob1 enrichment at an ectopically located rRFB in the absence of Top1 or Tof2, suggesting that Top1 and Tof2 stabilize a ribosomal Fob1 complex. The hypothesis of Fob1 recruiting a complex to the rRFB could be further tested by artificially targeting Fob1 to genomic locations to test its capacity of recruiting Top1 or Tof2 or even to stabilize Top1 nicks at such locations. We attempted to recruit Fob1 fused to the Gal4 binding domain to the *GAL4<sub>UAS</sub>* elements of the *GAL2* promoter, but failed to detect the Fob1 fusion protein at this site (data not shown). Experiments using this or another targeting setup should be pursued to successfully target Fob1 to an rDNA-unrelated site. In addition, the effect of *FOB1* or *TOP1* deletion on Tof2 association with the rRFB should be addressed using CHIP experiments. Furthermore, biochemical reconstitution assays would be valuable to further dissect the prerequisites for and the mechanistic principles of Top1cc



formation and stabilization. Using recombinant purified proteins, it could be tested whether Fob1 is sufficient to target Top1 and to induce site-specific DNA nicks and how Top2 further stimulates such activity.

A misalignment of the second DNA end making it unavailable to Top1 for religation may also contribute to or even underlie Top1cc stabilization at the rRFB. Binding of Fob1, Top1 and Top2 to the rRFB site might indeed alter the DNA structure in a way that prevents nick resealing. The proposed DNA wrapping around Fob1 dimers (Kobayashi, 2003) and the resulting DNA bend could serve as an additional explanation for DNA misalignment, in addition to potentially providing site-specificity for the previously observed Top1-dependent nicks at this site. Bended DNA was shown to be substrate for the calf thymus topoisomerase 1 (Caserta et al., 1989) and wrapping of the DNA might additionally expose preferential Top1 nicking sites (Edwards et al., 1982; Shuman & Prescott, 1990), which could provide the frame for site-specific DNA cleavage by Top1.

We hypothesized that the Top1 nicks at the rRFB evolved to remove DNA torsion resulting from the constant and high transcription of the *35S* rRNA gene. This could prove particularly valuable in the context of the anchoring of the rDNA repeats to the inner nuclear membrane that constrains diffusion of DNA torsion (Chan et al., 2011; Huang et al., 2006; Mekhail et al., 2008). Interestingly, we found that Top1cc accumulation is independent from RF stalling and transcription as inferred from Top1cc enrichment at non-blocking eRFBs and the low transcriptional activity in the eRFB region, respectively. Consistently, presence of the previously observed Top1-dependent nicks was not altered by changing RNA polymerase I activity (Vogelauer & Camilloni, 1999). Despite being now independent of replication and transcription, a constitutive Top1cc-dependent system for relieving torsional stress could have evolved in response to these processes in the rDNA to provide a permanent possibility of normalizing DNA torsion.

In our study, we could also clarify the nature of the DSBs previously observed at the rRFB (Burkhalter & Sogo, 2004; Kobayashi et al., 2004; Weitao et al., 2003). Top1 is required for the majority of the nicks observed in wild-type cells, implying that the apparent DSBs are rather Top1 nicks at the rRFB. These nicks are close to the rRFB, and thus, the stretch of DNA between them and the RF could melt during DNA isolation, giving rise to the molecules appearing as DSBs on Southern blot. In contrast, we observed additional DSBs in the absence of the RF-stabilizing RecQ helicase Sgs1 that were independent of Top1. Thus, these breaks are probably real DSBs that also contribute to rDNA recombination as suggested from the rDNA instability in Sgs1-deficient cells.

Additionally, we observed an unstable rDNA and increased production of recombination products in the absence of Top1. Though Top1 might well provide rDNA stability due to its accumulation at the rRFB, we were not able to distinguish between the rRFB-related function and its canonical functions during replication and transcription.

Stabilization of Top1cc might not be restricted to the rDNA context of budding yeast. Scenarios similar to the situation at the rRFB are conceivable in other complex genomic regions with an increased need for the release of torsional stress. In higher eukaryotes, such a mechanism could contribute to the regulation of DNA torsion in their highly complex and repetitive genome. The high order organization of the chromatin might create additional need for DNA torsion removal and, thus, the stabilization of a topoisomerase in its cleavage complex could serve as a paradigm of topological regulation also in higher organisms. Persistent removal of torsional stress could involve topoisomerases 1 or other topoisomerases and additional proteins that take over the role of the budding yeast nuclear proteins Fob1 and Tof2 in the stabilization of the cleavage complex.

In the second project of my thesis, we showed that despite Ung1 provides the major uracil excision activity *in vitro*, both UDGs of *S. pombe*, Ung1 and Thp1, contribute to mutation avoidance. Indeed, increased mutation rates became only apparent in the absence of both proteins. This is similar to experiments in a non-peer-reviewed report stating increased mutation rates in the double mutant *thp1Δ ung1Δ* (Ikeda et al., 2009). Similarly, we saw a significant accumulation of uracil in genomic DNA only in the absence of Ung1 and Thp1. Despite these apparent compensatory functions of the two fission yeast UDGs, we present evidence for adverse effects of base excision by Thp1. We showed, that Thp1 expression is toxic in cells exposed to the anti-cancer drug 5-FU or facing increased cytosine deamination by human AID. Similar to human TDG, Thp1 is strongly product-inhibited (Hardeland et al., 2003) and the longer lifetime of Thp1-generated AP sites could explain the observed suboptimal performance as a DNA repair protein. However, adverse Thp1-mediated repair was not only observed under 5-FU generated stress, but also under unperturbed growth conditions. We found Thp1-dependent base excision, most likely uracil excision, to account for a large part (>60%) of spontaneous mitotic recombination. This strongly indicates that Thp1-generated long-lived AP sites induce spontaneous genomic instability in wild-type cells. Despite all this, however, Thp1 might still be required for uracil removal under specific circumstances as the co-evolution of two UDG suggests important functions for either protein. For instance, Thp1 might be important in the removal of specific lesions that are refractory to Ung1 as suggested from its broad substrate spectrum (Hardeland et al., 2003). Alternatively, Thp1 and Ung1 may act during different cell cycle stages. Since quick repair is important in the context of replicating DNA, Ung1 might be the main enzyme for uracil removal in the S-phase of the cell cycle, as it was suggested for mammalian UNG2 (Hagen et al., 2008) (Hu et al., 2008; Otterlei et al., 1999). In contrast, the slow turnover of Thp1 might be less of a problem, or perhaps even useful, outside of S-phase and in non-dividing cells. In non-replicating DNA, where repair proteins might not be readily available, the association of Thp1 with the AP site might protect the site from breakage and at the same time mark it for the

recruitment of the appropriate repair machine. In line with this, artificial expression of mammalian TDG during S-phase interferes with cell proliferation (Hardeland et al., 2007) and *E. coli* Mug is important for mutation avoidance in stationary phase cells (Mokkapati et al., 2001). It will thus be of high interest to test whether the cell cycle-dependent regulation observed in mammalian cells, also applies for the *S. pombe* ortholog. To that end, protein expression and/or localization will have to be determined in the course of the cell cycle.

Finally, given the essential epigenetic function of mammalian TDG, we considered and tested the possibility of a similar specialization of the *S. pombe* Thp1. Besides the presence of different conserved epigenetic pathways, fission yeast only expresses two UDGs, whereas five are present in mammals, therefore presenting a simplified model to separate UDG functions. We thus examined the possibility to use *S. pombe* as a model to investigate epigenetic functions of Thp1-dependent BER in the absence of DNA methylation. Comparing RNA transcription of wild-type and Thp1-deficient cells using tiling arrays, we did not find pronounced dysregulated expression patterns in Thp1-deficient cells. Despite this lack of changes, we noted a well-spread suppression of transcriptional activity in Thp1-deficient cells. Furthermore, Thp1-deficiency led to a higher variability of gene expression across replicas as compared to that across wild-type replicas, a typical feature of epigenetic instability.

Although the observed differences in the expression of single genes were not statistically significant, the transcriptome-wide trend towards lower and more variable gene expression indicates that Thp1 contributes to stable gene expression. However, this assumption requires further testing. In our experiments, RNA for each replica was pooled from three cell populations derived from independent spores. This offers the possibility to increase the dynamic range of expression variability by the use of non-pooled spore-derived cell populations for transcriptome analysis. Identification of involved mechanisms would be also of great interest. We checked Thp1-association with the genome by ChIP paired with next-generation sequencing (ChIP-seq), but could not identify specific sites or regions showing significant Thp1-enrichment. This either reflects lack of specific Thp1-association with the genome or failure of the ChIP procedure, as there were no established Thp1-binding sites in the genome that we could use as a positive control to optimize the ChIP protocol. Altogether, however, our findings suggest an epigenetic function in the maintenance of active transcription for Thp1.

As a next step, the potential gene regulatory function of Thp1 could be examined by analyzing differences in histone marks in cells expressing or not Thp1, as differences in gene expression are normally accompanied by altered modification patterns of the histone tails (Kouzarides, 2007). Such analysis could be performed either genome-wide or targeted. In the latter case, those genes from the tiling array exhibiting the highest variation and/or downregulation would be the most promising candidates to investigate. In case of any detectable differences in histone modification between wild-

type and Thp1-deficient cells, it would be interesting to know whether Ung1 or downstream BER enzymes contribute to this Thp1 function, a possibility that could be tested by comparing gene expression or histone modifications in the presence or absence of different BER enzymes.

Another potential mechanism for activating transcription is the active nucleosome eviction at promoters, facilitating transcription factor binding and transcription initiation. Nucleosome depletion in promoter and/or gene bodies has been observed in budding and fission yeast upon induced gene expression (Lee et al., 2004; Shivaswamy et al., 2008; Zawadzki et al., 2009). It is a tempting, but speculative, idea that Thp1 by performing slow repair could increase the occurrence of single-stranded nicks resulting from AP site hydrolysis. Assuming Thp1 is targeted to gene promoters through transcription factor interactions, this could, similar to the reported action of topoisomerases (Durand-Dubief et al., 2010; Gavin et al., 2001; Pedersen et al., 2012), lead to decreased torsion, offering the possibility of nucleosome eviction by chromatin remodelers.

During my studies, I have identified a novel concept of Top1 activity that seems to be important for rDNA stability. As similar concepts might apply to other genomic contexts, our work provides a basis to further investigate the mechanisms underlying the regulation of DNA torsion in higher eukaryotes. Moreover, our work on the separation of UDG functions unraveled an interesting difference between Ung1 and Thp1 in their qualitative repair outcome. The efficient Ung1-initiated repair and the non-productive and toxic repair by Thp1 suggest a spatial and temporal separation of the two uracil excision activities, such as potential functions in different cell cycle stages or in the immune system. It will be of great interest to identify such UDG-specific processes. The presence of a seemingly conserved Thp1-mediated epigenetic effect on gene regulation offers the possibility of analyzing this TDG/Thp1 function in a model organism lacking DNA methylation.

## 6 References

- Abdel-Rahman, W. M., Knuutila, S., Peltomäki, P., Harrison, D. J., & Bader, S. A. (2008). Truncation of MBD4 predisposes to reciprocal chromosomal translocations and alters the response to therapeutic agents in colon cancer cells. *DNA Repair*, 7(2), 321–328.
- Ahn, J., Osman, F., & Whitby, M. (2005). Replication fork blockage by RTS1 at an ectopic site promotes recombination in fission yeast. *EMBO J*, 24, 2011–2023.
- Almeida, K. H., & Sobol, R. W. (2007). A unified view of base excision repair: Lesion-dependent protein complexes regulated by post-translational modification. *DNA Repair*, 6(6), 695–711.
- An, Q., Robins, P., Lindahl, T., & Barnes, D. E. (2005). C → T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases. *EMBO J*, 24(12), 2205–2213.
- Antequera, F., Tamame, M., Villanueva, J. R., & Santos, T. (1984). DNA methylation in the fungi. *J Biol Chem*, 259(13), 8033–8036.
- Aravind, L., & Koonin, E. V. (2000). The alpha/beta fold uracil DNA glycosylases: a common origin with diverse fates. *Genome Biology*, 1(4), RESEARCH0007.
- Arcangioli, B., & Klar, A. J. (1991). A novel switch-activating site (SAS1) and its cognate binding factor (SAP1) required for efficient mat1 switching in *Schizosaccharomyces pombe*. *EMBO J*, 10(10), 3025–3032.
- Baba, D., Maita, N., Jee, J.-G., Uchimura, Y., Saitoh, H., Sugasawa, K., et al. (2005). Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature*, 435(7044), 979–982.
- Baba, D., Maita, N., Jee, J.-G., Uchimura, Y., Saitoh, H., Sugasawa, K., et al. (2006). Crystal structure of SUMO-3-modified thymine-DNA glycosylase. *J Mol Biol*, 359(1), 137–147.
- Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, 21(3), 381–395.
- Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., & Pearl, L. H. (1998). Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell*, 92(1), 117–129.
- Barrett, T. E., Schärer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L., & Pearl, L. H. (1999). Crystal structure of a thwarted mismatch glycosylase DNA repair complex. *EMBO J*, 18(23), 6599–6609.
- Benbrook, D. M., Chambon, P., Rochette-Egly, C., & Asson-Batres, M. A. (2014). History of Retinoic Acid Receptors. In *Subcellular Biochemistry* (Vol. 70, pp. 1–20). Dordrecht: Springer Netherlands.
- Benguría, A., Hernández, P., Krimer, D. B., & Schwartzman, J. B. (2003). Sir2p suppresses recombination of replication forks stalled at the replication fork barrier of ribosomal DNA in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, 31(3), 893–898.
- Bermejo, R., Doksan, Y., Capra, T., Katou, Y.-M., Tanaka, H., Shirahige, K., & Foiani, M. (2007). Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation. *Gene Dev*, 21(15), 1921–1936.
- Bermejo, R., Lai, M. S., & Foiani, M. (2012). Preventing Replication Stress to Maintain Genome Stability: Resolving Conflicts between Replication and Transcription. *Mol Cell*, 45(6), 710–718.
- Bétermier, M., Bertrand, P., & Lopez, B. S. (2014). Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? *Plos Genetics*, 10(1), e1004086.
- Bhalla, N., Biggins, S., & Murray, A. W. (2002). Mutation of YCS4, a budding yeast condensin subunit, affects mitotic and nonmitotic chromosome behavior. *Mol Biol Cell*, 13(2), 632–645.
- Bhutani, N., Brady, J. J., Damian, M., Sacco, A., Corbel, S. Y., & Blau, H. M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature*, 463(7284), 1042–1047.
- Biswas, S., & Bastia, D. (2008). Mechanistic Insights into Replication Termination as Revealed by Investigations of the Reb1-Ter3 Complex of *Schizosaccharomyces pombe*. *Mol Cell Biol*, 28(22), 6844–6857.

- Boiteux, S. & Jinks-Robertson, S. (2013). DNA Repair Mechanisms and the Bypass of DNA Damage in *Saccharomyces cerevisiae*. *Genetics*, 193(4), 1025–1064.
- Boland, M. J., & Christman, J. K. (2008). Characterization of Dnmt3b : thymine-DNA glycosylase interaction and stimulation of thymine glycosylase-mediated methyltransferase(s) repair by DNA and RNA. *J Mol Biol*, 379(3), 492–504.
- Borys-Brzywczy, E., Arczewska, K. D., Saparbaev, M., Hardeland, U., Schär, P., & Kuśmierk, J. T. (2005). Mismatch dependent uracil/thymine-DNA glycosylases excise exocyclic hydroxyethano and hydroxypropano cytosine adducts. *Acta Biochimica Polonica*, 52(1), 149–165.
- Branzei, D., & Foiani, M. (2010). Maintaining genome stability at the replication fork. *Nature Reviews Mol Cell Biology*, 11(3), 208.
- Brewer, B. J., & Fangman, W. L. (1988). A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell*, 55(4), 637–643.
- Brewer, B. J., Lockshon, D., & Fangman, W. L. (1992). The arrest of replication forks in the rDNA of yeast occurs independently of transcription. *Cell*, 71(2), 267–276.
- Brégeon, D., Doddridge, Z. A., You, H. J., Weiss, B., & Doetsch, P. W. (2003). Transcriptional Mutagenesis Induced by Uracil and 8-Oxoguanine in *Escherichia coli*. *Mol Cell*, 12(4), 959–970.
- Brooks, S. C., Adhikary, S., Rubinson, E. H., & Eichman, B. F. (2013). Recent advances in the structural mechanisms of DNA glycosylases. *Biochimica Et Biophysica Acta (BBA) - Proteins and Proteomics*, 1834(1), 247–271.
- Burkhalter, M. D., & Sogo, J. M. (2004). rDNA enhancer affects replication initiation and mitotic recombination: Fob1 mediates nucleolytic processing independently of replication. *Mol Cell*, 15(3), 409–421.
- Bühler, M. (2009). RNA turnover and chromatin-dependent gene silencing. *Chromosoma*, 118(2), 141–151.
- Caldecott, K. W. (2008). Single-strand break repair and genetic disease. *Nat Rev Genet*, 9(8), 619–631.
- Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A., & Labib, K. (2005). Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. *Gene Dev*, 19(16), 1905–1919.
- Capuano, F., Mülleder, M., Kok, R., Blom, H. J., & Ralser, M. (2014). Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. *Analytical Chemistry*, 86(8), 3697–3702.
- Caserta, M., Amadei, A., Di Mauro, E., & Camilloni, G. (1989). In vitro preferential topoisomerization of bent DNA. *Nucleic Acids Res*, 17(21), 8463–8474.
- Chagin, V. O., Stear, J. H., & Cardoso, M. C. (2010). Organization of DNA replication. *Cold Spring Harb Perspect Biol*, 2(4), a000737.
- Champoux, J. J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem*, 70, 369–413.
- Chan, J. N. Y., Poon, B. P. K., Salvi, J., Olsen, J. B., Emili, A., & Mekhail, K. (2011). Perinuclear cohibin complexes maintain replicative life span via roles at distinct silent chromatin domains. *Dev Cell*, 20(6), 867–879.
- Chen, D., Lucey, M. J., Phoenix, F., Lopez-Garcia, J., Hart, S. M., Losson, R., et al. (2003). T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha. *J Biol Chem*, 278(40), 38586–38592.
- Chen, S. H., Chan, N.-L., & Hsieh, T.-S. (2013). New mechanistic and functional insights into DNA topoisomerases. *Annu Rev Biochem*, 82, 139–170.
- Chen, T., & Dent, S. Y. R. (2013). Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet*, 15(2), 93–106.
- Chen, Z.-X., & Riggs, A. D. (2011). DNA methylation and demethylation in mammals. *J Biol Chem*, 286(21), 18347–18353.
- Chevray, P. M., & Nathans, D. (1992). Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun. *Proc Natl Acad Sci USA*, 89(13),

5789–5793.

- Christman, M. F., Dietrich, F. S., & Fink, G. R. (1988). Mitotic recombination in the rDNA of *S. cerevisiae* is suppressed by the combined action of DNA topoisomerases I and II. *Cell*, *55*(3), 413–425.
- Cobb, J. A., Shimada, K., & Gasser, S. M. (2004). Redundancy, insult-specific sensors and thresholds: unlocking the S-phase checkpoint response. *Curr Opin Genet Dev*, *14*(3), 292–300.
- Cortázar, D., Kunz, C., Saito, Y., Steinacher, R., & Schar, P. (2007). The enigmatic thymine DNA glycosylase. *DNA Repair*, *6*(4), 489–504.
- Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., Macdougall, E., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*, *470*(7334), 419–423.
- Cortellino, S., Turner, D., Masciullo, V., Schepis, F., Albino, D., Daniel, R., et al. (2003). The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. *Proc Natl Acad Sci USA*, *100*(25), 15071–15076.
- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, *146*(1), 67–79.
- Cotta-Ramusino, C., Fachinetti, D., Lucca, C., Dokhani, Y., Lopes, M., Sogo, J., & Foiani, M. (2005). Exo1 processes stalled replication forks and counteracts fork reversal in checkpoint-defective cells. *Mol Cell*, *17*(1), 153–159.
- Dalgaard, J. Z., & Klar, A. J. (2000). *swi1* and *swi3* perform imprinting, pausing, and termination of DNA replication in *S. pombe*. *Cell*, *102*(6), 745–751.
- Dammann, R., Lucchini, R., Koller, T., & Sogo, J. M. (1993). Chromatin structures and transcription of rDNA in yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res*, *21*(10), 2331–2338.
- de Lahondès, R., Ribes, V., & Arcangioli, B. (2003). Fission Yeast Sap1 Protein Is Essential for Chromosome Stability. *Eukaryot Cell*, *2*(5), 910–921.
- Defossez, P. A., Prusty, R., Kaeberlein, M., Lin, S. J., Ferrigno, P., Silver, P. A., et al. (1999). Elimination of replication block protein Fob1 extends the life span of yeast mother cells. *Mol Cell*, *3*(4), 447–455.
- Deng, C., Brown, J. A., You, D., & Brown, J. M. (2005). Multiple endonucleases function to repair covalent topoisomerase I complexes in *Saccharomyces cerevisiae*. *Genetics*, *170*(2), 591–600.
- Dengg, M., Garcia-Muse, T., Gill, S. G., Ashcroft, N., Boulton, S. J., & Nilsen, H. (2006). Abrogation of the CLK-2 checkpoint leads to tolerance to base-excision repair intermediates. *Embo Reports*, *7*(10), 1046–1051.
- Di Felice, F., Cioci, F., & Camilloni, G. (2005). FOB1 affects DNA topoisomerase I in vivo cleavages in the enhancer region of the *Saccharomyces cerevisiae* ribosomal DNA locus. *Nucleic Acids Res*, *33*(19), 6327–6337.
- Di Noia, J. M., Rada, C., & Neuberger, M. S. (2006). SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *EMBO J*, *25*(3), 585–595.
- Dianov, G. L., & Hübscher, U. (2013). Mammalian base excision repair: the forgotten archangel. *Nucleic Acids Res*, *41*(6), 3483–3490.
- Dianov, G. L., Thybo, T., Dianova, I. I., Lipinski, L. J., & Bohr, V. A. (2000). Single nucleotide patch base excision repair is the major pathway for removal of thymine glycol from DNA in human cell extracts. *J Biol Chem*, *275*(16), 11809–11813.
- Dinh, T. T., Gao, L., Liu, X., Li, D., Li, S., Zhao, Y., et al. (2014). DNA Topoisomerase 1 $\alpha$  Promotes Transcriptional Silencing of Transposable Elements through DNA Methylation and Histone Lysine 9 Dimethylation in *Arabidopsis*. *Plos Genetics*, *10*(7), e1004446.
- Dizdaroglu, M., Karakaya, A., Jaruga, P., SLUPPHAUG, G., & Krokan, H. E. (1996). Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage. *Nucleic Acids Res*, *24*(3), 418–422.
- Dong, L., Mi, R., Glass, R. A., Barry, J. N., & Cao, W. (2008). Repair of deaminated base damage by *Schizosaccharomyces pombe* thymine DNA glycosylase. *DNA Repair*, *7*(12), 1962–1972.

- Doseth, B., Visnes, T., Wallenius, A., Ericsson, I., Sarno, A., Pettersen, H. S., et al. (2011). Uracil-DNA glycosylase in base excision repair and adaptive immunity: species differences between man and mouse. *J Biol Chem*, *286*(19), 16669–16680.
- Dubois, E., Córdoba-Cañero, D., Massot, S., Siaud, N., Gakière, B., Domenichini, S., et al. (2011). Homologous recombination is stimulated by a decrease in dUTPase in Arabidopsis. *PLoS One*, *6*(4), e18658.
- Durand-Dubief, M., Jenna, P., (null), Norman, U., Hartsuiker, E., & Ekwall, K. (2010). Topoisomerase I regulates open chromatin and controls gene expression in vivo. *EMBO J*, *29*(13), 2126–2134.
- Edwards, K. A., Halligan, B. D., Davis, J. L., Nivera, N. L., & Liu, L. F. (1982). Recognition sites of eukaryotic DNA topoisomerase I: DNA nucleotide sequencing analysis of topo I cleavage sites on SV40 DNA. *Nucleic Acids Res*, *10*(8), 2565–2576.
- Eftedal, I., Guddal, P. H., Slupphaug, G., Volden, G., & Krokan, H. E. (1993). Consensus sequences for good and poor removal of uracil from double stranded DNA by uracil-DNA glycosylase. *Nucleic Acids Res*, *21*(9), 2095.
- Egel, R. (2005). Fission yeast mating-type switching: programmed damage and repair. *DNA Repair*, *4*(5), 525–536.
- el-Hajj, H. H., Zhang, H., & Weiss, B. (1988). Lethality of a Dut (Deoxyuridine Triphosphatase) Mutation in Escherichia-Coli. *J Bacteriol*, *170*(3), 1069–1075.
- Elder, R. T., Zhu, X., Priet, S., Chen, M., Yu, M., Navarro, J. M., et al. (2003). A fission yeast homologue of the human uracil-DNA-glycosylase and their roles in causing DNA damage after overexpression. *Biochemical and Biophysical Research Communications*, *306*(3), 693–700.
- Eydmann, T., Sommariva, E., Inagawa, T., Mian, S., Klar, A. J. S., & Dalgaard, J. Z. (2008). Rtf1-mediated eukaryotic site-specific replication termination. *Genetics*, *180*(1), 27–39.
- Fadda, E., & Pomès, R. (2011). On the molecular basis of uracil recognition in DNA: comparative study of T-A versus U-A structure, dynamics and open base pair kinetics. *Nucleic Acids Res*, *39*(2), 767–780.
- Franco, D., Sgrignani, J., Bussi, G., & Magistrato, A. (2013). Structural Role of Uracil DNA Glycosylase for the Recognition of Uracil in DNA Duplexes. Clues from Atomistic Simulations. *J Chem Infn and Model*, *53*(6), 1371–1387.
- French, S. L., Osheim, Y. N., Cioci, F., Nomura, M., & Beyer, A. L. (2003). In exponentially growing *Saccharomyces cerevisiae* cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. *Mol Cell Biol*, *23*(5), 1558–1568.
- Friedberg, E. C., Walker, G. C., Siede, W., Wood, R. D. (2006). DNA Repair And Mutagenesis. American Society for Microbiology Press.
- Friedman, J. I., Majumdar, A., & Stivers, J. T. (2009). Nontarget DNA binding shapes the dynamic landscape for enzymatic recognition of DNA damage. *Nucleic Acids Res*, *37*(11), 3493–3500.
- Fritsch, O., Burkhalter, M. D., Kais, S., Sogo, J. M., & Schär, P. (2010). DNA ligase 4 stabilizes the ribosomal DNA array upon fork collapse at the replication fork barrier. *DNA Repair*, *9*(8), 879–888.
- Gadsden, M. H., McIntosh, E. M., Game, J. C., Wilson, P. J., & Haynes, R. H. (1993). dUTP pyrophosphatase is an essential enzyme in *Saccharomyces cerevisiae*. *EMBO J*, *12*(11), 4425–4431.
- Galashevskaya, A., Sarno, A., Vågbø, C. B., Aas, P. A., Hagen, L., Slupphaug, G., & Krokan, H. E. (2013). A robust, sensitive assay for genomic uracil determination by LC/MS/MS reveals lower levels than previously reported. *DNA Repair*, *12*(9), 699–706.
- Gallinari, P., & Jiricny, J. (1996). A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature*, *383*(6602), 735–738.
- Gavin, I., Horn, P. J., & Peterson, C. L. (2001). SWI/SNF chromatin remodeling requires changes in DNA topology. *Mol Cell*, *7*(1), 97–104.
- Gerber, J., Gogel, E., Berger, C., Wallisch, M., Muller, F., Grummt, I., & Grummt, F. (1997). Termination of mammalian rDNA replication: Polar arrest of replication fork movement by transcription termination factor TTF-I. *Cell*, *90*(3), 559–567.
- Gilbert, N., & Allan, J. (2014). Supercoiling in DNA and chromatin. *Curr Opin Genet Dev*, *25*, 15–21.



- Globisch, D., Münzel, M., Müller, M., Michalakis, S., Wagner, M., Koch, S., et al. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One*, *5*(12), e15367.
- Gomez-González, B., García-Rubio, M., Bermejo, R., Gaillard, H., Shirahige, K., Marín, A., et al. (2011). Genome-wide function of THO/TREX in active genes prevents R-loop-dependent replication obstacles. *EMBO J*, *30*(15), 3106–3119.
- Goodman, R. H., & Smolik, S. (2000). CBP/p300 in cell growth, transformation, and development. *Gene Dev*, *14*(13), 1553–1577.
- Greer, E. L., & Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet*, *13*(5), 343–357.
- Gruber, M., Wellinger, R. E., & Sogo, J. M. (2000). Architecture of the replication fork stalled at the 3' end of yeast ribosomal genes. *Mol Cell Biol*, *20*(15), 5777–5787.
- Guillet, M., Van der Kemp, P. A., & Boiteux, S. (2006). dUTPase activity is critical to maintain genetic stability in *Saccharomyces cerevisiae*. *Nucleic Acids Res*, *34*(7), 2056–2066.
- Hagen, L., Kavli, B., Sousa, M. M. L., Torseth, K., Liabakk, N. B., Sundheim, O., et al. (2008). Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA. *EMBO J*, *27*(1), 51–61.
- Hakem, R. (2008). DNA-damage repair; the good, the bad, and the ugly. *EMBO J*, *27*(4), 589–605.
- Hardeland, U., Bentele, M., Jiricny, J., & Schar, P. (2000). Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J Biol Chem*, *275*(43), 33449–33456.
- Hardeland, U., Bentele, M., Jiricny, J., & Schar, P. (2003). The versatile thymine DNA-glycosylase: a comparative characterization of the human, *Drosophila* and fission yeast orthologs. *Nucleic Acids Res*, *31*(9), 2261–2271.
- Hardeland, U., Kunz, C., Focke, F., Szadkowski, M., & Schar, P. (2007). Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2. *Nucleic Acids Res*, *35*(11), 3859–3867.
- Hardeland, U., Steinacher, R., Jiricny, J., & Schar, P. (2002). Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *EMBO J*, *21*(6), 1456–1464.
- Hashash, N., Johnson, A. L., & Cha, R. S. (2011). Regulation of fragile sites expression in budding yeast by MEC1, RRM3 and hydroxyurea. *J of Cell Sci*, *124*(Pt 2), 181–185.
- Hashash, N., Johnson, A. L., & Cha, R. S. (2012). Topoisomerase II- and condensin-dependent breakage of MEC1ATR-sensitive fragile sites occurs independently of spindle tension, anaphase, or cytokinesis. *Plos Genetics*, *8*(10), e1002978.
- He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, *333*(6047), 1303–1307.
- Heldring, N., Pike, A., Andersson, S., Matthews, J., Cheng, G., Hartman, J., et al. (2007). Estrogen Receptors: How Do They Signal and What Are Their Targets. *Physiological Reviews*, *87*(3), 905–931.
- Helland, D. E., Markussen, F.-H., Olsen, L. C., Aasland, R., Aarsaether, N., Bakke, O., & Krokan, H. E. (1993). Nuclear and mitochondrial forms of human uracil-DNA glycosylase are encoded by the same gene. *Nucleic Acids Res*, *21*(11), 2579.
- Helmrich, A., Ballarino, M., Nudler, E., & Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. *Nat Struct Mol Biol*, *20*(4), 412–418.
- Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., & Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature*, *401*(6750), 301–304.
- Heyer, W.-D., Ehmsen, K. T., & Liu, J. (2010). Regulation of homologous recombination in eukaryotes. *Ann Rev of Genet*, *44*, 113–139.
- Hodgson, B., Calzada, A., & Labib, K. (2007). Mrc1 and Tof1 regulate DNA replication forks in different ways during normal S phase. *Mol Biol Cell*, *18*(10), 3894–3902.
- Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature*, *411*(6835), 366–374.
- Horváth, A., Békési, A., Muha, V., Erdélyi, M., & Vértessy, B. G. (2013). Expanding the DNA alphabet

- in the fruit fly: Uracil enrichment in genomic DNA. *Fly*, 7(1), 23.
- Hu, J., Sun, L., Shen, F., Chen, Y., Hua, Y., Liu, Y., et al. (2012). The intra-S phase checkpoint targets Dna2 to prevent stalled replication forks from reversing. *Cell*, 149(6), 1221–1232.
- Hu, J., Ma, A., & Dinner, A. R. (2008). A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proc Natl Acad Sci US A*, 105(12), 4615–4620.
- Huang, G. S., & Keil, R. L. (1995). Requirements for activity of the yeast mitotic recombination hotspot HOT1: RNA polymerase I and multiple cis-acting sequences. *Genetics*, 141(3), 845–855.
- Huang, J., & Moazed, D. (2003). Association of the RENT complex with nontranscribed and coding regions of rDNA and a regional requirement for the replication fork block protein Fob1 in rDNA silencing. *Gene Dev*, 17(17), 2162–2176.
- Huang, J., Brito, I. L., Villen, J., Gygi, S. P., Amon, A., & Moazed, D. (2006). Inhibition of homologous recombination by a cohesin-associated clamp complex recruited to the rDNA recombination enhancer. *Gene Dev*, 20(20), 2887–2901.
- Ikeda, M., Ikeda, R., Ikeda, S. (2009). Spontaneous mutation in uracil DNA glycosylase-deficient cells of a fission yeast *Schizosaccharomyces pombe*. *Current Topics in Biochemical Research*. 11(1), 55–60.
- Interthal, H., & Champoux, J. J. (2011). Effects of DNA and protein size on substrate cleavage by human tyrosyl-DNA phosphodiesterase (TDP1). *The Biochemical Journal*, 436(3), 559–566.
- Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 466(7310), 1129–1133.
- Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., et al. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science*, 333(6047), 1300–1303.
- Ivessa, A. S., Lenzmeier, B. A., Bessler, J. B., Goudsouzian, L. K., Schnakenberg, S. L., & Zakian, V. A. (2003). The *Saccharomyces cerevisiae* helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Mol Cell*, 12(6), 1525–1536.
- Ivessa, A. S., Zhou, J. Q., & Zakian, V. A. (2000). The *Saccharomyces* Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell*, 100(4), 479–489.
- Jacobs, A. L., & Schar, P. (2012). DNA glycosylases: in DNA repair and beyond. *Chromosoma*, 121(1), 1–20.
- Jagelská, E. B., Brázda, V., Pecinka, P., Palecek, E., & Fojta, M. (2008). DNA topology influences p53 sequence-specific DNA binding through structural transitions within the target sites. *The Biochemical Journal*, 412(1), 57–63.
- Jia, Y., Nie, F., Du, A., Chen, Z., Qin, Y., Huang, T., et al. (2014). Thymine DNA glycosylase promotes transactivation of  $\beta$ -catenin/TCFs by cooperating with CBP. *Journal of Mol Cell Biology*, 6(3), 231–239.
- Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nature Reviews Mol Cell Biology*, 7(5), 335–346.
- Jobert, L., Skjeldam, H. K., Dalhus, B., Galashevskaya, A., Vågbo, C. B., Bjørås, M., & Nilsen, H. (2013). The Human Base Excision Repair Enzyme SMUG1 Directly Interacts with DKC1 and Contributes to RNA Quality Control. *Mol Cell*, 49(2), 339–345.
- Johzuka, K., & Horiuchi, T. (2002). Replication fork block protein, Fob1, acts as an rDNA region specific recombinator in *S. cerevisiae*. *Genes Cell*, 7(2), 99–113.
- Johzuka, K., Terasawa, M., Ogawa, H., Ogawa, T., & Horiuchi, T. (2006). Condensin loaded onto the replication fork barrier site in the rRNA gene repeats during S phase in a FOB1-dependent fashion to prevent contraction of a long repetitive array in *Saccharomyces cerevisiae*. *Mol Cell Biol*, 26(6), 2226–2236.
- Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*, 13(7), 484–492.
- Impellizzeri, KJ, B. A. P. M. B. (1991). The spectrum of spontaneous mutations in a *Saccharomyces cerevisiae* uracil-DNA-glycosylase mutant limits the function of this enzyme to cytosine deamination repair. *J Bacteriol*, 173(21), 6807.

- Kaerberlein, M., McVey, M., & Guarente, L. (1999). The SIR2/3/4 complex and SIR2 alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Gene Dev*, *13*(19), 2570–2580.
- Kamileri, I., Karakasiloti, I., & Garinis, G. A. (2012). Nucleotide excision repair: new tricks with old bricks. *Trends Genet*, *28*(11), 566–573.
- Kanamitsu, K., & Ikeda, S. (2010). Early Steps in the DNA Base Excision Repair Pathway of a Fission Yeast *Schizosaccharomyces pombe*. *J Nucleic Acids*, *2010*(7), 1–9.
- Kangaspeska, S., Stride, B., Métivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R. P., et al. (2008). Transient cyclical methylation of promoter DNA. *Nature*, *452*(7183), 112–115.
- Kaplan, D. L., & Bastia, D. (2009). Mechanisms of polar arrest of a replication fork. *Mol Microbiol*, *72*(2), 279–285.
- Karumbati, A. S., Deshpande, R. A., Jilani, A., Vance, J. R., Ramotar, D., & Wilson, T. E. (2003). The role of yeast DNA 3'-phosphatase Tpp1 and rad1/Rad10 endonuclease in processing spontaneous and induced base lesions. *J Biol Chem*, *278*(33), 31434–31443.
- Kavli, B., Otterlei, M., Slupphaug, G., & Krokan, H. E. (2007). Uracil in DNA—General mutagen, but normal intermediate in acquired immunity. *DNA Repair*, *6*(4), 505–516.
- Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., et al. (2002). 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*(42), 39926–39936.
- Keeney, S. (2008). Spo11 and the Formation of DNA Double-Strand Breaks in Meiosis. *Genome Dyn Stab*, *2*, 81–123.
- Kegel, A., Betts-Lindroos, H., Kanno, T., Jeppsson, K., Ström, L., Katou, Y., et al. (2011). Chromosome length influences replication-induced topological stress. *Nature*, *471*(7338), 392–396.
- Keil, R. L., & McWilliams, A. D. (1993). A gene with specific and global effects on recombination of sequences from tandemly repeated genes in *Saccharomyces cerevisiae*. *Genetics*, *135*(3), 711–718.
- Keil, R. L., & Roeder, G. S. (1984). Cis-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of *S. cerevisiae*. *Cell*, *39*(2 Pt 1), 377–386.
- Kelley, M. R., Kow, Y. W., & Wilson, D. M. (2003). Disparity between DNA base excision repair in yeast and mammals: translational implications. *Cancer Research*, *63*(3), 549–554.
- Kemmerich, K., Dingler, F. A., Rada, C., & Neuberger, M. S. (2012). Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung<sup>-/-</sup>Msh2<sup>-/-</sup> mice. *Nucleic Acids Res*, *40*(13), 6016–6025.
- Kim, M.-S., Kondo, T., Takada, I., Youn, M.-Y., Yamamoto, Y., Takahashi, S., et al. (2009). DNA demethylation in hormone-induced transcriptional derepression. *Nature*, *461*(7266), 1007–1012.
- Kim, N. & Jinks-Robertson, S. (2012). Transcription as a source of genome instability. *Nat Rev Genet*, *13*(3), 204–214.
- Kim, N., Huang, S. N., Williams, J. S., Li, Y. C., Clark, A. B., et al. (2011). Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science*, *332*(6037), 1561–1564.
- Kim, R. A., & Wang, J. C. (1989). A subthreshold level of DNA topoisomerases leads to the excision of yeast rDNA as extrachromosomal rings. *Cell*, *57*(6), 975–985.
- Kim, Y.-J., Wilson D. M. III (2012). Overview of Base Excision Repair Biochemistry. *Current Molecular Pharmacology*, *5*(1), 3.
- Kobayashi, T. (2003). The replication fork barrier site forms a unique structure with Fob1p and inhibits the replication fork. *Mol Cell Biol*, *23*(24), 9178–9188.
- Kobayashi, T. (2005). Recombination Regulation by Transcription-Induced Cohesin Dissociation in rDNA Repeats. *Science*, *309*(5740), 1581–1584.
- Kobayashi, T. (2006). Strategies to maintain the stability of the ribosomal RNA gene repeats--collaboration of recombination, cohesion, and condensation. *Genes Genet Syst*, *81*(3), 155–161.
- Kobayashi, T. (2011). How does genome instability affect lifespan? *Genes Cells* *16*(6), 617–624.
- Kobayashi, T., & Horiuchi, T. (1996). A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities. *Genes Cells* *1*(5), 465–474.
- Kobayashi, T., Heck, D. J., Nomura, M., & Horiuchi, T. (1998). Expansion and contraction of ribosomal

- DNA repeats in *Saccharomyces cerevisiae*: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Gene Dev*, *12*(24), 3821–3830.
- Kobayashi, T., Horiuchi, T., Tongaonkar, P., Vu, L., & Nomura, M. (2004). SIR2 regulates recombination between different rDNA repeats, but not recombination within individual rRNA genes in yeast. *Cell*, *117*(4), 441–453.
- Kohli, R. M., & Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, *502*(7472), 472–479.
- Kondo, E., Gu, Z., Horii, A., & Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. *Mol Cell Biol*, *25*(11), 4388–4396.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, *128*(4), 693–705.
- Kouzine, F., Sanford, S., Elisha-Feil, Z., & Levens, D. (2008). The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat Struct Mol Biol*, *15*(2), 146–154.
- Krejci, L., Altmannova, V., Spirek, M., & Zhao, X. (2012). Homologous recombination and its regulation. *Nucleic Acids Res*, *40*(13), 5795–5818.
- Kriaucionis, S., & Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*, *324*(5929), 929–930.
- Krings, G., & Bastia, D. (2004). swi1- and swi3-dependent and independent replication fork arrest at the ribosomal DNA of *Schizosaccharomyces pombe*. *Proc Natl Acad Sci US A*, *101*(39), 14085–14090.
- Krings, G., & Bastia, D. (2005). Sap1p binds to Ter1 at the ribosomal DNA of *Schizosaccharomyces pombe* and causes polar replication fork arrest. *J Biol Chem*, *280*(47), 39135–39142.
- Krokan, H. E., Drabløs, F., & Slupphaug, G. (2002). Uracil in DNA—occurrence, consequences and repair. *Oncogene*, *21*(58), 8935–8948.
- Kunz, C., Focke, F., Saito, Y., Schuermann, D., Lettieri, T., Selfridge, J., et al. (2009). Base Excision by Thymine DNA Glycosylase Mediates DNA-Directed Cytotoxicity of 5-Fluorouracil. *PLoS Biology*, *7*(4), e91.
- Kuramoto, Y., Hata, K., Koyanagi, S., Ohdo, S., Shimeno, H., & Soeda, S. (2006). Circadian regulation of mouse topoisomerase I gene expression by glucocorticoid hormones. *Biochem Pharmacol*, *71*(8), 1155–1161.
- Laerdahl, J. K., Korvald, H., Nilsen, L., Dahl-Michelsen, K., Rognes, T., Bjørås, M., & Alseth, I. (2011). *Schizosaccharomyces pombe* encodes a mutated AP endonuclease 1. *DNA Repair*, *10*(3), 296–305.
- Laloraya, S., Guacci, V., & Koshland, D. (2000). Chromosomal addresses of the cohesin component Mcd1p. *J Cell Biol*, *151*(5), 1047–1056.
- Lambert, S., & Carr, A. M. (2013). Impediments to replication fork movement: stabilisation, reactivation and genome instability. *Chromosoma*, *122*(1-2), 33–45.
- Lambert, S., Watson, A., Sheedy, D., Martin, B., & Carr, A. (2005). Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell*, *121*(5), 689–702.
- Lee, C.-K., Shibata, Y., Rao, B., Strahl, B. D., & Lieb, J. D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature Genetics*, *36*(8), 900–905.
- Leman, A. R., & Noguchi, E. (2013). The replication fork: understanding the eukaryotic replication machinery and the challenges to genome duplication. *Genes*, *4*(1), 1–32.
- Leppard, J. B., & Champoux, J. J. (2005). Human DNA topoisomerase I: relaxation, roles, and damage control. *Chromosoma*, *114*(2), 75–85.
- Li, G.-M. (2008). Mechanisms and functions of DNA mismatch repair. *Cell Research*, *18*(1), 85–98.
- Li, Y.-Q., Zhou, P.-Z., Zheng, X.-D., Walsh, C. P., & Xu, G.-L. (2007). Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. *Nucleic Acids Res*, *35*(2), 390–400.
- Lindahl, T., Ljungquist, S., Siegert, W., Nyberg, B., & Sperens, B. (1977). DNA N-glycosidases: properties of uracil-DNA glycosidase from *Escherichia coli*. *J Biol Chem*, *252*(10), 3286–3294.
- Lippert, M. J., Kim, N., Cho, J. E., Larson, R. P., Schoenly, N. E., et al. (2011). Role for topoisomerase 1

- in transcription-associated mutagenesis in yeast. *Proc Natl Acad Sci US A*, 108(2), 698–703.
- Liu, C., Pouliot, J. J., & Nash, H. A. (2002). Repair of topoisomerase I covalent complexes in the absence of the tyrosyl-DNA phosphodiesterase Tdp1. *Proc Natl Acad Sci USA*, 99(23), 14970–14975.
- Liutkeviciute, Z., Lukinavicius, G., Masevicius, V., Daujotyte, D., & Klimasauskas, S. (2009). Cytosine-5-methyltransferases add aldehydes to DNA. *Nature Chemical Biology*, 5(6), 400–402.
- Longley, D. B., Harkin, D. P., & Johnston, P. G. (2003). 5-fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer*, 3(5), 330–338.
- Lopes, M., Cotta-Ramusino, C., Pelliccioli, A., Liberi, G., Plevani, P., Muzi-Falconi, M., et al. (2001). The DNA replication checkpoint response stabilizes stalled replication forks. *Nature*, 412(6846), 557–561.
- Lühnsdorf, B., Epe, B., & Khobta, A. (2014). Excision of Uracil from Transcribed DNA Negatively Affects Gene Expression. *J Biol Chem*, 289(32), 22008–22018.
- Lyu, Y. L., Lin, C.-P., Azarova, A. M., Cai, L., Wang, J. C., & Liu, L. F. (2006). Role of topoisomerase IIbeta in the expression of developmentally regulated genes. *Mol Cell Biol*, 26(21), 7929–7941.
- Ma, J., Bai, L., & Wang, M. D. (2013). Transcription Under Torsion. *Science*, 340(6140), 1580–1583.
- Madabushi, A., Hwang, B. J., Jin, J., & Lu, A.-L. (2013). Histone deacetylase SIRT1 modulates and deacetylates DNA base excision repair enzyme thymine DNA glycosylase. *The Biochemical Journal*, 456(1), 89–98.
- Magnani, L., & Lupien, M. (2014). Chromatin and epigenetic determinants of estrogen receptor alpha (ESR1) signaling. *Mol Cell Endocrinol*, 382(1), 633–641.
- Maiti, A., & Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem*, 286(41), 35334–35338.
- Maiti, A., Morgan, M. T., Pozharski, E., & Drohat, A. C. (2008). Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proc Natl Acad Sci US A*, 105(26), 8890–8895.
- Malanga, M., Czuby, A., Girstun, A., Staron, K., & Althaus, F. R. (2008). Poly(ADP-ribose) binds to the splicing factor ASF/SF2 and regulates its phosphorylation by DNA topoisomerase I. *J Biol Chem*, 283(29), 19991–19998.
- Marques, M., Laflamme, L., & Gaudreau, L. (2013). Estrogen receptor  $\alpha$  can selectively repress dioxin receptor-mediated gene expression by targeting DNA methylation. *Nucleic Acids Res*, 41(17), 8094–8106.
- Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205), 766–770.
- Mejía-Ramírez, E., Sánchez-Gorostiaga, A., Krimer, D. B., Schwartzman, J. B., & Hernández, P. (2005). The mating type switch-activating protein Sap1 is required for replication fork arrest at the rRNA genes of fission yeast. *Mol Cell Biol*, 25(19), 8755–8761.
- Mekhail, K., Seebacher, J., Gygi, S. P., & Moazed, D. (2008). Role for perinuclear chromosome tethering in maintenance of genome stability. *Nature*, 456(7222), 667–670.
- Merino, A., Madden, K. R., Lane, W. S., Champoux, J. J., & Reinberg, D. (1993). DNA topoisomerase I is involved in both repression and activation of transcription. *Nature*, 365(6443), 227–232.
- Métivier, R., Gallais, R., Tiffocche, C., Le Péron, C., Jurkowska, R. Z., Carmouche, R. P., et al. (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature*, 452(7183), 45–50.
- Métivier, R., Penot, G., Hübner, M. R., Reid, G., Brand, H., Kos, M., & Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell*, 115(6), 751–763.
- Miao, Z.-H., Player, A., Shankavaram, U., Wang, Y.-H., Zimonjic, D. B., Lorenzi, P. L., et al. (2007). Nonclassic functions of human topoisomerase I: genome-wide and pharmacologic analyses. *Cancer Research*, 67(18), 8752–8761.
- Millar, C. B., & Grunstein, M. (2006). Genome-wide patterns of histone modifications in yeast. *Nature Reviews Mol Cell Biology*, 7(9), 657–666.
- Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., Macdougall, E., Hendrich, B., et al. (2002). Enhanced

- CpG mutability and tumorigenesis in MBD4-deficient mice. *Science*, 297(5580), 403–405.
- Mirkin, E. V., & Mirkin, S. M. (2007). Replication fork stalling at natural impediments. *Microbiol Mol Biol Rev*, 71(1), 13–35.
- Missero, C., Pirro, M. T., Simeone, S., Pischetola, M., & Di Lauro, R. (2001). The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription. *J Biol Chem*, 276(36), 33569–33575.
- Mohanty, B. K., Bairwa, N. K., & Bastia, D. (2006). The Tof1p-Csm3p protein complex counteracts the Rrm3p helicase to control replication termination of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*, 103(4), 897–902.
- Mokkapati, S. K., Fernández de Henestrosa, A. R., & Bhagwat, A. S. (2001). *Escherichia coli* DNA glycosylase Mug: a growth-regulated enzyme required for mutation avoidance in stationary-phase cells. *Mol Microbiol*, 41(5), 1101–1111.
- Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., & Tainer, J. A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: Structural basis for specificity and catalysis. *Cell*, 80(6), 869–878.
- Morgan, H. D., Dean, W., Coker, H. A., Reik, W., & Petersen Mahrt, S. K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem*, 279(50), 52353–52360.
- Muha, V., Horváth, A., Békési, A., Pukáncsik, M., Hodoscsek, B., Merényi, G., et al. (2012). Uracil-containing DNA in *Drosophila*: stability, stage-specific accumulation, and developmental involvement. *Plos Genetics*, 8(6), e1002738.
- Müller, U., Bauer, C., Siegl, M., Rottach, A., & Leonhardt, H. (2014). TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res*, 42(13), 8592–604.
- Muller, M. T., Pfund, W. P., Mehta, V. B., Trask, D. K. (1985). Eukaryotic type I topoisomerase is enriched in the nucleolus and catalytically active on ribosomal DNA. *EMBO J*, 4(5), 1237.
- Nabel, C. S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H. L., Stivers, J. T., et al. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nature Chemical Biology*, 8(9), 751–758.
- Nagaria, P., Svilar, D., Brown, A. R., Wang, X.-H., Sobol, R. W., & Wyatt, M. D. (2013). SMUG1 but not UNG DNA glycosylase contributes to the cellular response to recovery from 5-fluorouracil induced replication stress. *Mutation Research*, 743-744, 26–32.
- Naughton, C., Avlonitis, N., Corless, S., Prendergast, J. G., Mati, I. K., Eijk, P. P., et al. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat Struct Mol Biol*, 20(3), 387–395.
- Neddermann, P., & Jiricny, J. (1994). Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. *Proc Natl Acad Sci USA*, 91(5), 1642–1646.
- Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L., & Lindahl, T. (2001). Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. *EMBO J*, 20(15), 4278–4286.
- Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C. F., Andersen, S., SLUPPHAUG, G., et al. (2000). Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. *Mol Cell*, 5(6), 1059–1065.
- Nilsen, H., Stamp, G., Andersen, S., Hrivnak, G., Krokan, H. E., Lindahl, T., & Barnes, D. E. (2003). Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas. *Oncogene*, 22(35), 5381–5386.
- O'Neill, R. J., Vorob'eva, O. V., Shahbakhti, H., Zmuda, E., Bhagwat, A. S., & Baldwin, G. S. (2003). Mismatch Uracil Glycosylase from *Escherichia coli*: A GENERAL MISMATCH OR A SPECIFIC DNA GLYCOSYLASE? *J Biol Chem*, 278(23), 20526–20532.
- Olsen, L. C., Aasland, R., Wittwer, C. U., Krokan, H. E., & Helland, D. E. (1989). Molecular cloning of human uracil-DNA glycosylase, a highly conserved DNA repair enzyme. *EMBO J*, 8(10), 3121–3125.
- Olsson, I., Bjerling, P. (2011). Advancing our understanding of functional genome organisation

- through studies in the fission yeast. *Curr Genet*, 57(1), 1-12.
- Onishi, Y., & Kawano, Y. (2012). Rhythmic binding of Topoisomerase I impacts on the transcription of Bmal1 and circadian period. *Nucleic Acids Res*, 40(19), 9482–9492.
- Otterlei, Haug, T., Nagelhus, T. A., SLUPPHAUG, G., Lindmo, T., & Krokan, H. E. (1998). Nuclear and mitochondrial splice forms of human uracil-DNA glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively. *Nucleic Acids Res*, 26(20), 4611–4617.
- Otterlei, M., Warbrick, E., Nagelhus, T., Haug, T., Slupphaug, G., Akbari, M., et al. (1999). Post-replicative base excision repair in replication foci. *EMBO J*, 18(13), 3834–3844.
- Pardo, B., Gomez-González, B., & Aguilera, A. (2009). DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. *Cell Mol Life Sci*, 66(6), 1039–1056.
- Park, H., & Sternglanz, R. (1999). Identification and characterization of the genes for two topoisomerase I-interacting proteins from *Saccharomyces cerevisiae*. *Yeast*, 15(1), 35–41.
- Parker, J. B., Bianchet, M. A., Krosky, D. J., Friedman, J. I., Amzel, L. M., & Stivers, J. T. (2007). Enzymatic capture of an extrahelical thymine in the search for uracil in DNA. *Nature*, 449(7161), 433–437.
- Pedersen, J. M., Fredsoe, J., Roedgaard, M., Andreasen, L., Mundbjerg, K., Kruhøffer, M., et al. (2012). DNA Topoisomerases Maintain Promoters in a State Competent for Transcriptional Activation in *Saccharomyces cerevisiae*. *Plos Genetics*, 8(12), e1003128.
- Pommier, Y., Barcelo, J. M., Rao, V. A., Sordet, O., Jobson, A. G., Thibaut, L., et al. (2006). Repair of topoisomerase I-mediated DNA damage. *Prog Nucleic Acid Res Mol Biol*, 81, 179–229.
- Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., et al. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*.
- Postow, L., Crisona, N. J., Peter, B. J., Hardy, C. D., & Cozzarelli, N. R. (2001). Topological challenges to DNA replication: conformations at the fork. *Proc Natl Acad Sci USA*, 98(15), 8219–8226.
- Pouliot, J. J., Yao, K. C., Robertson, C. A., & Nash, H. A. (1999). Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. *Science*, 286(5439), 552–555.
- Putter, V., & Grummt, F. (2002). Transcription termination factor TTF-I exhibits contrahelicase activity during DNA replication. *Embo Reports*, 3(2), 147–152.
- Rai, K., Huggins, I. J., James, S. R., Karpf, A. R., Jones, D. A., & Cairns, B. R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell*, 135(7), 1201–1212.
- Reid, R. J., Benedetti, P., & Bjornsti, M. A. (1998). Yeast as a model organism for studying the actions of DNA topoisomerase-targeted drugs. *Biochimica Et Biophysica Acta*, 1400(1-3), 289–300.
- Robertson, A. B., Klungland, A., Rognes, T., & Leiros, I. (2009). DNA Repair in Mammalian Cells. *Cell Mol Life Sci*, 66(6), 981–993.
- Rogstad, D. K., Liu, P., Burdzy, A., Lin, S. S., & Sowers, L. C. (2002). Endogenous DNA lesions can inhibit the binding of the AP-1 (c-Jun) transcription factor. *Biochemistry*, 41(25), 8093–8102.
- Rosche, W. A., Trinh, T. Q., & Sinden, R. R. (1995). Differential DNA secondary structure-mediated deletion mutation in the leading and lagging strands. *J Bacteriol*, 177(15), 4385–4391.
- Rossi, F., Labourier, E., Forné, T., Divita, G., Derancourt, J., Riou, J. F., et al. (1996). Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature*, 381(6577), 80–82.
- Rüegg, J., Cai, W., Karimi, M., Kiss, N. B., Swedenborg, E., Larsson, C., et al. (2011). Epigenetic regulation of glucose transporter 4 by estrogen receptor  $\beta$ . *Mol Endocrinol*, 25(12), 2017–2028.
- Sansom, O. J., Bishop, S. M., Bird, A., & Clarke, A. R. (2004). MBD4 deficiency does not increase mutation or accelerate tumorigenesis in mice lacking MMR. *Oncogene*, 23(33), 5693–5696.
- Sansom, O. J., Zabkiewicz, J., Bishop, S. M., Guy, J., Bird, A., & Clarke, A. R. (2003). MBD4 deficiency reduces the apoptotic response to DNA-damaging agents in the murine small intestine. *Oncogene*, 22(46), 7130–7136.
- Santos, F., Peat, J., Burgess, H., Rada, C., Reik, W., & Dean, W. (2013). Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin*, 6(1), 39.
- Savva, R., & Pearl, L. H. (1995). Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase–uracil glycosylase inhibitor protein complex. *Nat Struct Mol Biol*, 2(9), 752–757.

- Sánchez-Gorostiaga, A., López-Estraño, C., Krimer, D. B., Schwartzman, J. B., & Hernández, P. (2004). Transcription termination factor reb1p causes two replication fork barriers at its cognate sites in fission yeast ribosomal DNA in vivo. *Mol Cell Biol*, 24(1), 398–406.
- Schärer, O. D. (2003). Chemistry and biology of DNA repair. *Angewandte Chemie (International Ed. in English)*, 42(26), 2946–2974.
- Schärer, O. D. (2013). Nucleotide excision repair in eukaryotes. *Cold Spring Harb Perspect Biol*, 5(10), a012609.
- Schrader, C. E., Linehan, E. K., Mochegova, S. N., Woodland, R. T., & Stavnezer, J. (2005). Inducible DNA breaks in Ig S regions are dependent on AID and UNG. *J Exp Med*, 202(4), 561–568.
- Sekiguchi, J. & Shuman, S. (1997). Site-specific ribonuclease activity of eukaryotic DNA topoisomerase I. *Mol Cell*, 1(1), 89–97.
- Serizawa, N., Horiuchi, T., & Kobayashi, T. (2004). Transcription-mediated hyper-recombination in HOT1. *Genes Cells*, 9(4), 305–315.
- Sheinin, M. Y., Li, M., Soltani, M., Luger, K., & Wang, M. D. (2013). Torque modulates nucleosome stability and facilitates H2A/H2B dimer loss. *Nat Commun*, 4, 2579.
- Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A. C., Fung, H.-L., et al. (2013). Genome-wide Analysis Reveals TET- and TDG-Dependent 5-Methylcytosine Oxidation Dynamics. *Cell*, 153(3), 692–706.
- Shibata, E., Dar, A., & Dutta, A. (2014). CRL4Cdt2 E3 Ubiquitin Ligase and PCNA Cooperate to Degrade Thymine DNA Glycosylase in S-phase. *J Biol Chem*, 289(33), 23056–23064.
- Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., & Iyer, V. R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biology*, 6(3), e65.
- Shuman, S., & Prescott, J. (1990). Specific DNA cleavage and binding by vaccinia virus DNA topoisomerase I. *J Biol Chem*, 265(29), 17826–17836.
- Shykind, B. M., Kim, J., Stewart, L., Champoux, J. J., & Sharp, P. A. (1997). Topoisomerase I enhances TFIIID-TFIIA complex assembly during activation of transcription. *Gene Dev*, 11(3), 397–407.
- Sinclair, D. A., & Guarente, L. (1997). Extrachromosomal rDNA circles--a cause of aging in yeast. *Cell*, 91(7), 1033–1042.
- Sjolund, A. B., Senejani, A. G., & Sweasy, J. B. (2013). MBD4 and TDG: Multifaceted DNA glycosylases with ever expanding biological roles. *Mutation Research*, 743–744, 12–25.
- Slenn, T. J., Morris, B., Havens, C. G., Freeman, R. M., Takahashi, T. S., & Walter, J. C. (2014). Thymine DNA Glycosylase is a CRL4Cdt2 Substrate. *J Biol Chem*, 289(33), 23043–23055.
- Slupphaug, G., Eftedal, I., Kavli, B., Bharati, S., Helle, N. M., Haug, T., et al. (1995). Properties of a Recombinant Human Uracil-DNA Glycosylase from the UNG Gene and Evidence that UNG Encodes the Major Uracil-DNA Glycosylase. *Biochemistry*, 34(1), 128–138.
- Smith, J. S., Caputo, E., & Boeke, J. D. (1999). A genetic screen for ribosomal DNA silencing defects identifies multiple DNA replication and chromatin-modulating factors. *Mol Cell Biol*, 19(4), 3184–3197.
- Sogo, J. M., (null), Lopes, M., & Foiani, M. (2002). Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science*, 297(5581), 599–602.
- Song, C.-X., Szulwach, K. E., Dai, Q., Fu, Y., Mao, S.-Q., Lin, L., et al. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell*, 153(3), 678–691.
- Song, C.-X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., et al. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnology*, 29(1), 68–72.
- Sood, V., & Brickner, J. H. (2014). Nuclear pore interactions with the genome. *Curr Opin Genet Dev*, 25, 43–49.
- Sordet, O., Khan, Q. A., & Pommier, Y. (2004a). Apoptotic topoisomerase I-DNA complexes induced by oxygen radicals and mitochondrial dysfunction. *Cell Cycle*, 3(9), 1095–1097.
- Sordet, O., Khan, Q. A., Plo, I., Pourquier, P., Urasaki, Y., Yoshida, A., et al. (2004b). Apoptotic topoisomerase I-DNA complexes induced by staurosporine-mediated oxygen radicals. *The J Biol Chem*, 279(48), 50499–50504.



- Sordet, O., Liao, Z., Liu, H., Antony, S., Stevens, E. V., Kohlhagen, G., et al. (2004c). Topoisomerase I-DNA complexes contribute to arsenic trioxide-induced apoptosis. *J Biol Chem*, *279*(32), 33968–33975.
- Soret, J., Gabut, M., Dupon, C., Kohlhagen, G., Stévenin, J., Pommier, Y., & Tazi, J. (2003). Altered serine/arginine-rich protein phosphorylation and exonic enhancer-dependent splicing in Mammalian cells lacking topoisomerase I. *Cancer Research*, *63*(23), 8203–8211.
- Sousa, M. M. L., Krokan, H. E., & Slupphaug, G. (2007). DNA-uracil and human pathology. *Molecular Aspects of Medicine*, *28*(3-4), 276–306.
- Sperling, A. S., Jeong, K. S., Kitada, T., & Grunstein, M. (2011). Topoisomerase II binds nucleosome-free DNA and acts redundantly with topoisomerase I to enhance recruitment of RNA Pol II in budding yeast. *Proc Natl Acad Sci US A*, *108*(31), 12693–12698.
- Stavnezer, J., Guikema, J. E. J., & Schrader, C. E. (2008). Mechanism and Regulation of Class Switch Recombination. *Ann Rev Immunol*, *26*(1), 261–292.
- Steinacher, R., & Schar, P. (2005). Functionality of Human Thymine DNA Glycosylase Requires SUMO-Regulated Changes in Protein Conformation. *Curr Biol*, *15*(7), 616–623.
- Sullivan, M., Higuchi, T., Katis, V. L., & Uhlmann, F. (2004). Cdc14 phosphatase induces rDNA condensation and resolves cohesin-independent cohesion during budding yeast anaphase. *Cell*, *117*(4), 471–482.
- Sung, J.-S., & Demple, B. (2006). Roles of base excision repair subpathways in correcting oxidized abasic sites in DNA. *FEBS J*, *273*(8), 1620–1629.
- Suzuki, T., Ide, H., Yamada, M., Endo, N., Kanaori, K., Tajima, K., et al. (2000). Formation of 2'-deoxyoxanosine from 2'-deoxyguanosine and nitrous acid: mechanism and intermediates. *Nucleic Acids Res*, *28*(2), 544–551.
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, *324*(5929), 930–935.
- Takahashi, T., Burguiere-Slezak, G., Van der Kemp, P. A., & Boiteux, S. (2011). Topoisomerase 1 provokes the formation of short deletions in repeated sequences upon high transcription in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA*, *108*(2), 692–697.
- Takeuchi, Y. (2003). Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Gene Dev*, *17*(12), 1497–1506.
- Tammimies, K., Tapia-Páez, I., Rüegg, J., Rosin, G., Kere, J., Gustafsson, J.-Å., & Nalvarte, I. (2012). The rs3743205 SNP is important for the regulation of the dyslexia candidate gene DYX1C1 by estrogen receptor  $\beta$  and DNA methylation. *Mol Endocrinol*, *26*(4), 619–629.
- Tercero, J. A., & Diffley, J. F. (2001). Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature*, *412*(6846), 553–557.
- Teves, S. S., & Henikoff, S. (2014a). Transcription-generated torsional stress destabilizes nucleosomes. *Nat Struct Mol Biol*, *21*(1), 88–94.
- Teves, S., & Henikoff, S. (2014b). DNA torsion as a feedback mediator of transcription and chromatin dynamics. *Nucleus*, *5*(3), 1-8.
- Thomassin, H., Flavin, M., Espinás, M. L., & Grange, T. (2001). Glucocorticoid-induced DNA demethylation and gene memory during development. *EMBO J*, *20*(8), 1974–1983.
- Tini, M., Benecke, A., Um, S.-J., Torchia, J., Evans, R. M., & Chambon, P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol Cell*, *9*(2), 265–277.
- Tsang, E., & Carr, A. M. (2008). Replication fork arrest, recombination and the maintenance of ribosomal DNA stability. *DNA Repair*, *7*(10), 1613–1623.
- Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., et al. (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature Cell Biology*, *11*(11), 1315–1324.
- Udvardy, A., Schedl, P., Sander, M., & Hsieh, T. S. (1985). Novel partitioning of DNA cleavage sites for *Drosophila* topoisomerase II. *Cell*, *40*(4), 933–941.
- Um, S., Harbers, M., Benecke, A., Pierrat, B., Losson, R., & Chambon, P. (1998). Retinoic acid

- receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J Biol Chem*, 273(33), 20728–20736.
- Vance, J. R., & Wilson, T. E. (2002). Yeast Tdp1 and Rad1-Rad10 function as redundant pathways for repairing Top1 replicative damage. *Proc Natl Acad Sci USA*, 99(21), 13669–13674.
- Verri, A., Mazzarello, P., Biamonti, G., Spadari, S., & Focher, F. (1990). The specific binding of nuclear protein(s) to the cAMP responsive element (CRE) sequence (TGACGTCA) is reduced by the misincorporation of U and increased by the deamination of C. *Nucleic Acids Res*, 18(19), 5775.
- Visnes, T., Doseth, B., Pettersen, H. S., Hagen, L., Sousa, M. M. L., Akbari, M., et al. (2009). Uracil in DNA and its processing by different DNA glycosylases. *Philos Trans R Soc Lond, B, Biol Sci*, 364(1517), 563–568.
- Voelkel-Meiman, K., Keil, R. L., & Roeder, G. S. (1987). Recombination-stimulating sequences in yeast ribosomal DNA correspond to sequences regulating transcription by RNA polymerase I. *Cell*, 48(6), 1071–1079.
- Vogelauer, M., & Camilloni, G. (1999). Site-specific in vivo cleavages by DNA topoisomerase I in the regulatory regions of the 35 S rRNA in *Saccharomyces cerevisiae* are transcription independent. *J Mol Biol*, 293(1), 19–28.
- Vos, S. M., Lyubimov, A. Y., Hershey, D. M., Schoeffler, A. J., Sengupta, S., Nagaraja, V., & Berger, J. M. (2014). Direct control of type IIA topoisomerase activity by a chromosomally encoded regulatory protein. *Gene Dev*, 28(13), 1485–1497.
- Wang, J. H. (2013). The role of activation-induced deaminase in antibody diversification and genomic instability. *Immunol Res*, 55(1-3), 287–297.
- Ward, T. R., Hoang, M. L., Prusty, R., Lau, C. K., Keil, R. L., Fangman, W. L., & Brewer, B. J. (2000). Ribosomal DNA Replication Fork Barrier and HOT1 Recombination Hot Spot: Shared Sequences but Independent Activities. *Mol Cell Biol*, 20(13), 4948–4957.
- Warner, H. R., & Rockstroh, P. A. (1980). Incorporation and excision of 5-fluorouracil from deoxyribonucleic acid in *Escherichia coli*. *J Bacteriol*, 141(2), 680–686.
- Warner, H. R., Duncan, B. K., Garrett, C., & Neuhard, J. (1981). Synthesis and metabolism of uracil-containing deoxyribonucleic acid in *Escherichia coli*. *J Bacteriol*, 145(2), 687–695.
- Waters, T. R., & Swann, P. F. (1998). Kinetics of the Action of Thymine DNA Glycosylase. *J Biol Chem*, 273(32), 20007–20014.
- Waters, T. R., Gallinari, P., Jiricny, J., & Swann, P. F. (1999). Human Thymine DNA Glycosylase Binds to Apurinic Sites in DNA but Is Displaced by Human Apurinic Endonuclease 1. *J Biol Chem*, 274(1), 67–74.
- Weitao, T., Budd, M., Hoopes, L., & Campbell, J. (2003). Dna2 helicase/nuclease causes replicative fork stalling and double-strand breaks in the ribosomal DNA of *Saccharomyces cerevisiae*. *J Biol Chem*, 278(25), 22513–22522.
- Wiebauer, K., & Jiricny, J. (1989). In vitro correction of G.T mispairs to G.C pairs in nuclear extracts from human cells. *Nature*, 339(6221), 234–236.
- Wilkinson, C. R. M., Bartlett, R., Nurse, P., & Bird, A. P. (1995). The fission yeast gene *pmt1+* encodes a DNA methyltransferase homologue. *Nucleic Acids Res*, 23(2), 203.
- Wilson, S. H., & Kunkel, T. A. (2000, March). Passing the baton in base excision repair. *Nat Struct Mol Biol*, pp. 176–178.
- Wu, H., & Zhang, Y. (2011). Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Gene Dev*, 25(23), 2436–2452.
- Yang, B., Chen, K., Zhang, C., Huang, S., & Zhang, H. (2007). Virion-associated uracil DNA glycosylase-2 and apurinic/aprimidinic endonuclease are involved in the degradation of APOBEC3G-edited nascent HIV-1 DNA. *J Biol Chem*, 282(16), 11667–11675.
- Yang, F., Nakajima, Y., Kumagai, M., Ohmiya, Y., & Ikeda, M. (2009). The molecular mechanism regulating the autonomous circadian expression of Topoisomerase I in NIH3T3 cells. *Biochemical and Biophysical Research Communications*, 380(1), 22–27.
- Yang, S. W., Burgin, A. B., Huizenga, B. N., Robertson, C. A., Yao, K. C., & Nash, H. A. (1996). A eukaryotic enzyme that can disjoin dead-end covalent complexes between DNA and type I topoisomerases. *Proc Natl Acad Sci USA*, 93(21), 11534–11539.

- Yang, X., Li, W., Prescott, E. D., Burden, S. J., & Wang, J. C. (2000). DNA topoisomerase II $\beta$  and neural development. *Science*, 287(5450), 131–134.
- Zawadzki, K. A., Morozov, A. V., & Broach, J. R. (2009). Chromatin-dependent transcription factor accessibility rather than nucleosome remodeling predominates during global transcriptional restructuring in *Saccharomyces cerevisiae*. *Mol Biol Cell*, 20(15), 3503–3513.
- Zeitlin, S. G., Chapados, B. R., Baker, N. M., Tai, C., Slupphaug, G., & Wang, J. Y. J. (2011). Uracil DNA N-Glycosylase Promotes Assembly of Human Centromere Protein A. *PLoS One*, 6(3), e17151.
- Zeitlin, S. G., Patel, S., Kavli, B., & Slupphaug, G. (2005). *Xenopus* CENP-A assembly into chromatin requires base excision repair proteins. *DNA Repair*, 4(7), 760–772.
- Zharkov, D. O., Mechetin, G. V., & Nevinsky, G. A. (2010). Uracil-DNA glycosylase: Structural, thermodynamic and kinetic aspects of lesion search and recognition. *Mutation Research*, 685(1-2), 11–20.
- Zhou, J., Blue, E. K., Hu, G., & Herring, B. P. (2008). Thymine DNA glycosylase represses myocardin-induced smooth muscle cell differentiation by competing with serum response factor for myocardin binding. *J Biol Chem*, 283(51), 35383–35392.
- Zhu, B., Zheng, Y., Hess, D., Anglikar, H., Schwarz, S., Siegmann, M., et al. (2000). 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. *Proc Natl Acad Sci USA*, 97(10), 5135–5139.
- Zhu, J., & Schiestl, R. H. (2004). Human topoisomerase I mediates illegitimate recombination leading to DNA insertion into the ribosomal DNA locus in *Saccharomyces cerevisiae*. *Mol Genet and Genomics : MGG*, 271(3), 347–358.

## 7 Appendix

- I. Reversible Top1 Cleavage Complexes are Stabilized Strand-Specifically at the Ribosomal Replication Fork Barrier and Contribute to Ribosomal DNA Stability
- II. Uracil Repair Causes DNA Glycosylase-Dependent Genome Instability
- III. Estrogen Receptor  $\beta$  Regulates Epigenetic Patterns at Specific Genomic Loci Through Interaction with Thymine DNA Glycosylase

# Reversible Top1 cleavage complexes are stabilized strand-specifically at the ribosomal replication fork barrier and contribute to ribosomal DNA stability

Claudia Krawczyk<sup>1</sup>, Vincent Dion<sup>2</sup>, Primo Schär<sup>1,\*</sup> and Olivier Fritsch<sup>1,\*</sup>

<sup>1</sup>Department of Biomedicine, University of Basel, 4058 Basel, Switzerland and <sup>2</sup>Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland

Received July 17, 2013; Revised January 14, 2014; Accepted January 31, 2014

## ABSTRACT

Various topological constraints at the ribosomal DNA (rDNA) locus impose an extra challenge for transcription and DNA replication, generating constant torsional DNA stress. The topoisomerase Top1 is known to release such torsion by single-strand nicking and re-ligation in a process involving transient covalent Top1 cleavage complexes (Top1cc) with the nicked DNA. Here we show that Top1ccs, despite their usually transient nature, are specifically targeted to and stabilized at the ribosomal replication fork barrier (rRFB) of budding yeast, establishing a link with previously reported Top1 controlled nicks. Using ectopically engineered rRFBs, we establish that the rRFB sequence itself is sufficient for induction of DNA strand-specific and replication-independent Top1ccs. These Top1ccs accumulate only in the presence of Fob1 and Tof2, they are reversible as they are not subject to repair by Tdp1- or Mus81-dependent processes, and their presence correlates with Top1 provided rDNA stability. Notably, the targeted formation of these Top1ccs accounts for the previously reported broken replication forks at the rRFB. These findings implicate a novel and physiologically regulated mode of Top1 action, suggesting a mechanism by which Top1 is recruited to the rRFB and stabilized in a reversible Top1cc configuration to preserve the integrity of the rDNA.

## INTRODUCTION

The DNA of a cell is constantly participating in molecular transactions associated with gene transcription, DNA replication and repair. During S-phase, all these processes

operate in parallel; RNA polymerases share their template with DNA polymerases moving along chromosomes and DNA repair proteins fixing DNA damage. Not only does this generate opportunities for collisions between the transcription and replication machineries, but also local DNA topological stress that requires constant release by specialized DNA topoisomerases. Topoisomerase malfunction therefore leads to an accumulation of aberrant DNA structures and, thereby, threatens genome integrity.

DNA replication encounters various obstacles such as DNA lesions or DNA-bound proteins that may impede or block replication fork (RF) progression. At programmed replication fork barriers (RFB), replication-blocking proteins bind to specific DNA sequences. In *Escherichia coli*, binding of the Tus protein to a terminator DNA sequence ensures replication termination in a defined region (1). Apart from replication termination, programmed RFBs may also function to prevent collisions between replication and transcription machineries, a potential cause of genome instability. The eukaryotic ribosomal DNA (rDNA) is particularly interesting in this respect, given that rRNA genes are continuously transcribed at high levels, irrespective whether the rDNA happens to be simultaneously replicated. The rDNA generally consists of numerous repeat units organized in tandem arrays within the nucleolus. Each rDNA unit harbours an RFB at the 3'-end of a highly transcribed rRNA gene, a feature that is highly conserved among eukaryotes [reviewed in (2)].

In budding yeast, the rDNA consists of a single array of 150–200 identical repeats on chromosome (Chr) XII, each comprising the transcriptional units 35S and 5S, an origin of replication (ARS) and a unidirectional ribosomal RFB (rRFB). A core rRFB sequence of ~100 bp is sufficient to recruit and bind sequence-specifically Fob1, a protein essential for the RF-blocking activity of the rRFB (3,4). The rRFB consists of one major (RFB1) and two minor (RFB2, RFB3) blocking sites that arrest altogether ≥90%

\*To whom correspondence should be addressed. Tel: +41 616 953 060; Fax: +41 612 673 566; Email: olivier.fritsch@unibas.ch  
Correspondence may also be addressed to Primo Schär. Tel: +41 61267 0767; Fax: +41 61267 3566; Email: primo.schaer@unibas.ch  
Present address:

Vincent Dion, Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, 1015 Lausanne, Switzerland.

of encountering RFs (5–7) without eliciting a checkpoint response (3) and force them to progress codirectionally with transcription through the highly transcribed *35S* gene (8). Due to its repetitive structure, the rDNA is prone to non-conservative recombination that can generate gains or losses of repeat units (9). While such events are crucial to restore the wild-type repeat number after accidental rDNA expansion or contraction, recombination between homologous rDNA repeats needs to be tightly regulated to prevent rDNA instability. Fob1 plays a major but ambivalent role in this context, having pro- and anti-recombinogenic properties (10). Consistently, Fob1-dependent DNA double-strand breaks (DSBs) were observed at the rRFB (9,11–13), but their molecular nature has remained elusive, and we showed previously that these DSBs are not substrate for canonical DSB repair (14).

A number of studies suggest a complex structural organisation of the rDNA, including the anchoring of the repeats to the nuclear membrane (15–17). This anchoring, mediated through the rRFB, involves Fob1 association with Tof2 and the cohibin complex, which in turn interacts with proteins of the inner nuclear membrane (15–17). Disruption of this association by deleting cohibin or one of the inner nuclear membrane proteins Heh1 or Nur1 destabilizes the rDNA repeat structure, suggesting that perinuclear attachment is crucial for the integrity of the locus (15,16). However, such anchorage is likely to impose mobility constraints to the rDNA, thereby increasing the topological stress created by replication and transcription and, hence, creating a need for continuous relaxation. Accordingly, DNA topoisomerases are specifically required for rDNA maintenance and functionality (18). In budding yeast, the conserved type IB DNA topoisomerase Top1 has been associated with the appearance of DNA nicks near the RFB (19), plays an important role in rRNA-gene transcription (18,20,21), PolII silencing (22–24) and, interestingly, in suppressing mitotic recombination in the rDNA (25–27). However, the mode of action underlying the specialized function of Top1 in the rDNA has not been addressed.

It is generally accepted that Top1 senses DNA-topological stress by direct recognition of the torque in the DNA (28). The protein then incises one DNA strand through a reversible transesterification involving a 3'-phosphodiester-tyrosyl bond, thereby engaging in a transient Top1 cleavage complex (Top1cc) intermediate with the 3'-end of the nick. Following relaxation of the DNA helix by rotation of the unbound end, Top1 reseals the nick through the reverse transesterification reaction (29). Top1ccs can be stabilized or even converted to irreversible intermediates under the influence of topoisomerase inhibitors like camptothecin (CPT), or in the presence of DNA lesions preventing the re-ligation of the single-stranded ends (29). Exactly how the cleavage and re-ligation reactions of topoisomerases are coordinated under physiological conditions is poorly understood, but the notion is that cleavage complexes are very short-lived as their resolution is triggered immediately following relaxation of the DNA helix. Here we sought to characterize the function of Top1 at the rRFB and found that Top1ccs accumulate at the rRFB, independently of its RF-stalling activity. A 450-

bp rRFB sequence is sufficient to trigger Top1cc formation independently of the rDNA context and the enrichment of Top1ccs at such sites, mediated by Fob1 and Tof2, reflects a stabilization of reversible cleavage complexes. The formation of these targeted Top1cc complexes explains the previously reported occurrence of DSBs at the rRFB in wild-type cells and implicates a physiologically regulated mode of Top1 action.

## MATERIALS AND METHODS

### Yeast strains and plasmids

Yeast strains are listed in Supplementary Table S1. All strains are isogenic derivatives of the closely related FF18733, FF18734 and FF18984 congenic series in an A364 background. Strains carrying eRFBs on Chr IV were constructed by PCR-mediated gene targeting using plasmids harbouring a 450-bp rRFB sequence insert. Gene deletions and tagging were achieved by standard methods. Strains with a LacO array tagged eRFB locus were constructed as described (30). For strains usage and plasmids construction see Supplementary Materials.

### Chromatin immunoprecipitation with or without crosslink

For chromatin immunoprecipitation (ChIP) with crosslink, whole-cell extracts were prepared from formaldehyde-fixed cell cultures as described (31,32) with modifications (Supplementary Materials). For Top1cc immunoprecipitation (IP), non-crosslinked samples were processed as described previously (33), with some modifications. Shortly, 100 ml of an exponentially growing culture was washed once with cold 20 mM Tris-HCl pH 8.0, twice with cold FA buffer (50 mM Hepes-KOH, 150 mM NaCl, 1 mM EDTA, 1% Triton-X, 0.1% Na-deoxycholate, 0.1% SDS, 1 mM PMSF) and the cell pellet was frozen at  $-80^{\circ}\text{C}$ . Cold FA (1 ml) was added to the pellet prior to cell disruption in a Fastprep-24 (MP) bead-beater (twice 40 sec at 6.5 m/s) with 0.5 mm Zirconia/silica beads. Beads were discarded and 1 ml of FA buffer was added before centrifugation for 20 min at 17 000 g,  $4^{\circ}\text{C}$ . The pellet was re-suspended in 800  $\mu\text{l}$  FA buffer and kept rolling (30 min,  $4^{\circ}\text{C}$ ) before sonication to an average size of 400 bp (Bioruptor, Diagenode). After adding 800  $\mu\text{l}$  of FA buffer, samples were centrifuged (30 min, 10 000 g,  $4^{\circ}\text{C}$ ). The supernatant was stored at  $-80^{\circ}\text{C}$  and an aliquot saved as Input. IP was performed with 500  $\mu\text{l}$  of supernatant (2 h,  $4^{\circ}\text{C}$ ) with Dynabeads (Dyna, Life Technologies) precoated with mouse anti-Myc (9E10) antibody. Beads were then washed twice for 10 s and once for 10 min with 500  $\mu\text{l}$  of modified FA buffer with 500 mM NaCl (shaking, 1400 rpm). After successive washes (shaking, 1400 rpm) with 500  $\mu\text{l}$  of 10 mM Tris-HCl pH 8.0, 0.25 M LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% Na-deoxycholate and 500  $\mu\text{l}$  of TE pH 8.0, DNA was eluted from the beads (20 min,  $65^{\circ}\text{C}$ ). IP and Input DNA were treated with 1 mg/mL Pronase (Roche) in 25 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5% NP-40, 0.5% Na-deoxycholate for 1 h at  $37^{\circ}\text{C}$ . DNA was then recovered by phenol-chloroform and isopropanol precipitation.

DNA from ChIP of crosslinking and non-crosslinking conditions were analyzed by real-time PCR (qPCR) with SYBR-Green (QuantiTect, Qiagen) using a Rotor-Gene RG3000 system (primer sequences are available upon request) and values from IP samples were related to that of their relative Input. Samples from one ChIP experiment were run two to three times in duplicates. Mean duplicate values within each run were used for calculation.

### Live-cell microscopy and image analysis

Live-cell-imaging experiments were performed using cultures of yeast cells in the appropriate SC drop out to select for the presence of the CFP-Nop1 (pFN5 and pFN6) or GFP-Nup49 (pUN100-GFP-Nup49) expressing vectors as described (34–36) Two independent cultures were imaged on different days for each genotype. A single user analyzed the images without knowledge of the genotypes. Deconvolution was achieved using the Huygens Remote Manager (37) (see Supplementary Materials for imaging and analysis details).

### Cell growth and CPT treatments

Cell-cycle arrest at the end of G1 was induced in liquid YPD cultures ( $7.5 \times 10^6$  cells/ml) with  $2 \mu\text{g/ml}$   $\alpha$ -factor (GenScript) for 2 h at  $30^\circ\text{C}$ . Cells were released from G1-arrest by the addition of  $50 \mu\text{g/ml}$  Pronase. G1-arrest and release were validated by FACS.

### 1D- and 2D-gel electrophoresis and Southern blots

Isolation of DNA in agarose plugs was done as described (11), with some modifications (see Supplementary Materials). Gels and alkaline Southern blotting were done as described (13,14). 1D-gel conditions: 24 h, 70 V, 1% agarose without EtBr. PCR-amplified probes were radioactively labelled with ( $\alpha$ - $^{32}\text{P}$ )-dCTP (6000 Ci/mmol, PerkinElmer). rDNA copy number was estimated from BamHI-digested genomic DNA of late-exponential cultures isolated in agarose plugs as recommended in the Biorad's CHEF-DR manual. PFGE was performed as described (14), with the exception that the length of the run was 48 h.

### Quantification of ERC formation

Genomic DNA was isolated from late logarithmic-phase cultures in YPD media using Qiagen genomic tips. Undigested DNA was loaded on 0.8% agarose gels and run in  $1 \times$  TAE buffer for 17 h at 65 V without EtBr. Southern blots with rDNA probe1 and quantitation were performed as for 2D-gels. The prominent signal corresponding to the bulk of rDNA was used for normalization across genotypes.

### Quantitation of DSBs and statistical methods

Quantitation of 1D- and 2D-gels was performed as described (14). See Supplementary Materials for RI definition and further details. Statistical analysis was performed with the Prism software. When not else mentioned, unpaired 2-tailed T-tests were performed for comparisons. For Figure 4A and C paired 2-tailed T-tests

were performed. For Figure 5C, one-sample 2-tailed T-tests against a theoretical mean were applied for mutants versus wild-type comparisons.

## RESULTS

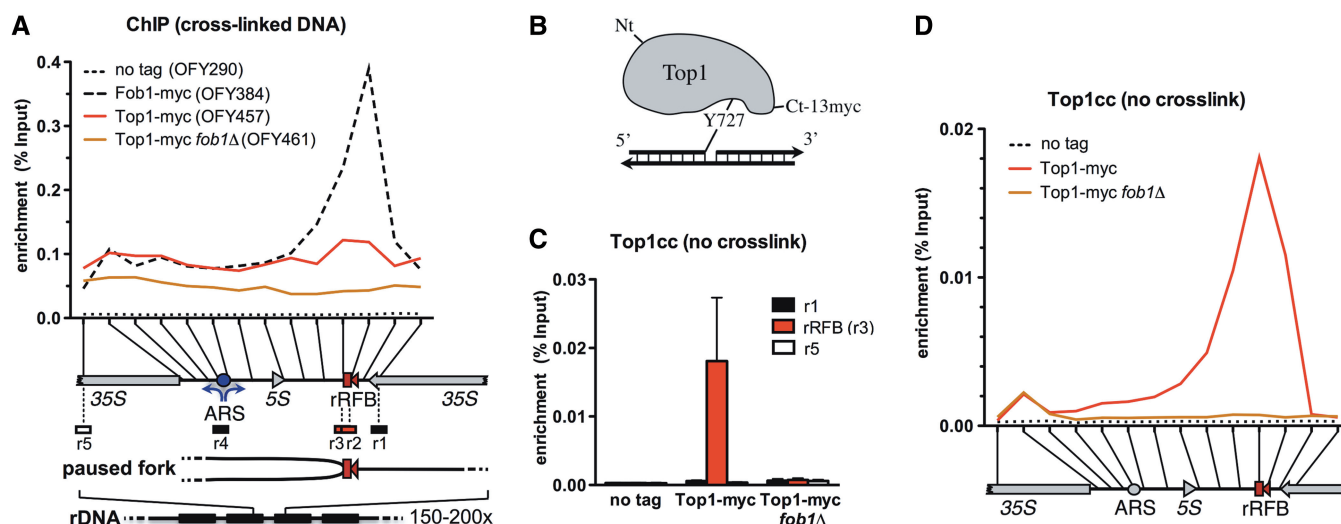
### Fob1-dependent Top1ccs are enriched at the rRFB

Based on the observation that Top1 controls the appearance of DNA nicks near the rRFB (19), we investigated more specifically the interaction between Top1 and the rRFB in budding yeast. We first compared Top1 and Fob1 occupancies along the rDNA unit by ChIP following crosslinking in asynchronous cultures. As observed previously (17,38), Fob1 was highly enriched at the rRFB with a mild secondary peak in the *35S* promoter region (Figure 1A). By contrast, Top1 enrichment spread all over the rDNA unit showing only a slight increase at the rRFB (Figure 1A). In a *fob1* $\Delta$  background, Top1 association with the rDNA was generally reduced, most strongly ( $\sim 3$ -fold) at the rRFB (Figure 1A). Thus, cross-linked ChIP supported an association of Top1 with the rDNA but did not reveal any preferential sites of enrichment.

We then used ChIP under non-crosslinking conditions to localize specifically Top1ccs, i.e. Top1 molecules that are actively engaged in DNA nicking and covalently bound to the 3'-end of the DNA via their Tyr727 residue (Figure 1B) (33). With this method, we detected a strong enrichment of Top1cc at the rRFB relative to sites in the *35S* (Figure 1C and D) and a mild enrichment in the *35S* promoter region, similar to crosslinked Fob1 ChIP (Figure 1D). In the absence of Fob1, Top1cc association with the rRFB was completely lost, whereas the small enrichment in the *35S* promoter region remained unaffected (Figure 1D). From these results, we conclude that Top1cc formation occurs with different specificity and kinetics along the rDNA unit with the rRFB being a hotspot. Importantly, the different enrichment patterns of Top1cc (without crosslinking) and Top1 (with crosslinking) suggest that Top1 associates with chromatin throughout the rDNA but is preferentially engaged in nicking at the rRFB. This may reflect an increased enzymatic turnover and/or a stabilization of the Top1cc intermediate specifically at the rRFB.

### Top1cc is detected at ectopic rRFBs outside the rDNA context

To gain further insight into the nature and context-dependency of Top1cc at the rRFB, we established a novel, strategically designed ectopic rRFB (eRFB) model. We introduced a 450-bp sequence containing the rRFB region near an early firing origin of replication (ARS453) on Chr IV (Figure 2A). As this region lacks detectable replication-origin firing within 50 kb on the distal side of the resulting eRFB (eRFB1) (39), Fob1 binding to eRFB1 should pause RFs originating from ARS453 before they encounter a converging fork. The introduced eRFB sequence was indeed sufficient to recruit Fob1 (Supplementary Figure S1) and assemble a functional RF-pausing site (Figure 2B), as observed previously at a different locus (3). To distinguish Fob1-



**Figure 1.** Top1ccs accumulate at the rRFB. (A) Mapping of Top1 in the rDNA by standard ChIP. Exponentially growing cells were crosslinked with formaldehyde before ChIP. Shown is the mean enrichment ( $N = 3$ ). The position of qPCR primer sets is indicated on one rDNA unit, together with those used in (C) and (D) with the unidirectional rRFB blocking RFs as depicted. 35S and 5S, rRNA transcription units. (B) Schematic representation of myc-tagged Top1cc. While nicking one strand, Tyr727 engages transiently but covalently with the 3'-phosphodiester at the nick. (C) Top1cc is enriched at the rRFB. Top1cc ChIP was conducted from non-crosslinked exponentially growing cells of yeast strains as in (A). The mean ( $N = 4$ ) with SEM is shown. Primer-sets location, see (A). (D) Mapping of Top1cc along the rDNA unit by non-crosslinked ChIP as in (C). The mean ( $N = 4$ ) is shown. qPCR primer set positions are indicated. Strains as in (A).

mediated from stalling-triggered effects, we introduced a second eRFB (eRFB2) 6-kb apart from the first, in either the same (direct repeat) or the opposite (indirect repeat) orientation relative to eRFB1 (Figure 2A). The eRFB2 in the same orientation as eRFB1 will pause progression of forks emanating from ARS453 (and leaking from eRFB1), whereas the eRFB2 in the opposite orientation will be permissive and, thus, not accumulate pausing forks. Although Fob1 was enriched at the eRFB2 in either orientation (Supplementary Figure S1), paused RFs only accumulated in the direct-repeat configuration (Figure 2C), confirming that no or only very few RFs were approaching from the ARS453-distal side (39). The eRFB2 constructs thus allow the assessment of RF stalling-independent functions of Fob1.

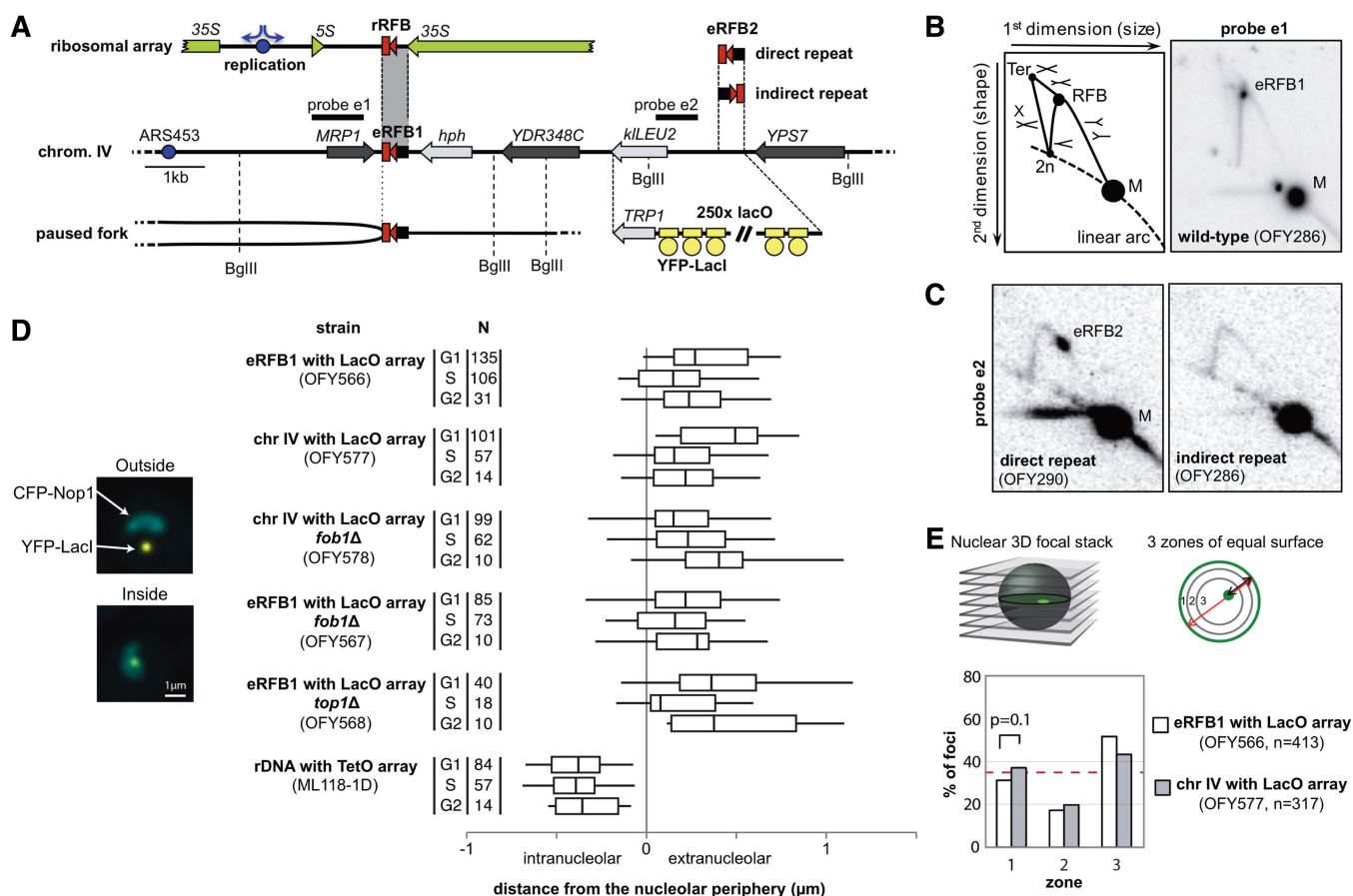
To confirm that the eRFB is outside of the rDNA context and not anchored to the nuclear membrane despite the insertion of rDNA sequence—this was not addressed for previously published eRFB strains (3)—we assessed the position of the eRFB in the nucleus by fluorescence microscopy. We inserted  $\sim 250$  lacO repeats near the eRFB1, allowing for the measurement of its relative distance to the nucleolar periphery upon ectopic expression of YFP-LacI, visualizing the location of the lacO array and of CFP-Nop1 fusions marking the nucleolus (Figure 2A). In the wild-type, but also in *fob1Δ* and *top1Δ* backgrounds, the eRFB1 sub-nuclear localization was clearly extranucleolar, irrespective of the cell-cycle stage and deletion of eRFB1 confirmed that its ectopic insertion did not alter the authentic localisation of this Chr IV locus relative to the nucleolus (Figure 2D). In contrast, a wild-type strain carrying a TetO array in the rDNA and an integrated TetI-mRFP1, highlighting an internal rDNA locus (40) produced fluorescence signals

clearly located in the nucleolus. We then assessed the position of the eRFB1 locus relative to the nuclear periphery as visualized by a GFP-Nup49 fusion marking the nuclear periphery. The eRFB1 locus was not associated with the periphery irrespective of the presence or absence of the barrier (Figure 2E), indicating that eRFB1 is not sufficient to mediate perinuclear anchoring. These results show that the eRFB insertion preserves the authentic location of the locus with respect to the nucleolus and the nuclear periphery making it a suitable model to analyze rRFB functions outside the complex organisation of the rDNA. Using this eRFB system, we detected Fob1-dependent peaks of Top1cc enrichment at the eRFB1 and eRFB2 (Figure 3B and C), similar to the situation at the rRFB in the rDNA (Figure 1D). Given the spatial separation of the eRFB from the nucleolus, these results suggest that Top1cc at the rRFB in wild-type cells is formed independently of the rDNA context and that the 450-bp rRFB sequence is sufficient to trigger Top1cc formation outside the rDNA.

#### Top1cc associates strand-specifically with the eRFB independently of RF-stalling

Investigating the link between RF-pausing and Top1cc formation at the rRFB, we found the levels of Top1cc enrichment at the rRFB to be similar in G1- and S-phase of the cell cycle (Figure 3B). This was true both at ectopic and ribosomal sites, establishing that Top1cc formation at the rRFB is not restricted to S-phase. To exclude that RF pausing induces nicks that persist throughout the cell cycle, we mapped Top1cc along the eRFB1/eRFB2 locus in the direct- and indirect-repeat configuration (Figure 3C). Top1cc was strongly enriched at both eRFBs, independent of their orientation,

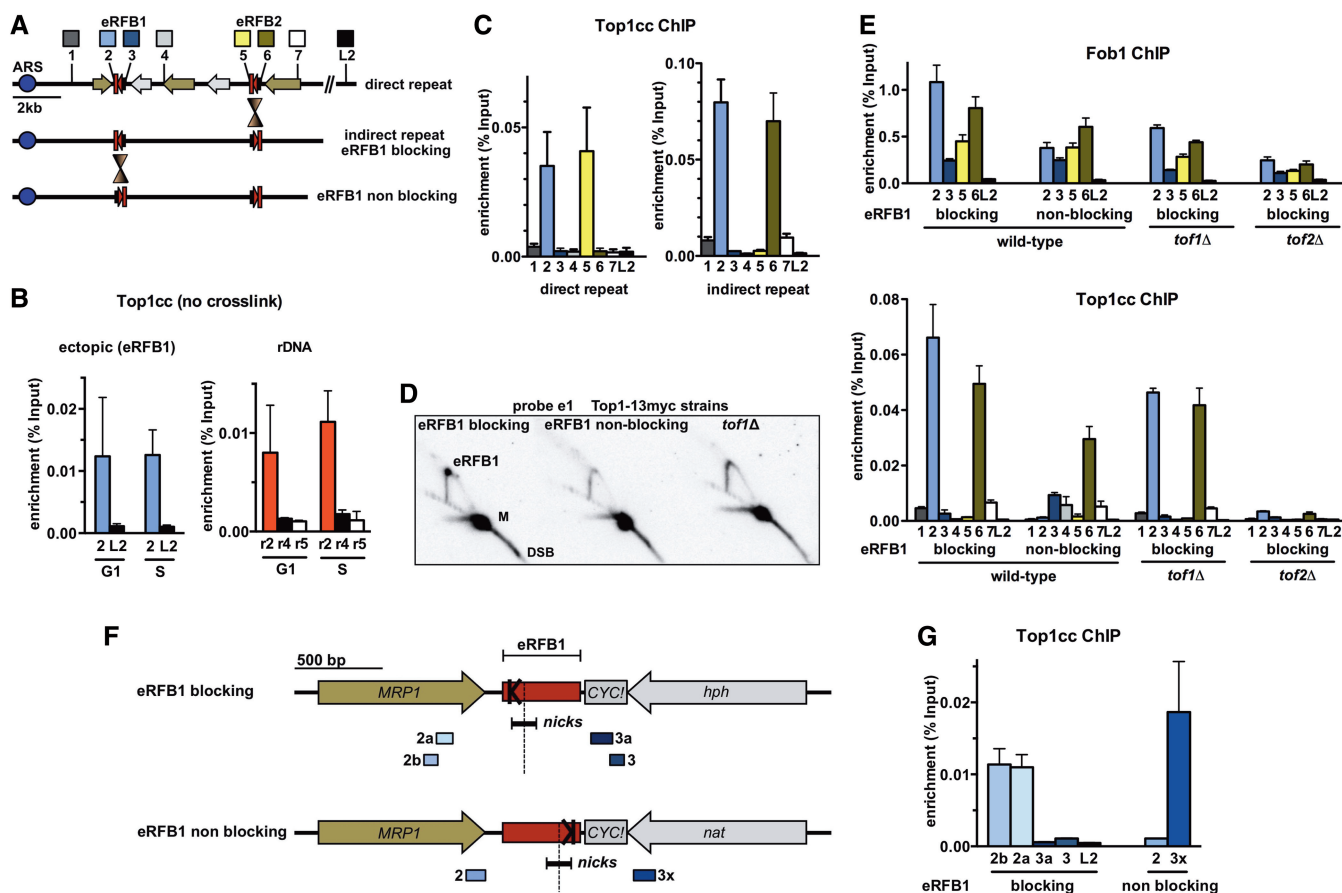




**Figure 2.** A 450-bp rRFB sequence is sufficient to trigger a Top1cc signal outside the rDNA. (A) The eRFB model on Chr IV. One rDNA unit is shown as reference. The 450-bp rRFB sequence (grey) used as eRFB was inserted at two ectopic sites (eRFB1 and eRFB2) near the early firing origin ARS453. For live-cell microscopy, an array of 250 LacO repeats was inserted instead of eRFB2. (B) The eRFB1 is proficient for RF stalling. Scheme and Southern blot of DNA species obtained after 2D-gel electrophoresis of BglII-digested genomic DNA from exponential cultures of a wild-type strain carrying the eRFB1. Probe e1, see (A) M, monomer; RFB, rRFB-stalled RFs; Ter, converging RFs; 2n, molecules of twice the monomer size. Note that faint spots on the left of the M and eRFB1 signals correspond to incomplete digestion of the BglII site between *PAL1* and *hph*. (C) Fork-pausing activity at the eRFB2. Southern blots of 2D-gels as in (B) except for probe e2 (A). (D) The eRFB1 locus is not recruited to the nucleolus. eRFB1 strains harbour a LacO array with an integrated YFP-LacI gene (A) and a CFP-Nop1 expression plasmid for nucleolar staining. To assess the localization of the ectopic locus without a barrier, eRFB1 was exchanged for a *natNT2* cassette. (Left) Representative images for eRFB1 localization relative to the nucleolus. Shown are examples of strain OFY566 with extra- and intranucleolar localization, respectively. (Right) Quantification of nucleolar localization. The distance distribution between the tagged genomic locus and the nucleolar periphery, as obtained from the CFP-Nop1 signal, is plotted as a box plot (boxes, 25th–75th percentiles; whiskers, 5th and 95th percentiles; vertical bars, medians). (E) A single eRFB is not sufficient to anchor a locus to the nuclear periphery. (Top) Optical slices were taken in cells expressing GFP-NUP49 fusion that demarcates the nuclear periphery. Three zones of equal surface were defined for distance measurements (35). A locus that is randomly distributed throughout the nucleus is expected to have an equal chance of being in any one zone. (Bottom) Location of the eRFB1 locus on Chr IV in the presence or absence of the barrier. *P*-value is for a  $\chi^2$  test with *df* = 1 for zone 1 between the two strains.

indicating that Top1cc accumulates at the eRFB2 even in the absence of RF-pausing at this site. In line with this, deletion of *TOF1*, encoding a factor required for RF stalling at an RFB (Figure 3D) (3), did not significantly affect Top1cc enrichment at eRFB1 and eRFB2 (Figure 3E). Further, to exclude genomic context effects, we also inverted the eRFB1 and analyzed resulting strains now carrying two non-blocking eRFBs (Figure 3D and E). Top1 was indeed associated with the non-blocking eRFB1, corroborating that Top1cc at the eRFB forms independently of RF stalling. Since Top1cc accumulation requires Fob1 but not RF pausing (Figures 1D and 3C), these findings strongly suggest that Fob1 recruits and/or

stabilizes Top1cc at the eRFB independently of DNA replication. Importantly, Top1cc enrichment was asymmetric with respect to the 450-bp eRFB insertion. While Top1cc was enriched next to the side that is capable of RF stalling, it was absent from the opposite side of the insertion (Figure 3C and E; compare primer sets 2/3 for eRFB1 and 5/6 for eRFB2). This is best explained by a preferential breakage of Top1 nicked sites during DNA sonication and a subsequent enrichment of DNA fragments with the 3'-attached Top1cc by Top1-ChIP. However, the data do not exclude the possibility that the asymmetry of Top1cc enrichment over the 450-bp eRFB sequence (Figure 3F) is caused by the unequal distance of the PCR probes to the



**Figure 3.** Top1cc enrichment at an eRFB is strand-specific and independent of RF stalling. (A) Orientation of eRFBs and localization of qPCR primer sets in eRFB strains. The control L2 set is located in a late replicated region of Chr IV. (B) Top1cc is enriched at the rRFB and eRFB1 in both G1- and S-phases. Wild-type cells expressing Top1-myc (OFY457) were arrested in G1 and released into S-phase for 30 min prior to non-crosslinked ChIP. The mean enrichment ( $N = 2$ ) is shown with SD. Primer-sets location, see (A) and Figure 1A. (C) Asymmetry of the Top1cc signal at eRFB1 and eRFB2. Top1cc ChIP with primer sets as indicated in (A). The mean with SEM is shown ( $N = 3$ ). (D) RF-stalling activity at the eRFB1 of *tof1Δ* and eRFB1 blocking and non-blocking strains. Southern blot of a 2D-gel from exponentially growing cells as in Figure 2B. (E) Requirements for Top1cc accumulation at eRFBs. Mean enrichments from ChIP experiments for Fob1-myc (crosslink,  $N = 4$ ) and Top1cc (no crosslink,  $N = 3$ ) are shown. Primer sets for qPCR, see (A). (F and G) The Top1cc signal at the eRFB1 is strand-specific. (F) Scheme of the eRFB1 region in the blocking versus non-blocking orientation (as referred to ARS453). Black symbols |< and >| indicate the main RF pausing site (RFB1) with blocking RFs coming from the left or right, respectively. For each eRFB1 orientation, qPCR probes at equidistant sites from the centre of Top1 induced nicks are indicated (black line). (G) Mean enrichment values ( $N = 3$ , SEM) for non-crosslinked Top1cc ChIPs with strains and primers as depicted in (F).

area where Top1 nicks are located. We addressed this by qPCR of Top1cc ChIP experiments using pairs of primer-sets equidistant to the Top1 nicks (Figure 3F). The detection of a strong Top1cc signal on the blocking side and no signal on the non-blocking side in each of the eRFB1 orientation, allowed us to exclude this hypothesis (Figure 3G). This given, the asymmetric enrichment is strongly indicative for Top1 being positioned to the rRFB to specifically incise the DNA strand corresponding to the parental lagging strand ahead of a potential rRFB-paused RF, consistent with results from the mapping of Top1 controlled nicks (19,41).

### Top1 accumulation at the rRFB reflects stabilization of a transient cleavage complex

The enrichment of Top1ccs at both ribosomal and ectopic RFBs might reflect an accumulation of irreversibly bound Top1ccs. Processing of irreversible Top1ccs involves both

proteolytic degradation of Top1 and cleavage of the DNA 3'-end that is covalently attached to Tyr727. The enzyme dedicated to Top1cc processing is Tdp1, but other factors provide complementary activities and can even substitute for Tdp1, most prominently the Rad1 and Mus81 nucleases (42,43). Given the poor growth and pleiotropic phenotype of the triple *tdp1Δ rad1Δ mus81Δ* mutant, we resorted to monitor the potential accumulation of irreversible Top1cc in *tdp1Δ rad1Δ* and *tdp1Δ mus81Δ* double mutants, where defects in Top1cc repair are detectable (43). We generated these strains, confirmed their sensitivity to CPT (Supplementary Figure S2) and assessed Top1cc enrichment at eRFB1 by non-crosslinked ChIP (Figure 4A). We found no significant change in Top1cc levels in these double mutant strains compared to the wild type. To rule out that Top1cc enrichment at the eRFB is saturated in the wild type, we added CPT to permeabilized cells prior to ChIP; permeabilization allowing for a rapid

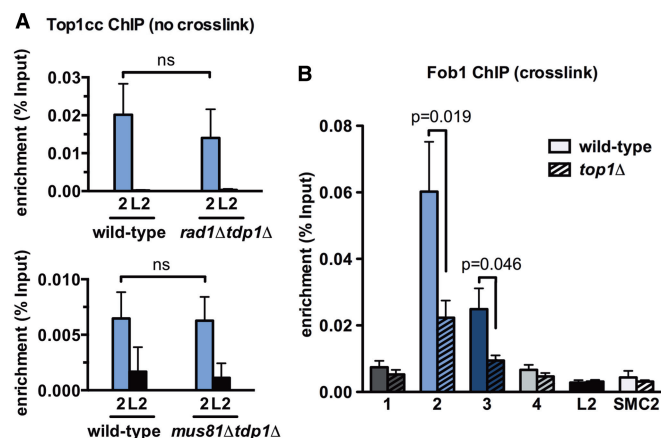
and efficient CPT uptake. Under these conditions, Top1cc levels were clearly increased (Supplementary Figure S3), suggesting that potential effects of Top1cc repair defects at the eRFB could be detected. The unchanged level of Top1cc in *tdp1Δ rad1Δ* and *tdp1Δ mus81Δ* strains thus indicates that Top1ccs formed at the eRFB are not subject to repair but remain reversible with the nicked configuration stabilized but prone for re-ligation. We therefore conclude that stabilized Top1ccs at the rRFB are physiological intermediates.

### Tof2 mediates Top1cc accumulation at eRFBs and stabilizes Fob1

The Fob1 dependency of Top1cc formation at the rRFB raises the possibility that Fob1 positioning also requires Top1, and that the two proteins engage in a Top1cc-stabilizing functional interaction. To assess the existence of such a complex, we performed Fob1 ChIP from *TOP1* wild-type and deleted strains. Because of the repetitive structure and the nuclear context of the endogenous rRFB, we focused on the eRFB. Fob1 enrichment at the eRFB was indeed reduced in *top1Δ* cells as compared to wild-type cells (Figure 4B and Supplementary Figure S4), suggesting that Top1 stabilizes the Fob1-eRFB interaction and, consequently, that the two proteins are part of a functional complex. Fob1 seems to be the central organizer of this complex as it is a prerequisite for Top1cc association with the eRFB but remains detectable in the absence of Top1. Searching for additional components of the complex, we considered Tof2. Tof2 physically interacts with both Fob1 and Top1 (17,44) and is therefore implicated as a mediator of an rRFB-associated function of Top1. We assessed the influence of *TOF2* deletion on Top1cc and Fob1 association with the eRFBs. Fob1 enrichment at eRFB1 and eRFB2 in *tof2Δ* cells was reduced to less than half of that in wild-type cells, whereas Top1cc enrichment was almost completely lost (~5% of the wild-type level, Figure 3E). Thus both Fob1 and Top1cc enrichment at the eRFBs is modulated by Tof2, suggesting that the three proteins are part of an rRFB binding protein complex that mediates targeted and strand-specific Top1 activity.

### DSBs detected at ectopic and endogenous rRFBs depend on Top1 activity

RFs paused at the rRFB have been proposed to break under physiological conditions (9,11–13), based on the detection of DNA-replication associated DSB signals in 1D and 2D-gel experiments. As these apparent DSBs did not appear to be substrate for the canonical DSB-repair systems (14), we wondered whether they might originate from Top1-induced nicks at the rRFB (Figure 5A), as the DNA duplex between the nick and the arrested RF might melt during DNA isolation (orange fragment on Figure 5A). We performed 2D-gels using DNA isolated from asynchronous cell cultures (Figure 5B), quantified the DSB signal and normalized it to replication intermediates (RIs) (14). In *top1Δ* cells, the level of broken RFs at the rRFB decreased to 35% of that in wild-type cells ( $P = 0.032$ ; Figure 5C). We also detected a DSB signal

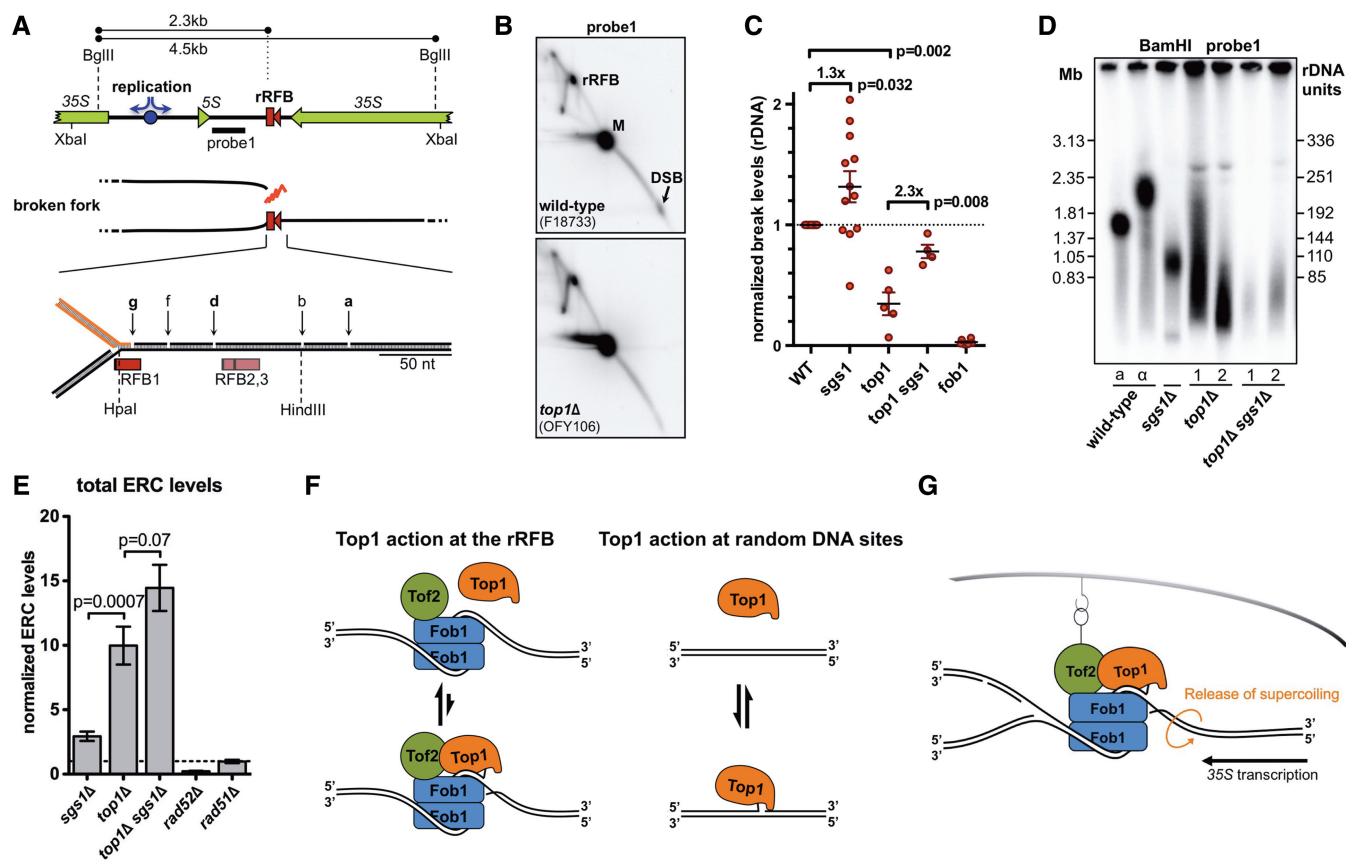


**Figure 4.** Stabilization of Top1cc at the eRFB1. (A) Contribution of irreversible-Top1cc repair pathways to Top1cc enrichment at the eRFB1. Top1cc ChIP from exponentially growing cells. Matching wild-type/mutant pairs were isolated by crosses. Location of qPCR primer sets, see Figure 3A. (B) Fob1 enrichment at the eRFB1 is partially dependent on Top1. Mean values from five cross-linked ChIP experiments are shown for one representative wild-type/*top1Δ* strains pair. SMC2, qPCR primer set in the *SMC2* gene. Location of other primer sets, see Figure 3A.

at the eRFB1 in wild-type cells (Figure 3A and Supplementary Figure S5A); as in the rDNA, the DSB signal was absent in *fob1Δ* cells (Supplementary Figure S5A) and reduced in *top1Δ* cells (Supplementary Figure S5B). Thus, most of the DNA breaks at these RFBs arise in a Top1-dependent manner. To further differentiate between an enzymatic and a structural contribution of Top1, we tested the effect of CPT treatment on the DSB signal. As CPT stabilizes Top1cc, it should increase the steady-state level of Top1-dependent nicks at the rRFB (19). We exposed permeabilized spheroplasts to CPT prior to 1D-gel analysis (Supplementary Figure S6). Whereas the level of broken RFs at the rRFB increased 1.8-fold upon CPT treatment in wild-type cells, it remained unaltered in *top1Δ* cells, showing that Top1ccs and, thus, the catalytic activity of Top1 are responsible for the appearance of the DSB signal at the rRFB.

### DSBs induced at the rRFB upon *SGS1* disruption occur independently of Top1

In the absence of the RecQ helicase Sgs1, the level of broken RFs at the rRFB increases on 1D- and 2D-gels (12,14). We reasoned that, in contrast to the breaks in wild-type cells, these additional breaks result from insufficient stabilization of the rRFB-paused RF and, thus, may be of a different molecular nature. As Top1cc accounted for most of the DSB signal on 2D-gels in wild-type cells (Figure 5C), we asked whether the additional breaks observed in *sgs1Δ* cells are also Top1-dependent. We measured DSBs at the endogenous rRFB that exhibits a better dynamic range for DSB quantitation than the eRFB. Disruption of *SGS1* led to a moderate but significant increase of DNA breaks at the rRFB both in the wild-type (1.3-fold) and *top1Δ* (2.3-fold) backgrounds (Figure 5C). Thus, Top1-independent breaks accumulate



**Figure 5.** Function of Top1cc at the rRFB and link with previously reported DSBs. (A) Map of one rDNA unit and close-up on the DNA nicks (*a*, *b*, *d*, *f* and *g*) observed at the rRFB on the upper strand (13), *a* and *d* showing Top1 dependency (41). (B) Top1 contributes to the DSB signal on 2D-gels. Representative Southern blots of 2D-gels from exponential cultures. M, monomer; rRFB, rRFB-stalled RFs; DSB, broken RFs. (C) Quantification of DSB levels from experiments as in (B). Dots represent DSB levels relative to RIs (see Materials and methods section) and normalized to DSB levels from wild-type DNA on the same gel. Means  $\pm$  SEM are indicated. (D) Impact of Top1 on rDNA copy-number maintenance. Southern blot of genomic DNA digested with BamHI, which does not cut within the rDNA, analysed by PFGE. (E) ERC formation. Undigested genomic DNA from young cells was separated on agarose gels, which were subjected to Southern blot analysis using rDNA probe1 (Figure 5A). All ERC signals, see (13) for details, were quantitated from at least three experiments in duplicate, corrected for rDNA loading and normalized to wild-type. The mean  $\pm$  SEM is plotted. (F and G) A model for Top1cc formation and function at the rRFB. See text for details. (F) Top1ccs are preferentially stabilized at the rRFB as compared to genome-wide occurring Top1ccs. (G) Top1ccs stabilize the rDNA by relieving torsional stress deriving from the perinuclear anchoring of the rRFBs.

in the absence of Sgs1; probably reflecting collapsed RFs that are prone to recombination.

### Reduced rDNA stability in the absence of Top1

We showed that Top1 nicks are stabilized at the rRFB, providing a mechanism for the continuous release of torsional stress. Assuming that stabilized nicks prevent unscheduled and possibly recombinogenic breaks, rDNA recombination should increase upon *TOP1* deletion. Increased rDNA recombination in *top1Δ* cells is supported by the reduced and variable rDNA copy number (Figure 5D) as well as an increased failure of Chr XII of exponentially growing cells to enter the gel (45) in pulse-field gel electrophoresis (PFGE). rDNA recombination also gives rise to elevated excision of extra-chromosomal ribosomal circles (ERCs). Indeed, ERC levels are increased in *top1Δ* cells (Figure 5E) (27), suggesting that Top1-dependent nicks contribute to the preservation of DNA integrity. Similar to *top1Δ* cells, *sgs1Δ* cells possess an unstable rDNA copy number

(Figure 5D) (46) and an increased level of ERCs (Figure 5E) (47). Normalized Top1cc levels at the rRFB, however, remain essentially unaltered upon *SGS1* disruption (Supplementary Figure S7), whether at the eRFB1 or the authentic rRFB in the rDNA, suggesting that Top1ccs at the rRFB do not contribute to the rDNA instability of *sgs1Δ* cells. PFGE of DNA from *top1Δ sgs1Δ* cells did not reveal any further instability than that seen in either single mutant (Figure 5D). Thus, both Top1 and Sgs1 function in rDNA homeostasis. While positioned DNA incision by Top1 at the rRFB promotes rDNA stability, DSBs generated in *sgs1Δ* cells cause instability.

### DISCUSSION

Top1 releases DNA torsional stress generated by DNA replication, transcription or repair by nicking and religating DNA single strands. This involves the formation of Top1ccs, transient covalent intermediates linking Top1

with the DNA. We show that Top1ccs are strongly enriched at the yeast rRFB, a site where DNA nicks were previously correlated with Top1 activity, in the absence of topoisomerase poisons, suggesting that persistent Top1-mediated strand cleavage is targeted to the rRFB under physiological conditions. Whereas ChIP without prior chemical crosslinking yielded a strong peak of Top1cc enrichment at the rRFB, crosslinking resulted in only little specific Top1 enrichment at this site (Figure 1). Thus, although Top1 associates with the rDNA repeat along its entire length, the steady-state level of cleavage complexes is specifically increased at the rRFB. Given that these Top1ccs are not subject to topoisomerase repair by the canonical pathways (Figure 4A) and that the loss of Top1 destabilizes the rDNA structure (Figure 5D and E), the data suggest that Top1 is targeted to the rRFB where it accumulates in reversible cleavage complexes in a physiological process of rDNA maintenance. Although we cannot strictly rule out the possibility that the enrichment of Top1cc at the RFB reflects an increased turn-over of Top1 cleavage cycles, the data favour a model in which the normally transient Top1cc intermediates are stabilized at the RFB. Both scenarios support that Top1ccs, and hence the DNA nicks, are subject to physiological regulation in the context of genome maintenance.

The stabilization of Top1cc would implicate some form of regulation of the re-ligation step of the of Top1 reaction. This may provide a mechanism for controlled release of the DNA torsional stress building up in the rDNA due to transcription and replication activities and aggravated by the mobility constraints imposed by the perinuclear anchoring of the rDNA repeats (15–17). The purpose of such strategic positioning of persistent but reversible Top1 nicks at the RFB would thus be to continuously normalize DNA topology and, thereby, increase the stability of this special genomic region. Consistently, recombination in the rDNA increased upon *TOP1* deletion, as we observe an unstable and decreased rDNA length associated with increased formation of ERCs, an rDNA recombination product (Figure 5D and E) (25). This phenotype caused by the absence of Top1 could be well explained by a build-up of torsional stress in the rDNA, which is then compensated for by the random induction of unscheduled DNA breaks, by Top2 or Top3 for example. Such unscheduled breaks could occasionally trigger recombination and finally the formation of ERCs.

Top1 does not only prevent unscheduled recombination in the rDNA, it also stabilizes the Fob1 complex as implicated by a reduced Fob1 ChIP signal upon *TOP1* deletion (Figure 4B). Gains or losses in rDNA copy number are quickly normalized by unequal sister-chromatid exchange in wild type but not in *fob1Δ* yeast cells, in which also ERC formation is abolished (48,49). While *TOP1* deletion results in an unstable rDNA repeat number, *FOB1* deletion completely abolishes copy number changes (10). The different effects of *TOP1* and *FOB1* deletion on rDNA recombination are consistent with Fob1 being the main regulator of the rRFB complex. In the absence of Fob1, neither RF stalling nor perinuclear anchoring will occur. In this situation,

less torsional stress is created, lowering the need for Top1-dependent release of DNA torsion. Thus, both Top1 and Fob1 are important for rDNA stability, but they impact at different levels of a common underlying pathway.

Using ectopically integrated rRFBs (eRFB1 and eRFB2), we show that Top1cc accumulation at the rRFB requires neither the authentic nucleolar rDNA context nor the perinuclear anchoring. Thus, the 450-bp rRFB-containing sequence carries sufficient information to mediate Top1cc accumulation. The eRFBs on Chr IV reside in a genomic context with low transcriptional activity as opposed to the situation on the rRFB. The presence of Top1 nicks at eRFB1 thus suggests that Top1cc enrichment is independent of the proximity to the highly transcribed 35S. Moreover, the comparable enrichment of Top1cc at eRFB1 and eRFB2 as well as in strains with inverted orientations of the eRFBs indicates that Top1cc levels are not influenced by the local transcriptional activity (Figure 3). This is in accordance with previous work showing that the occurrence of CPT-inducible nicks at the rRFB is not affected by altering the transcription efficiency of RNA PolII (19).

Remarkably, we show that the formation of Top1ccs at the rRFB is restricted to the DNA strand defined as the one that will form the lagging strand template of a potential RF blocked at the rRFB consistent with a previously reported strand-preference of Top1-dependent nicks at the rRFB (19,13). Inversion of the eRFB1 and eRFB2 sequences resulted in an inversion of the Top1cc asymmetry pattern. Together with the S-phase independence of Top1cc formation, this strongly suggests that the eRFB sequence itself directly controls the positioning of the Top1cc (Figure 3C and 3G). The binding of Fob1 and associated proteins to the rRFB may induce a DNA conformation or topology, exposing preferential Top1 cleavage sites (50,51) to Top1.

Despite the Top2-dependence of Top1ccs at the eRFBs (Figure 3E), we show that the perinuclear anchoring of the rDNA is not maintained at the eRFBs (Figure 2E). Thus, although the rRFB-associated Top1 nicking activity seems to have evolved as a provisional mechanism to normalize topological stress in the rDNA, the nicking itself does not appear to be triggered by rDNA-specific topological conditions, but is mediated directly by the assembly of a Fob1 dependent functional protein complex at the rRFB. As we show here, the formation of the Top1cc at the rRFB requires at least the presence of Fob1 and Top2. In line with this, Top1 and Top2 were co-purified with Tap-tagged Fob1 (17) and physical interactions have been reported between Top1 and Top2 (44). The co-existence of Top1, Top2 and Fob1 in a complex suggests that these proteins collectively coordinate rDNA transcription, replication and recombination (52) to maintain rDNA stability.

Finally, we show that DNA breaks at the rRFB that were observed on 2D gels (7,11–13) relate to Top1ccs rather than DSBs. The strict S-phase dependency of the DSBs on agarose gels (13) but not of the Top1-dependent nicks (Figure 3B) (19) strongly suggest that these apparent DSBs arise during DNA isolation as a

consequence of the experimental melting of the parental duplex DNA between the branch point of a stalled RF and the nearby Top1 nick ahead of the RF. The nick mapped closest to the major RF-pausing site (RFB1) resides only 14 nt ahead of the expected RF (Figure 5A) (13,41), thus generating a relatively unstable duplex region that is likely to melt to generate a DSB fragment even under the mild conditions of DNA isolation used in our and previous experiments (Figure 5A, orange fragment). In contrast, the additional breaks observed in *sgs1*Δ cells—in which RFs pausing at the rRFB are likely to be destabilized and to break more easily—probably reflect canonical DSBs (Figure 5C) (12,14). Importantly, fission yeast *rqh1*<sup>sgs1</sup> cells, but not wild-type cells, display similar breaks at an ectopic RFB on 1D-gels (53).

Having shown that Top1 is stabilized at the rRFB, we propose a model for how strand-specific DNA nicking by Top1 is achieved. Recruitment to the rRFB complex by interactions with Fob1 and Tof2 (17,44) cannot fully explain the DNA strand and site-specific cleavage. Instead, the DNA structure imposed by wrapping around Fob1 (7), together with a preferential nicking-site sequence and the non-palindromic nature of the Fob1-binding site, might specify the position of Top1 cleavage—bent DNA was indeed shown to be a good substrate for Topoisomerase 1 *in vitro* (54). Following strand incision, Top1cc is stabilized by an rRFB-specific DNA conformation that delays the resealing of the nick (28), contrasting the normally short-lived Top1ccs (Figure 5F). This could be achieved by temporarily displacing the unbound 5'-DNA end from the Top1 catalytic site or from a direct stabilization of the cleavage complex by Fob1, Tof2 or other proteins present at the rRFB (17,44). Importantly, Top1cc stabilization does not require the presence of a pausing RF, but replication through the rRFB will require Top1cc removal to prevent the formation of a DSB. Our data suggest that this removal does not involve excision of Top1, but instead resealing of the nick by Top1 itself. We propose that the ability of Top1cc to complete the reaction and ligate the nick is finally achieved by the disassembly of the proteinaceous complex at the rRFB allowing the passage of the RF. Relieving the torsional stress at the rRFB is especially required assuming rRFB anchoring to the nuclear membrane (Figure 5G). In conclusion, our data not only provide insight into the mechanism of Top1 function in preserving rDNA integrity, but also they might present a general concept for how topoisomerases help at RF impediments.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Serge Boiteux and Michael Lisby for yeast strains and Frank Neumann for plasmids. We are also thankful for Susan M. Gasser's input and support. We

thank the Facility for Advanced Imaging and Microscopy from the FMI.

## FUNDING

Forschungsfonds of the University of Basel (to O.F. for C.K.) Novartis Research Foundation (to V.D.); Swiss National Science Foundation (to V.D. in SMG's laboratory). Funding for open access charge: University of Basel.

*Conflict of interest statement.* None declared.

## REFERENCES

- Hill, T.M., Tecklenburg, M.L., Pelletier, A.J. and Kuempel, P.L. (1989) *tus*, the trans-acting gene required for termination of DNA replication in *Escherichia coli*, encodes a DNA-binding protein. *Proc. Natl Acad. Sci. USA.*, **86**, 1593–1597.
- Mirkin, E.V. and Mirkin, S.M. (2007) Replication fork stalling at natural impediments. *Microbiol. Mol. Biol. Rev.*, **71**, 13–35.
- Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A. and Labib, K. (2005) Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. *Gene Dev.*, **19**, 1905–1919.
- Kobayashi, T. and Horiuchi, T. (1996) A yeast gene product, Fob1 protein, required for both replication fork blocking and recombinational hotspot activities. *Genes Cells.*, **1**, 465–474.
- Sogo, J.M., Lopes, M. and Foiani, M. (2002) Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects. *Science.*, **297**, 599–602.
- Brewer, B.J., Lockshon, D. and Fangman, W.L. (1992) The arrest of replication forks in the rDNA of yeast occurs independently of transcription. *Cell.*, **71**, 267–276.
- Kobayashi, T. (2003) The replication fork barrier site forms a unique structure with Fob1p and inhibits the replication fork. *Mol. Cell Biol.*, **23**, 9178–9188.
- Brewer, B.J. and Fangman, W.L. (1988) A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell.*, **55**, 637–643.
- Kobayashi, T., Horiuchi, T., Tongaonkar, P., Vu, L. and Nomura, M. (2004) SIR2 regulates recombination between different rDNA repeats, but not recombination within individual rRNA genes in yeast. *Cell.*, **117**, 441–453.
- Kobayashi, T., Heck, D.J., Nomura, M. and Horiuchi, T. (1998) Expansion and contraction of ribosomal DNA repeats in *Saccharomyces cerevisiae*: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Gene Dev.*, **12**, 3821–3830.
- Weitao, T., Budd, M., Hoopes, L. and Campbell, J. (2003) Dna2 helicase/nuclease causes replicative fork stalling and double-strand breaks in the ribosomal DNA of *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **278**, 22513–22522.
- Weitao, T., Budd, M. and Campbell, J.L. (2003) Evidence that yeast SGS1, DNA2, SRS2, and FOB1 interact to maintain rDNA stability. *Mut. Res.*, **532**, 157–172.
- Burkhalter, M.D. and Sogo, J.M. (2004) rDNA enhancer affects replication initiation and mitotic recombination: Fob1 mediates nucleolytic processing independently of replication. *Mol. Cell.*, **15**, 409–421.
- Fritsch, O., Burkhalter, M.D., Kais, S., Sogo, J.M. and Schär, P. (2010) DNA ligase 4 stabilizes the ribosomal DNA array upon fork collapse at the replication fork barrier. *DNA Repair.*, **9**, 879–888.
- Mekhail, K., Seebacher, J., Gygi, S.P. and Moazed, D. (2008) Role for perinuclear chromosome tethering in maintenance of genome stability. *Nature.*, **456**, 667–670.
- Chan, J.N.Y., Poon, B.P.K., Salvi, J., Olsen, J.B., Emili, A. and Mekhail, K. (2011) Perinuclear cohibin complexes maintain replicative life span via roles at distinct silent chromatin domains. *Dev. Cell.*, **20**, 867–879.
- Huang, J., Brito, I.L., Villen, J., Gygi, S.P., Amon, A. and Moazed, D. (2006) Inhibition of homologous recombination by a

- cohesin-associated clamp complex recruited to the rDNA recombination enhancer. *Gene Dev.*, **20**, 2887–2901.
18. French, S.L., Sikes, M.L., Hontz, R.D., Osheim, Y.N., Lambert, T.E., Hage, E.A., Smith, M.M., Tollervey, D., Smith, J.S. and Beyer, A.L. (2011) Distinguishing the roles of Topoisomerases I and II in relief of transcription-induced torsional stress in yeast rRNA genes. *Mol. Cell Biol.*, **31**, 482–494.
  19. Vogelaer, M. and Camilloni, G. (1999) Site-specific in vivo cleavages by DNA topoisomerase I in the regulatory regions of the 35S rRNA in *Saccharomyces cerevisiae* are transcription independent. *J. Mol. Biol.*, **293**, 19–28.
  20. Brill, S.J., DiNardo, S., Voelkel-Meiman, K. and Sternglanz, R. (1987) Need for DNA topoisomerase activity as a swivel for DNA replication for transcription of ribosomal RNA. *Nature.*, **326**, 414–416.
  21. Schultz, M.C., Brill, S.J., Ju, Q., Sternglanz, R. and Reeder, R.H. (1992) Topoisomerases and yeast rRNA transcription: negative supercoiling stimulates initiation and topoisomerase activity is required for elongation. *Gene Dev.*, **6**, 1332–1341.
  22. Bryk, M., Banerjee, M., Murphy, M., Knudsen, K.E., Garfinkel, D.J. and Curcio, M.J. (1997) Transcriptional silencing of Ty1 elements in the RDN1 locus of yeast. *Gene Dev.*, **11**, 255–269.
  23. Smith, J.S. and Boeke, J.D. (1997) An unusual form of transcriptional silencing in yeast ribosomal DNA. *Gene Dev.*, **11**, 241–254.
  24. Smith, J.S., Caputo, E. and Boeke, J.D. (1999) A genetic screen for ribosomal DNA silencing defects identifies multiple DNA replication and chromatin-modulating factors. *Mol. Cell Biol.*, **19**, 3184–3197.
  25. Christman, M.F., Dietrich, F.S. and Fink, G.R. (1988) Mitotic recombination in the rDNA of *S. cerevisiae* is suppressed by the combined action of DNA topoisomerases I and II. *Cell.*, **55**, 413–425.
  26. Kim, R.A. and Wang, J.C. (1989) A subthreshold level of DNA topoisomerases leads to the excision of yeast rDNA as extrachromosomal rings. *Cell.*, **57**, 975–985.
  27. Zhu, J. and Schiestl, R.H. (2004) Human topoisomerase I mediates illegitimate recombination leading to DNA insertion into the ribosomal DNA locus in *Saccharomyces cerevisiae*. *Mol. Genet. Genom.*, **271**, 347–358.
  28. Koster, D.A., Croquette, V., Dekker, C., Shuman, S. and Dekker, N.H. (2005) Friction and torque govern the relaxation of DNA supercoils by eukaryotic topoisomerase IB. *Nature.*, **434**, 671–674.
  29. Champoux, J.J. (2011) Human DNA Topoisomerase I: Structure, Enzymology and Biology. *DNA Topoisomerases and Cancer*. Springer New York, New York, NY, pp. 53–69.
  30. Rohner, S., Gasser, S.M. and Meister, P. (2008) Modules for cloning-free chromatin tagging in *Saccharomyces cerevisiae*. *Yeast.*, **25**, 235–239.
  31. 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.
  32. Martin, S.G., Laroche, T., Suka, N., Grunstein, M. and Gasser, S.M. (1999) Relocalization of telomeric Ku and SIR proteins in response to DNA strand breaks in yeast. *Cell.*, **97**, 621–633.
  33. Takahashi, T., Burguiere-Slezak, G., Van der Kemp, P.A. and Boiteux, S. (2011) Topoisomerase I provokes the formation of short deletions in repeated sequences upon high transcription in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA.*, **108**, 692–697.
  34. Dion, V., Kalck, V., Horigome, C., Towbin, B.D. and Gasser, S.M. (2012) Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery. *Nat. Cell Biol.*, **14**, 502–509.
  35. Meister, P., Gehlen, L.R., Varela, E., Kalck, V. and Gasser, S.M. (2010) Visualizing yeast chromosomes and nuclear architecture. *Meth. Enzymol.*, **470**, 535–567.
  36. Neumann, F.R., Dion, V., Gehlen, L.R., Tsai-Pflugfelder, M., Schmid, R., Taddei, A. and Gasser, S.M. (2012) Targeted INO80 enhances subnuclear chromatin movement and ectopic homologous recombination. *Gene Dev.*, **26**, 369–383.
  37. Ponti, A., Schwarb, P., Gulati, A. and Bäker, V. (2007) Huygens remote manager: a web interface for high-volume batch deconvolution. *Imag. Microscop.*, **9**, 57–58.
  38. Huang, J. and Moazed, D. (2003) Association of the RENT complex with nontranscribed and coding regions of rDNA and a regional requirement for the replication fork block protein Fob1 in rDNA silencing. *Gene Dev.*, **17**, 2162–2176.
  39. Nieduszynski, C.A., Hiraga, S.-I., Ak, P., Benham, C.J. and Donaldson, A.D. (2007) OriDB: a DNA replication origin database. *Nucleic Acids Res.*, **35**, D40–D46.
  40. Torres-Rosell, J., Sunjevaric, I., De Piccoli, G., Sacher, M., Eckert-Boulet, N., Reid, R., Jentsch, S., Rothstein, R., Aragón, L. and Lisby, M. (2007) The Smc5-Smc6 complex and SUMO modification of Rad52 regulates recombinational repair at the ribosomal gene locus. *Nat. Cell Biol.*, **9**, 923–931.
  41. Di Felice, F., Cioci, F. and Camilloni, G. (2005) FOB1 affects DNA topoisomerase I in vivo cleavages in the enhancer region of the *Saccharomyces cerevisiae* ribosomal DNA locus. *Nucleic Acids Res.*, **33**, 6327–6337.
  42. Vance, J.R. and Wilson, T.E. (2002) Yeast Tdp1 and Rad1-Rad10 function as redundant pathways for repairing Top1 replicative damage. *Proc. Natl Acad. Sci. USA.*, **99**, 13669–13674.
  43. Liu, C., Pouliot, J.J. and Nash, H.A. (2002) Repair of topoisomerase I covalent complexes in the absence of the tyrosyl-DNA phosphodiesterase Tdp1. *Proc. Natl Acad. Sci. USA.*, **99**, 14970–14975.
  44. Park, H. and Sternglanz, R. (1999) Identification and characterization of the genes for two topoisomerase I-interacting proteins from *Saccharomyces cerevisiae*. *Yeast.*, **15**, 35–41.
  45. Christman, M.F., Dietrich, F.S., Levin, N.A., Sadoff, B.U. and Fink, G.R. (1993) The rRNA-encoding DNA array has an altered structure in topoisomerase I mutants of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA.*, **90**, 7637–7641.
  46. Li, M., Li, T. and Brill, S.J. (2007) Mus81 functions in the quality control of replication forks at the rDNA and is involved in the maintenance of rDNA repeat number in *Saccharomyces cerevisiae*. *Mut. Res.*, **625**, 1–19.
  47. Versini, G., Comet, I., Wu, M., Hoopes, L., Schwob, E. and Pasero, P. (2003) The yeast Sgs1 helicase is differentially required for genomic and ribosomal DNA replication. *EMBO J.*, **22**, 1939–1949.
  48. Defossez, P.A., Prusty, R., Kaerberlein, M., Lin, S.J., Ferrigno, P., Silver, P.A., Keil, R.L. and Guarente, L. (1999) Elimination of replication block protein Fob1 extends the life span of yeast mother cells. *Mol. Cell.*, **3**, 447–455.
  49. Johzuka, K. and Horiuchi, T. (2002) Replication fork block protein, Fob1, acts as an rDNA region specific recombinator in *S. cerevisiae*. *Genes Cells.*, **7**, 99–113.
  50. Shuman, S. and Prescott, J. (1990) Specific DNA cleavage and binding by vaccinia virus DNA topoisomerase I. *J. Biol. Chem.*, **265**, 17826–17836.
  51. Edwards, K.A., Halligan, B.D., Davis, J.L., Nivera, N.L. and Liu, L.F. (1982) Recognition sites of eukaryotic DNA topoisomerase I: DNA nucleotide sequencing analysis of topo I cleavage sites on SV40 DNA. *Nucleic Acids Res.*, **10**, 2565–2576.
  52. Tsang, E. and Carr, A.M. (2008) Replication fork arrest, recombination and the maintenance of ribosomal DNA stability. *DNA Repair.*, **7**, 1613–1623.
  53. Ahn, J., Osman, F. and Whitby, M. (2005) Replication fork blockage by RTS1 at an ectopic site promotes recombination in fission yeast. *EMBO J.*, **24**, 2011–2023.
  54. Caserta, M., Amadei, A., Di Mauro, E. and Camilloni, G. (1989) In vitro preferential topoisomerization of bent DNA. *Nucleic Acids Res.*, **17**, 8463–8474.

## SUPPLEMENTARY MATERIALS

**Strains usage.** Strains usage is as follows, always from left to right. Figure 3C: direct repeat, OFY457; indirect repeat, OFY572. Figure 3D: OFY571, OFY595, OFY588. Figure 3E: (Fob1-9myc) OFY384, OFY592, OFY587, OFY583; (Top1-13myc) OFY571, OFY595, OFY588, OFY584. Figure 3G: OFY457, OFY595. Figure 4: wild-type/*rad1*  $\Delta$  *tdp1*  $\Delta$ , OFY511/OFY510; wild-type/*mus81*  $\Delta$  *tdp1*  $\Delta$ , OFY488/OFY494; wild-type/*top1*  $\Delta$ , OFY384/OFY499. Figure 5C: FF18733, GP100, OFY105/106, OFY436/437, OFY48. Figure 5D: FF18733, FF18734, GP100, OFY106, OFY105, OFY357, OFY435. Figure 5E: FF18733, GP100, OFY106, OFY436/OFY437, FF18742, WDHY201.

**CPT dependent stabilization of Top1 nicks (relates to Figure S6).** Cells were treated with CPT prior to plug preparation, as described (18) with minor modifications. Briefly,  $5 \times 10^8$  spheroplasts in 5 mL Nystatin buffer (18) were treated with 540  $\mu$ M CPT for 4 min, spun down, re-suspended in 300  $\mu$ L 50 mM EDTA, 10 mM Tris-HCl pH 8.0 and equilibrated at 50°C. Spheroplasts were then embedded in agarose, as described above for 2D-gels but CPT was added during the first incubation of the plugs at 37°C. Scanned Southern blots from 1D-gels were quantitated as described in the main Materials and Methods except that a manual baseline had to be used.

**Live-cell microscopy and image analysis.** Live cell imaging experiments were performed with yeast cells grown in the appropriate SC drop out to select for the presence of the CFP-Nop1 (pFN5 and pFN6) or GFP-Nup49 (pUN100-GFP-Nup49) expressing vectors at densities between  $2 \times 10^6$  cells/mL and  $4 \times 10^6$  cells/mL as described (33-35). Each genotype was done using two independent cultures imaged on different days. A single user analyzed the images without knowledge of the genotypes. The cell-cycle phase was inferred using the presence and size of the bud.

For Figure 2D, we used a spinning disc confocal to image cells over a 5  $\mu$ m range in z, using a step size of 100 nm and exposing each optical slice for 100ms. Both wavelengths were acquired at each z position before moving the piezo. Deconvolution was achieved using the Huygens Remote Manager (36). We took advantage of the spot and surface detection functions of the MathLab add-on to Imaris (version 7) as well as a tool made in house to calculate the shortest distance between the spot and the edge of the closest surface. This plug-in will be described in detail elsewhere.

For Figure 2E, optical slices were taken at every 200 nm over 4  $\mu$ m in cells expressing GFP-NUP49 fusion that demarcates the nuclear periphery. The nucleus in the slice where the LacO/LacI spot is in focus was divided into 3 zones of equal surface. The relative distance between the spot and the nuclear periphery was determined and the results were binned where zone 1 is the nuclear periphery and zone 3 is the nuclear interior. A locus that is randomly distributed throughout the nucleus is expected to have an equal chance of being in any one zone.

**Plasmids construction.** Plasmids constructed to create strains carrying ectopic barriers on chromosome IV are as follows. Plasmids pFA6ARFB and pFA6ABFR for eRFB1 and plasmids pUG73dBgRFB and pUG73dBgBFR for eRFB2. pFA6ARFB and pFA6ABFR were obtained by inserting a 450 bp rRFB sequence amplified with adapter-primers between the SacI and SpeI sites of pFA6A-hphNT1 with the blocking side of the RFB being *hph* distal and proximal, respectively. pUG73dBgRFB and pUG73dBgBFR were constructed from pUG73 by disrupting the BglII site of the *klRPL7* gene and subsequently inserting in SacI-SpeI the rRFB-containing SacI-SpeI fragment from pFA6ARFB or pFA6ABFR, respectively. The 3' sequences of the primers, defining the 450 bp ectopic RFB sequence are ttcgtagtattttttatcatc and tattctcttaagtggtactg.

**DNA isolation in agarose plugs for 1D- and 2D-gels and quantification of DSBs.** Isolation of DNA in agarose plugs was done as described (11), with some modifications. For the analysis of the ribosomal and ectopic rRFB,  $2.5 \times 10^8$  and  $1.25 \times 10^9$  cells from exponential cultures, respectively, were chilled in ice water. Sodium azide was added to a final concentration of 0.1% and samples were left in ice slurry for 2 min. After centrifugation (10 min, 3200 g, 4°C), cells were washed twice in cold 50 mM EDTA. For eRFB analysis, cells were additionally treated with 1 mg of Zymolyase-20T in 0.9 M sorbitol, 100 mM Tris-HCl pH 8.0, 50 mM EDTA at 37°C until spheroplasts were formed and harvested by centrifugation. Then, cells for rRFB or eRFB analysis were re-suspended in 300  $\mu$ L of 50 mM EDTA (pH 8.0) containing 400  $\mu$ g of zymolyase and heated to 50°C. 100  $\mu$ L of SEC buffer (10 mM K<sub>2</sub>HPO<sub>4</sub>, 10 mM citric acid, 1 M sorbitol, 100 mM EDTA pH 7.0) equilibrated at 50°C with 5  $\mu$ L  $\beta$ -mercaptoethanol, were mixed with the cell suspension and 400  $\mu$ L of 1% melted agarose (SeaKEM GTG Agarose, Cambrex) in 0.125 mM EDTA and filled into plug moulds. Solidified plugs were incubated in 0.5 M EDTA, 10 mM Tris-HCl pH 8.0, 500  $\mu$ g/mL zymolyase for 90 min at 37°C, washed in 50 mM EDTA and incubated for 22 h in 50 mM EDTA, 10 mM Tris-HCl pH 8.0, 20 mg/mL laurylsarcosyl, 2 mg/mL Proteinase K (PrK). For eRFB analysis, another 4 mg of PrK was added after 16 h and incubation was extended to 38 h. Plugs were successively washed with wash solution (WS; 0.1 mM EDTA, 10 mM Tris-HCl pH 8.0) for 1 h at room temperature (RT), with WS +200  $\mu$ g/ml RNase A (2 h, 37°C), with WS +1 mM PMSF (1-2 h, RT) and with WS (1 h, RT).

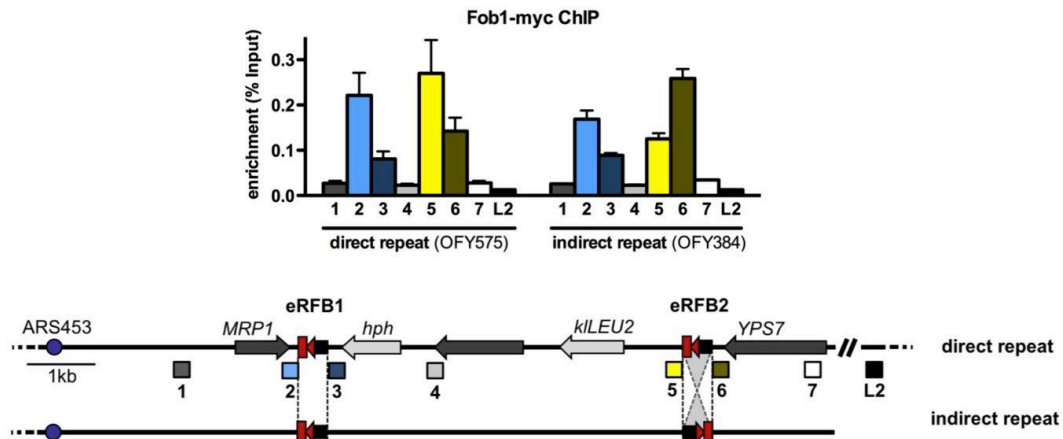
For 1D- and 2D-gels, plugs were digested with 50U (rDNA) or 350U (eRFB) of BglII or XbaI (NEB) for 20 h at 37°C. Scanning of Southern blots was done on a Typhoon 9400 (GEHealthcare) and quantitation with ImageQuant (version 7) applying a rolling-circle model for background subtraction. For 2D-gel electrophoresis, the DSB signal was related to the sum of all RI signals consisting of all BglII or XbaI fragments migrating differently than the linear monomer. RIs include RFs stalled at the RFB (RFB), those harboring an additional converging fork (spike between RFB and Ter) or resembling terminating structures (Ter, X-spike). For 1D-gels, the DSB signal was related to RIs consisting of all signals above background between and including Ter and 2n. When indicated, DSB levels were normalized to those of the wild-type strain of each membrane.

**ChIP from formaldehyde-crosslinked material.** Whole-cell extracts were prepared from formaldehyde-fixed cell cultures as described (30,31) except for cell disruption that was performed in a Fastprep-24 (MP) bead-beater (twice 40 sec at 6.5 m/s) with 0.5 mm Zirconia/silica beads. Whole-cell extracts were sonicated to an average size of 400 bp using a Bioruptor (Diagenode). An aliquot of each extract was reserved as Input before immunoprecipitation (2 h, 4°C) with Dynabeads (Dyna, Invitrogen) precoated with mouse anti-Myc (9E10) antibody. Beads were washed twice for 5 min at 4°C, shaking at

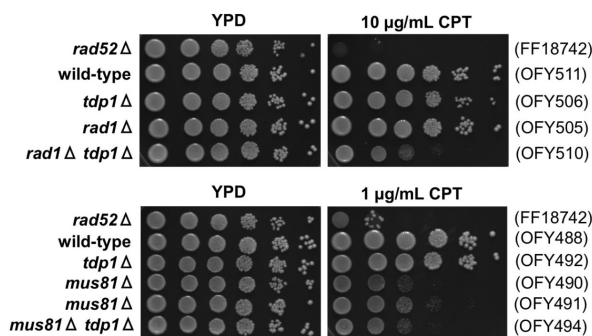


1400rpm, in 50 mM Hepes-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, protease inhibitors (1% Trasylol, 1 mM PMSF, 300  $\mu\text{g}/\text{mL}$  benzamidine, 10  $\mu\text{g}/\text{mL}$  TLCK, 20  $\mu\text{g}/\text{mL}$  TPCK, 2  $\mu\text{g}/\text{mL}$  antipain, 0.5  $\mu\text{g}/\text{mL}$  leupeptine, 1  $\mu\text{g}/\text{mL}$  pepstatinA) and once in 100 mM Tris-HCl pH 8.0, 0.25 M LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, 1 mM EDTA, protease inhibitors. After an additional wash in TE (1 min, shaking), the immunoprecipitate was eluted in 120  $\mu\text{L}$  TE/1% SDS at 65°C (10 min, shaking). After adding 130  $\mu\text{L}$  TE/1% SDS (or 100  $\mu\text{L}$  to 25  $\mu\text{L}$  Input), crosslinking was reversed at 65°C overnight. After Proteinase-K and RNase-A treatment, DNA was recovered by phenol-chloroform and isopropanol precipitation.

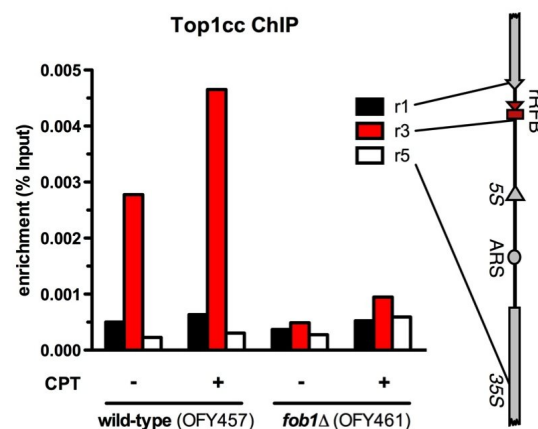
## SUPPLEMENTARY FIGURES



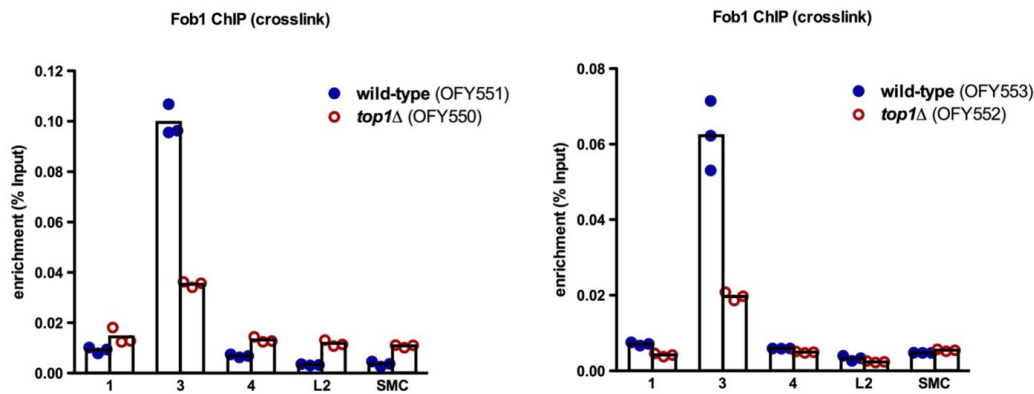
**Figure S1.** Fob1 is present at eRFB1 and eRFB2. Fob1-myc ChIP from asynchronous cultures that were crosslinked with formaldehyde. The mean enrichment with SEM from 4 runs from 2 experiments is shown. Primer positions are depicted on the scheme below. L2, control primer set at a late replicating region on Chr IV.



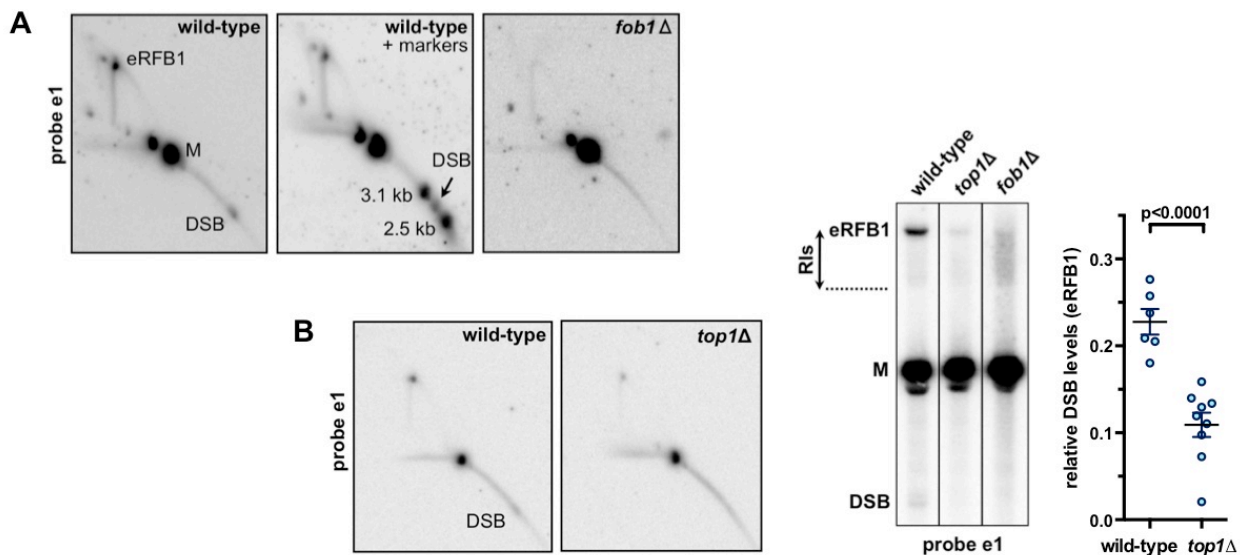
**Figure S2.** Drop tests for CPT sensitivity. Relates to Figure 4A. 10x serial dilutions of indicated strains were plated on YPD plates containing 25 mM Hepes-KOH pH 7.2, 0.25% DMSO and the indicated CPT concentration. Plates were incubated for 3 days at 30°C.



**Figure S3.** Effect of short CPT treatment on Top1cc detection by non-crosslinked ChIP. Top1cc ChIP after short exposure of the cells to CPT.  $10^9$  cells were treated with 5 mg of Zymolyase and the resulting spheroplasts were resuspended in nystatin buffer to which 100  $\mu\text{g}/\text{mL}$  CPT in DMSO, or DMSO alone, was added (18). After 4 min incubation, spheroplasts were resuspended in 800  $\mu\text{L}$  of FA buffer and sonicated to an average DNA size of 400 bp (Bioruptor, Diagenode). ChIP was then performed as described in the main Materials and Methods section. The mean enrichment value from two experiments is shown.

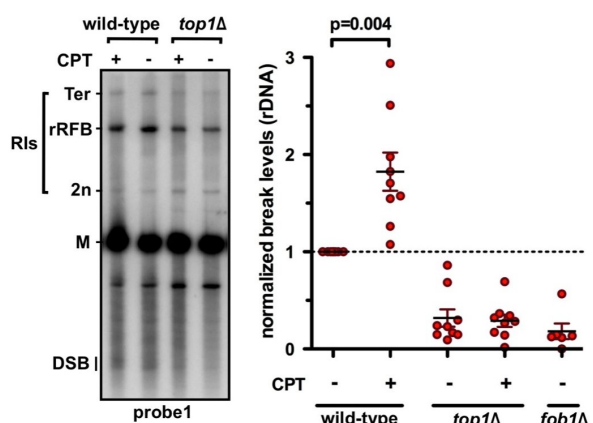


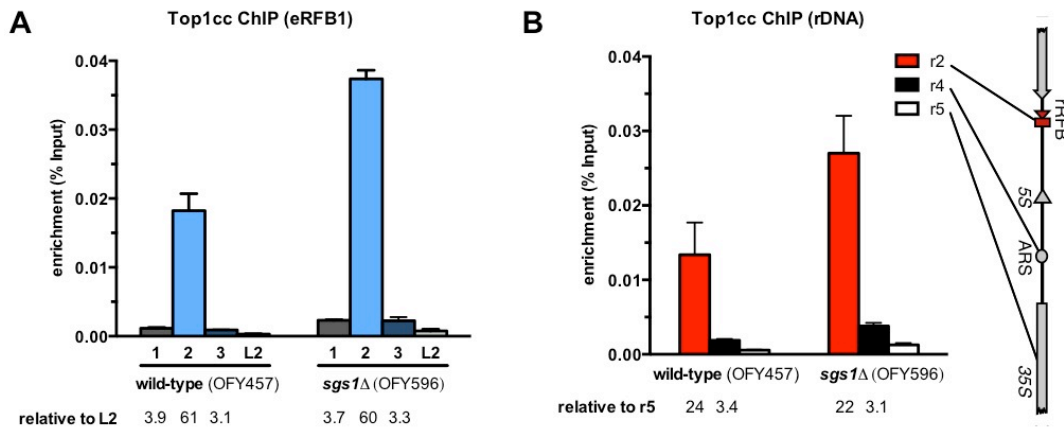
**Figure S4.** Fob1 enrichment at eRFB1 is partially Top1 dependent. Related to Figure 4B. Fob1 enrichment values from two additional wild-type/*top1*Δ strain pairs (left, OFY551/OFY550; right, OFY553/OFY552). Mean enrichment values from 3 qPCR runs of single ChIP experiments are shown with individual values. SMC, control qPCR primer set in the *SMC2* gene. Location of other qPCR primer sets, see Figure S3.



**Figure S5.** DSBs that relate to Top1 activity occur at ectopically located rRFBs. (A) Detection of the DSB signal at eRFB1-stalled RFs. 2D-gel electrophoresis and Southern Blot analysis were done as in Figure 2B, except for the area of the gel transferred for Southern blotting that was extended. PCR-amplified fragments encompassing probe e1 were used as size markers (3.1 kb and 2.5 kb). Note that faint spots on the left of the M and eRFB1 signals correspond to incomplete digestion of the *Bgl*III site between *YDR348C* and *hph*, see Figure 2A. Strains: wild-type, OFY165; *fob1*Δ, OFY56. (B) Breaks at eRFB1-stalled RFs are Top1 dependent. (Top and Bottom Left) Representative Southern blots of 2D- and 1D-gels of *Bgl*III-digested genomic DNA isolated in agarose plugs. (Bottom Right) Quantification of 1D-gels; data points represent the DSB signal normalized to RIs from single 1D-gel experiments. The mean  $\pm$ SEM is indicated. Strains: wild-type, OFY286; *top1*Δ, OFY441; *fob1*Δ, OFY288.

**Figure S6.** CPT increases the DSB signal at the rRFB. The quantification from 1D-gels is shown with a representative Southern blot (probe1). Nystatin-treated spheroplasts were exposed or not to CPT for 2 min prior to DNA isolation and digestion with *Xba*I. Individual values relative to RIs as defined on the left panel and normalized to DSB levels from wild-type DNA on the same gel are plotted together with the mean  $\pm$ SEM. Wild-type, FF18733/OFY286; *top1*Δ, OFY106/OFY441; *fob1*Δ, OFY48/OFY288.





**Figure S7.** The accumulation of Top1cc at the ectopic and endogenous rRFB is not dependent on Sgs1. While we see a trend for elevated Top1cc signal levels at all primer sets, the enrichment normalized to a control region (L2 for eRFB1 and r5 for the rRFB within the rDNA) is not affected. The mean enrichment for Top1cc from non-crosslinked ChIP experiments (N=3, with SEM) of asynchronous cultures is shown at the ectopic eRFB1 (A) and at the rRFB (B) within the rDNA array. Location of primer sets for eRFB1, see Figure S1.

**Table S1.** Yeast strains used in this study.

Strain	Genotype	Reference
FF18733	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1</i>	F. Fabre*
FF18734	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1</i>	F. Fabre
FF18742	<i>a leu2 trp1 ura3 his7-2 lys1-1 rad52::URA3</i>	F. Fabre
WDHY201	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 rad51::URA3</i>	F. Fabre
OFY48	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 fob1::URA3</i>	(13)
GP100	<i>a leu2 ura3 his7-1 lys2-1 sgs1::URA3</i>	(13)
OFY56	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 fob1::URA3 MRP1::eRFB-hphNT1</i>	this study
OFY105	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 top1::kanMX6</i>	this study
OFY106	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 top1::kanMX6</i>	this study
OFY165	<i>a leu2 trp1-289 ura3 his7-1/2 MRP1::eRFB-hphNT1</i>	this study
OFY286	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2</i>	this study
OFY288	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 fob1::URA3 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2</i>	this study
OFY290	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eBFR-kILEU2</i>	this study
OFY357	<i>a leu2-3 ura3-52 his7 lys trp1-289 sgs1::URA3 top1::kanMX6</i>	this study
OFY384	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 FOB1-9myc::kanMX</i>	this study
OFY435	<i>a leu2-3 ura3-52 his7-2 lys trp1-289 sgs1::URA3 top1::kanMX6</i>	this study
OFY436	<i>a leu2-3 ura3-52 his7-2 LYS trp1-289 sgs1::URA3 top1::kanMX6</i>	this study
OFY437	<i>a leu2-3 ura3-52 his7-2 lys trp1-289 sgs1::URA3 top1::kanMX6</i>	this study
OFY441	<i>a leu2-3 trp1 ura3-52 his7-2 lys top1::kanMX6 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2</i>	this study
OFY457	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eBFR-kILEU2 TOP1-13myc::kanMX</i>	this study
OFY461	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 fob1::URA3 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 TOP1-13myc::kanMX</i>	this study
OFY488	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 TOP1-13myc::kanMX MRP1::eRFB-hphNT1 YPS7::eBFR-kILEU2</i>	this study
OFY490	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 TOP1-13myc::kanMX MRP1::eRFB-hphNT1 YPS7::eBFR-kILEU2 mus81::kanMX6</i>	this study
OFY491	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 TOP1-13myc::kanMX MRP1::eRFB-hphNT1 YPS7::eBFR-kILEU2 mus81::kanMX6</i>	this study
OFY492	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 TOP1-13myc::kanMX MRP1::eRFB-hphNT1 YPS7::eBFR-kILEU2 tdp1::kanMX6</i>	this study
OFY494	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 TOP1-13myc::kanMX MRP1::eRFB-hphNT1 YPS7::eBFR-kILEU2 tdp1::kanMX6 mus81::kanMX6</i>	this study
OFY499	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 top1::kanMX6 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 FOB1-9myc::kanMX</i>	this study
OFY505	<i>a leu2-3,112 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 TOP1-13myc::kanMX rad1::caURA3</i>	this study
OFY506	<i>a leu2-3,112 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 TOP1-13myc::kanMX tdp1::kanMX6</i>	this study
OFY510	<i>a leu2-3,112 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 TOP1-13myc::kanMX tdp1::kanMX6 rad1::caURA3</i>	this study
OFY511	<i>a leu2-3,112 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 TOP1-13myc::kanMX</i>	this study
OFY550	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 top1::kanMX6 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 FOB1-9myc::kanMX</i>	this study
OFY551	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 FOB1-9myc::kanMX</i>	this study
OFY552	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 top1::kanMX6 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 FOB1-9myc::kanMX</i>	this study
OFY553	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 FOB1-9myc::kanMX</i>	this study
OFY566	<i>a leu2 trp1-289 URA3::HISpYFP-LacI::ura3-52 his7 lys MRP1::eRFB-hphNT1 YPS7::250xLacO-TRP1</i>	this study
OFY567	<i>a leu2 trp1-289 URA3::HISpYFP-LacI::ura3-52 his7 lys MRP1::eRFB-hphNT1 YPS7::250xLacO-TRP1 fob1::kanMX6</i>	this study
OFY568	<i>a leu2 trp1-289 URA3::HISpYFP-LacI::ura3-52 his7 lys MRP1::eRFB-hphNT1 YPS7::250xLacO-TRP1 top1::kanMX6</i>	this study
ML118-1D	<i>ADE2 can1-100 ura3-1 his3-11,15 leu2-3,112 trp1-1 RAD5 RAD52-YFP RDN25::224xtetO-URA3-I-SceI TetI-mRFP1-I-YGL119W</i>	(39)
OFY570	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eBFR-kILEU2 TOP1-13myc::kanMX fob1::URA3</i>	this study
OFY571	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 TOP1-13myc::kanMX</i>	this study
OFY572	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 TOP1-13myc::kanMX</i>	this study
OFY575	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eBFR-kILEU2 FOB1-9myc::kanMX</i>	this study
OFY577	<i>a leu2 trp1-289 URA3::HISpYFP-LacI::ura3-52 his7-1/2 lys1/2 MRP1::natNT2 YPS7::250xLacO-TRP1</i>	this study
OFY578	<i>a leu2 trp1-289 URA3::HISpYFP-LacI::ura3-52 his7-1/2 lys1/2 MRP1::natNT2 YPS7::250xLacO-TRP1 fob1::kanMX6</i>	this study
OFY579	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 TOP1-13myc::kanMX fob1::URA3</i>	this study
OFY583	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 FOB1-9myc::kanMX tof2::klTRP1</i>	this study
OFY584	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 TOP1-13myc::kanMX tof2::klTRP1</i>	this study
OFY587	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 FOB1-9myc::kanMX tof1::klTRP1</i>	this study
OFY588	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eRFB-hphNT1 YPS7::eRFB-kILEU2 TOP1-13myc::kanMX tof1::klTRP1</i>	this study
OFY592	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eBFR-natNT2 YPS7::eRFB-kILEU2 FOB1-9myc::kanMX</i>	this study
OFY595	<i>a leu2-3 trp1-289 ura3-52 his7-2 lys1-1 MRP1::eBFR-natNT2 YPS7::eRFB-kILEU2 TOP1-13myc::kanMX</i>	this study
OFY596	<i>a leu2-3 trp1-289 ura3-52 his7-2 LYS MRP1::eBFR-natNT2 YPS7::eRFB-kILEU2 TOP1-13myc::kanMX sgs1::URA3</i>	this study

\* Institut Curie, Paris, France

## **Uracil Repair Causes DNA Glycosylase Dependent Genome Instability**

Claudia Krawczyk<sup>1\*</sup>, Marc Bentele<sup>1\*</sup>, Vitaly Latypov<sup>3</sup>, Katrina Woolcock<sup>2</sup>, Robert Ivanek<sup>1</sup>,  
Faiza Noreen<sup>1</sup>, Olivier Fritsch<sup>1</sup>, Primo Schär<sup>1</sup>

1 Department of Biomedicine, University of Basel, Switzerland, 2 Friedrich Miescher Institute  
for Biomedical Research, Basel, Switzerland

\* authors contributed equally to this work

**Correspondence:** Primo Schär  
Department of Biomedicine,  
University of Basel  
Mattenstrasse 28  
CH-4048 Basel  
Switzerland  
Email: primo.schaer@unibas.ch

**Abstract**

Uracil in the DNA arises through misincorporation of dUMP during replication or from cytosine deamination, the latter being pro-mutagenic. Uracil DNA glycosylases (UDGs) excise uracil from DNA to initiate a base excision repair process that restores the canonical DNA base. Thereby, UDGs contribute to genome stability. Under different circumstances, they cooperate with DNA base modifying enzymes to effect mutagenesis or DNA demethylation. Hence their functions are manifold and, given that mammalian cells have four UDGs with similar activities, difficult to dissect. *Schizosaccaromyces pombe* has only two UDGs, Ung1 and Thp1. Using the fission yeast model towards resolving individual functions of these UDGs, we show that Ung1 provides the predominant uracil excision activity in cell-free extracts, while Thp1 activity is detectable only when overexpressed in the absence of active Ung1. By contrast, the two UDGs operate redundantly in living cells. It requires inactivation of both to trigger a significant accumulation of misincorporated uracil in DNA and a significant increase in the mutation rate. We further show that the qualitative outcome of uracil repair depends strongly on the DNA glycosylase engaged; Thp1 but not Ung1 dependent repair is cytotoxic when cells are exposed to 5-fluorouracil or high uracil levels generated by AID expression. Moreover, Thp1 but not Ung1 mediated repair is recombinogenic, accounting for more than 60% of spontaneous and 100% of non-lethal X-ray induced mitotic recombination in fission yeast. These differences most likely reflect the enzymatic properties of the two UDGs, with Ung1 being highly proficient and Thp1 being comparably slow and non-productive owing to its rate-limiting dissociation from AP-sites. Thus, while Ung1 shows all features required for a bona-fide DNA repair enzyme, Thp1 does not, implicating a non-canonical DNA repair function. Given the epigenetic roles of mammalian TDGs, we tested if Thp1 contributes to gene regulation. We found that *thp1<sup>+</sup>* deletion causes a slight but overall suppression of gene expression and a greater variability of transcript levels across replicates, implicating a role for this UDG in the maintenance of a transcriptionally active chromatin.

## Introduction

Genome integrity is threatened by exogenous and endogenous mutagens reacting with and modifying the DNA, but also by the intrinsic chemical instability of the DNA itself, both affecting its coding properties. Besides this, certain DNA modifications serve as substrates in specific genomic processes. Uracil in genomic DNA, for instance, is generated by the spontaneous deamination of cytosine, generating G•U mismatches that will give rise to C→T transition mutations during DNA replication unless corrected. Uracil also arises in DNA through the misincorporation of dUTPs during replication due to the inefficient discrimination between dUTP and dTTP (Goulian et al., 1980). Misincorporated uracils are not mutagenic, but they can interfere with transcription factor binding and thereby affect gene expression (Rogstad et al., 2002; Verri et al., 1990). In addition, uracil in DNA can be actively generated and used as a mark for targeted genome editing. In mammals, this occurs in the context of the fine-tuning of antibody specificity upon B-cell activation, where uracil is formed by the targeted action of the activation-induced cytidine deaminase (AID) to trigger somatic hypermutation (SHM) and class switch recombination (CSR) (Di Noia et al., 2007; Imai et al., 2003; Kavli et al., 2005; Nilsen et al., 2003; Rada et al., 2004; Rada et al., 2002). Strikingly, uracil accumulation is also a mediator of targeted cell death during the development of *D. melanogaster* (Horváth et al., 2013; Muha et al., 2012) and, more recently, 5-hydroxymethyluracil (5hmU) was proposed to be involved in active DNA demethylation in vertebrates, either as a demethylation intermediate arising from deamination of hydroxymethylcytosine (5hmC) (Cortellino et al., 2011; Guo et al., 2011) or by recruiting DNA repair to aid fast DNA demethylation (Pfaffeneder et al., 2014). Thus, uracil and its derivatives in DNA can have a variety of functions depending on the genomic context.

Specialized enzymes have evolved that recognize and excise genomic uracil. Uracil DNA glycosylases (UDGs) catalyze the cleavage of the N-glycosidic bond connecting the uracil base with the sugar moiety of the DNA backbone, thereby generating an abasic site (AP-site). To restore the original DNA sequence, the resulting AP-site is excised from the DNA and filled with a canonical nucleotide by a concerted enzymatic process known as DNA base excision repair (BER) (reviewed in Kim & Wilson, 2012). Given the diverse roles of uracil in DNA, UDGs are not only involved in preventing mutagenesis from cytosine deamination, but they participate in important biological functions including SHM and CSR in immune cells and

active DNA demethylation in developmental processes. Mammalian cells express four nuclear UDGs with different biochemical and biological properties (Brooks et al., 2013; Jacobs & Schar, 2012). UNG2 (uracil DNA glycosylase 2) is highly active on uracil and on the uracil analog 5-fluorouracil (Grogan et al., 2011) and provides the major activity for uracil excision following misincorporation during DNA replication (Otterlei et al., 1999). Consistently, UNG2 is upregulated during S-phase (Hagen et al., 2008), associates with proliferating cell nuclear antigen (PCNA) and replication protein A (RPA) (Ko & Bennett, 2005; Otterlei et al., 1999), and the disruption of *UNG* in mouse results in an accumulation of an appreciable amount of dUMPs in genomic DNA (Andersen et al., 2005; Nilsen et al., 2000). UNG2 is probably also the main enzyme initiating BER of deaminated cytosines, however, other UDGs, in particular Smug1 (single-strand-selective monofunctional uracil DNA glycosylase 1), likely fulfill redundant functions in U•G repair (Akbari et al., 2010; Doseeth et al., 2011; Kavli et al., 2002; Nilsen et al., 2001). UNG2 and SMUG1 initiated uracil excision exhibit different kinetics *in vitro*; UNG2 has a fast turnover and readily dissociates from AP-sites to stimulates incision by APE1, whereas SMUG1 has a higher affinity to AP-sites and delays further processing by APE1 (Pettersen et al., 2007), consistent with a predominant function of UNG2 at the moving replication fork. In addition to the classical error-free uracil repair, UNG2 also is prominently involved in the mutagenic and recombinogenic processing of AID-created uracils during SMH and CSR, respectively, which cannot be substituted for by endogenous SMUG1 levels (Di Noia et al., 2006; Imai et al., 2003; Kavli et al., 2005; Krijger et al., 2009; Rada et al., 2002). SMUG1, on the other hand, appears to contribute a major 5hmU excision activity (Kemmerich et al., 2012) and might be important for the stability of ribosomal RNAs (Jobert et al., 2013).

The two other mammalian UDGs, TDG (thymine DNA glycosylase) and MBD4 (methyl CpG binding domain protein 4), are mostly mismatch dependent and excise a broader spectrum of damaged bases paired or mispaired with guanine (Hardeland et al., 2003; Hendrich et al., 1999; Petronzelli et al., 2000a; Petronzelli et al., 2000b). While knocking out *MBD4* results in a threefold increase in C→T transition frequencies at CpG sites, probably reflecting targeted repair mediated by its methyl binding capacity (Hendrich et al., 1999; Millar et al., 2002), knocking out *TDG* does not increase mutation rates, implicating a minor role for this UDG in mutation avoidance (Cortázar et al., 2011; Goto et al., 2014). The inverse correlation of TDG and UNG2 protein levels throughout the cell cycle with TDG being degraded prior to S-phase

(Hardeland et al., 2007; Shibata et al., 2014) indicates a function of TDG outside of S-phase and/or in non-dividing cells. Notably, disruption of the TDG ortholog in *E. coli* was reported to generate a mutator phenotype in stationary phase rather than exponentially growing cells (Mokkapati et al., 2001). Interestingly, the loss of TDG in mammalian cells increases their resistance to the chemotherapeutic drug 5-fluorouracil (5-FU) (Kunz et al., 2009). This was interpreted to reflect TDG's rate-limiting dissociation from AP-sites (Hardeland et al., 2002; Waters & Swann, 1998) and, hence, an accumulation of this labile repair intermediate under UNG2 saturating conditions (Kunz 2009). Thus, TDG mediated uracil or 5-FU excision seems to be non-productive and toxic. Similarly, repair of 5-FU damage by MBD4 is toxic and its absence from cells confers hyper-resistance (Cortellino et al., 2003). Strikingly, however, TDG is the only UDG essential for embryonic development. Developmental failure of TDG knockout embryos is accompanied by aberrant gene expression and programming of epigenetic marks (Cortázar et al., 2011; Cortellino et al., 2011; Jacobs & Schar, 2012). Consistent with a role in epigenetic gene regulation, TDG was also shown to act as a co-activator of nuclear receptors (Chen et al., 2003; Missero et al., 2001; Um et al., 1998). Moreover, TDG, but also MBD4, have been implicated in active demethylation of 5-methylcytosine (5mC) (Dalton & Bellacosa, 2012; Jacobs & Schar, 2012). This function involves excision of demethylation intermediates, i.e. oxidation products of the ten eleven translocation (TET) proteins (He et al., 2011; Ito et al., 2011; Maiti & Drohat, 2011; Tahiliani et al., 2009) or AID/APOBEC initiated 5mC deamination products (Rai et al., 2008; Cortellino et al., 2011; Hashimoto et al., 2012; Petronzelli et al., 2000b). MBD4, however, has no essential developmental function, nor can it compensate for the loss of TDG in this context (Cortázar et al., 2011; Cortellino et al., 2011; Millar et al., 2002), but it does appear to be involved in targeted DNA demethylation as inferred from its activity at specific promoters of hormone-stimulated genes (Kim et al., 2009). Taken together, UDG functions are diverse and complex in mammalian cells. The four UDGs exert both error-free and mutagenic activities in DNA repair but partial functional redundancies make it difficult to dissect their specific biological roles and to investigate the underlying mechanistic features, i.e. to address the question why there is a need for more than one UDG in cells with similar substrates and different turnover kinetics.

The fission yeast *Schizosaccharomyces pombe* encodes only two different UDGs, the UNG2 ortholog Ung1 and a TDG ortholog Thp1, thus representing a less complex situation than in



mammals. Ung1 was shown to be an active UDG capable of excising uracil from DNA *in vitro* (Elder et al., 2003). Although deletion of *ung1*<sup>+</sup> did not reveal any obvious phenotypes, overexpression in *S. pombe* caused increased cell death, elevated mutation frequencies and an accumulation of AP-sites (Elder et al., 2003). Like its human counterpart, Ung1 was shown to interact with PCNA (Zamir et al., 2012). *S. pombe* Thp1 is active on a broader range of substrates than Ung1, but unlike mammalian TDG its activity is not restricted to double-stranded DNA. In addition, it processes hypoxanthine and 1,N<sup>6</sup>-ethenoadenine irrespective of the opposite base and guanine from G•G mismatches (Hardeland et al., 2003). Interestingly, it is incapable of excising the 5mC derivatives thymine and 5hmU opposite guanine (Hardeland et al., 2003; Borys-Brzywczy et al., 2005), which is in line with the lack of DNA methylation in fission yeast (Antequera et al., 1984; Capuano et al., 2014; Wilkinson et al., 1995). Implicating a potential role for Thp1 in gene regulation, Thp1 was identified in a proteomic screen to physically interact with Tap-tagged Cbf11, which binds sequence-specifically to RNA polymerase II core promoter proximal regions and is involved in gene activation (Oravcová et al., 2013; Převorovský et al., 2009). The fact that the comparably inefficient Thp1 co-evolved with the highly efficient Ung1 indicates a specific need for both types of UDGs in cells.

We used the less complex UDG system in *S. pombe* to address the functional and mechanistic characteristics of Ung1 and Thp1 dependent DNA repair and to separate potential enzyme-specific functions. We show that Ung1 and Thp1 act redundantly in the elimination of DNA uracil and in avoiding mutation by cytosine deamination. Yet, the respective loss of function phenotypes identified distinct modes of operation for the two UDGs. While Ung1 repairs 5-FU induced damage in a fast and predominantly error-free fashion, providing resistance to the drug, Thp1 mediated repair induces cell death. A similar adverse effect of Thp1 but not Ung1 mediated BER is observed in response to AID overexpression, which generates high levels of genomic G•U mispairs. On the grounds of its enzymatic properties, we interpret these cytotoxic effects to reflect the Thp1 dependent formation of long-lived AP-sites in DNA. Consistently, Thp1 induced repair is not only cytotoxic but also recombinogenic, accounting for more than 60% of spontaneous and induced mitotic recombination. Given this unproductive role in DNA repair, we considered additional functions of Thp1 and examined the influence of Thp1 on genome-wide gene expression. We found that Thp1 deficient cells exhibit a slight but overall suppression of

transcription and a higher variation of transcript levels across replicas when compared to wild-type cells. This points at a potential function of Thp1 in maintaining a transcriptionally active chromatin.

## Results

### **Ung1 accounts for the major uracil excision activity in *S. pombe* extracts.**

To assess the relative contributions of Ung1 and Thp1 to the uracil processing activity present in *S. pombe* cells, we measured efficiencies of uracil removal from A•U base pairs and G•U mismatches in cell-free extracts prepared from UDG proficient and deficient cells (Hardeland et al., 2000). As substrates, we used a synthetic, 5'-fluorescein-labeled 60-mer DNA duplex containing a single uracil paired with either an adenine or a guanine (Figure 1A). Uracil excision from these substrates followed by AP-site hydrolysis and electrophoretic separation gives rise to a labeled 23-mer. We also included recombinant Thp1, which we showed previously to excise uracil opposite from adenine and guanine (Dong et al, 2008; Hardeland et al., 2003). Extracts of wild-type cells showed high uracil excision activity on both substrates, none of which was detectable in extracts from *thp1Δung1Δ* double mutant cells. Thus, Ung1 and Thp1 account for all detectable UDG activity in *S. pombe* cells, consistent with the absence of additional UDG encoding genes in the fission yeast genome. To distinguish Ung1 from Thp1 activity in wild-type extracts, we inhibited Ung1 by adding Ugi peptide, a specific and potent inhibitor of the Ung family of UDGs (Wang & Mosbaugh, 1989), but not proteins of the Mug family, including Thp1 (Gallinari & Jiricny, 1996; Hardeland et al., 2003). Addition of Ugi eliminated detectable uracil processing in wild-type cells, both on A•U and G•U substrates. Considering a detection limit at 5% of total substrate used in the assay, we conclude that Ung1 accounts for more than 95% of uracil excision activity in wild-type cell extracts. The missing Thp1 activity could be explained either by lack or poor expression of Thp1 or by a comparably slow rate of uracil excision. We confirmed expression of endogenous TDG in exponentially growing cells by immunoblotting with a Thp1 directed antibody (Figure 1B), and tested the feasibility to measure Thp1 activity in extracts by ectopically overexpressing *thp1<sup>+</sup>* under the control of the inducible *nmt1* promoter (Maundrell, 1990). Induction of expression in wild-type cells yielded a 500- to 800-fold increase in Thp1 protein levels (Figure 1B) and a robust uracil excision activity on both A•U and G•U substrates under fully Ung1 inhibited conditions. Hence, Thp1, synthesized in

its authentic physiological environment, has the ability to excise uracil from DNA. Nevertheless, Ung1 constitutes the dominant uracil excision activity measurable in cell extracts, while endogenous Thp1 contributes only marginally, if anything, under these conditions.

### **Ung1 and Thp1 both remove uracil from genomic DNA**

While the nuclear localization of Ung1 was confirmed earlier (Elder et al., 2003), subcellular localization of Thp1 remained elusive. We identified two putative nuclear localization signals (NLSs) using the Wolf PSORT tool (Horton et al., 2007); a monopartite NLS consisting of four residues (pat4) and starting at amino acid 26 and a bipartite NLS at residue 115. We addressed the subcellular localization of Thp1 using fluorescence microscopy. To this end, we fused a sequence encoding an enhanced green fluorescent protein (EGFP) with the 3' end of the endogenous *thp1*<sup>+</sup> open reading frame to express a C-terminally tagged Thp1-EGFP fusion protein at endogenous levels. Indicative of a nuclear localization of Thp1 and consistent with the presence of NLSs, we observed a distinctive nuclear EGFP signal (Figure 2A). Given that both Ung1 and Thp1 are expressed and localize to the nucleus in *S. pombe* cells and that Ung1 constitutes the predominant uracil processing activity in biochemical assays, we next wanted to address the role of both UDGs in uracil removal in living cells by assessing the uracil content in genomic DNA isolated from wild-type, Ung1-, Thp1- and doubly deficient cells. We isolated DNA of exponentially growing cells in agarose plugs to obtain intact genomic DNA for analysis by pulse field gel electrophoresis (PFGE). This DNA was in-plug digested with *E. coli* Ung uracil-DNA glycosylase and/or human AP-endonuclease APE1 to excise potential uracil bases and incise the DNA at the resulting AP-sites, respectively. This treatment will thus generate single-stranded DNA breaks at uracil residues, which, if closely spaced on opposite strands, would give rise to breakage and fragmentation of the double-stranded DNA. To visualize such DNA fragmentation, we analyzed the digested DNA by PFGE (Figure 2B). Substantial smearing of chromosomal bands was apparent only in DNA derived from *thp1Δung1Δ* double mutant cells digested with both Ung and APE1, but not in DNA from either single mutant or wild-type cells at the same conditions. This result clearly indicates that uracil arises in DNA of vegetatively growing *S. pombe* cells, most likely through incorporation of dUMPs during DNA replication, but also by spontaneous cytosine deamination. As uracil accumulation is only detectable in the UDG depleted double mutant

cells, we conclude that in the cellular context both Ung1 and Thp1 contribute to the elimination of these spontaneously arising uracils and that both can compensate for the loss of the other.

### **Lack of uracil repair generates a moderate mutator phenotype**

Having established that uracil accumulates in the absence of Ung1 and Thp1, we wondered whether this coincides with an increased mutation rate as expected if a sizable fraction of these uracils resulted from deamination of cytosine. We thus made use of the powerful genetics of *S. pombe* and assessed mutation rates and spectra in wild-type and *ung1Δ*, *thp1Δ* and *thp1Δung1Δ* mutant cells. First, we measured reversion rates at the *ade6-M387* allele (Schar & Kohli, 1993), a G→C transversion that renders the *ade6* encoded protein non-functional and causes adenine auxotrophy. Any base substitution at *ade6-M387* will restore adenine prototrophy and the mutation events can be scored by sequencing of prototrophic clones. Mutation rate assessments in fluctuation analyses yielded an increase of the reversion rate by 3.4-fold in *thp1Δung1Δ* cells as compared to the wild-type, with non-overlapping 95% confidence intervals indicating statistical significance of the rate differences (Figure 2C, left panel). Importantly, disruption of only *ung1<sup>+</sup>* or *thp1<sup>+</sup>* did not result in increased mutation rates, indicating that the two UDGs can fully compensate for the loss of the other in the defense against mutagenesis by cytosine deamination. The analysis of mutation spectra by sequencing one randomly chosen adenine prototrophic clone per culture of the fluctuation test (to avoid effects of clonal amplification) confirmed that the *ade6-M387* reversion assay picks up both transversion and transition base substitutions (Figure 2C, right panel). The resulting spectra revealed a significant and specific increase of the C→T transition rate by a factor of 6.3 in *thp1Δung1Δ* double mutants relative to the wild-type ( $p < 0.004$ , Chi-square test), while there was no significant change in either single mutant. This result implicates that Thp1 and Ung1 operate synergistically in the removal of uracil from pre-mutagenic G•U mismatches in cells. Notably, Thp1 deficient cells displayed a tendency for elevated transversion mutations, implicating a function of Thp1 in the repair of base lesions other than cytosine deamination. This would be consistent with the broad substrate spectrum of this UDG (Hardeland et al., 2003). To corroborate the synergistic interaction of Ung1 and Thp1 in mutation avoidance, we performed an additional forward mutation assay, scoring for canavanine resistance (Kaur, Fraser, Freyer, Davey, & Doetsch,

1999). Similar to the result of the reversion assay, *thp1Δung1Δ* double mutant cells but neither of the single mutants showed a moderate mutator phenotype when compared to wild-type cells (Figure 2D). The observed 2.3-fold increase in forward mutation rate was small, but statistically significant by the criterion of non-overlapping 95% confidence intervals. From these results, we conclude that Thp1 and Ung1 have redundant functions in the repair of cytosine deamination damage in cells.

### **Ung1 and Thp1 operate differently on 5-Fluorouracil induced DNA lesions**

Having shown that Thp1 and Ung1 act synergistically in the prevention of uracil accumulation and cytosine deamination associated mutations, we further explored potential non-redundant functions of the two UDGs as implicated by the slightly different spectra of mutations arising in their absence. As *TDG* inactivation in mammalian cells was shown to increase the cellular resistance to the uracil analogue 5-FU (Kunz et al., 2009), we assessed the effect of *thp1<sup>+</sup>* and *ung1<sup>+</sup>* deletion on the 5-FU sensitivity of *S. pombe* cells. 5-FU exerts its cytotoxic effects by interfering with the cell metabolism at different levels. 5-FU inhibits the thymidylate synthase, which converts dUMP to dTMP. Exposure to 5-FU thus generates an imbalance in the nucleotide pool, in particular in the dUTP/dTTP ratio with the consequence of increased uracil incorporation into DNA (An et al., 2007; Copur et al., 1995; Peters et al., 2002). Besides this, 5-FU and its metabolites can also be directly incorporated into both DNA and RNA (Ardalan et al., 1980; Glazer & Peale, 1979; Kunz et al., 2009; Mojardín, Botet, Quintales, Moreno, & Salas, 2013). 5-FU in DNA is a substrate for BER (A. L. Jacobs & Schar, 2012; Matuo et al., 2010; Pettersen et al., 2011; Seiple, 2006) and will thus trigger repair events; the consequences of 5-FU in RNA are not entirely clear. To confirm 5-FU dependent uracil incorporation into DNA, we differentially digested DNA of 5-FU treated wild-type cells with uracil DNA glycosylase and AP-endonuclease and examined the resulting fragmentation of chromosomal DNA by PFGE. The observed DNA fragmentation was consistent with increased AP-site levels (APE1 digest) in DNA of 5-FU treated cells as compared to untreated cells (Figure 2B and 3A). No further increase in fragmentation became visible in doubly digested DNA (Ung and APE1), indicative of most uracils being excised by either UDG. In survival assays, we noticed a dose dependent 5-FU cytotoxicity in wild-type cells (Figure 3B), suggesting that at low concentrations 5-FU mediated DNA damage is readily repaired, while at higher doses it induces cell death. When testing the 5-

FU sensitivity of *ung1Δ* and *thp1Δ* single and double mutant cells, we observed diametrically opposite effects of defects of the two UDGs. While Thp1 deficient cells displayed a striking hyper-resistance towards 5-FU, Ung1 deficient cells are hyper-sensitive at the applied doses. Consistent with this 5-FU sensitizing effect of Thp1, we observed that Thp1 overexpression from the inducible *nmt1* promoter strongly increased 5-FU sensitivity in wild-type and *ung1Δthp1Δ* double mutant cells (Figure 3C). In the double mutant configuration, Ung1 mediated 5-FU hyper-sensitivity is largely cancelled out by the Thp1 related hyper-resistance. Thus, similar to mammalian cells but more pronounced, 5-FU sensitivity of fission yeast is mediated by Thp1, while Ung1 promotes cell survival. This clearly separates the two UDGs into different pathways generating different repair outcomes.

Next, we investigated whether Thp1 expression also has an effect on 5-FU induced mutations. To this end, we determined rates of spontaneous and 5-FU induced mutations that confer canavanine resistance and determined the mutation inducibility in Ung1 and/or Thp1 deficient cells relative to wild-type cells (Figure 3D, white boxes). Ung1 deficient cells displayed the highest 5-FU inducibility of mutations, indicating that Ung1 controlled DNA repair reduces the mutagenic effect of 5-FU. Notably, the hyper-inducibility of mutations in Ung1 deficient cells depends on functional Thp1 as it was fully suppressed in the *ung1Δthp1Δ* double mutant. Thus, Ung1 repairs 5-FU induced DNA lesions through a non-mutagenic pathway, while Thp1 mediated repair is mutagenic. From these results we concluded that Ung1 and Thp1 mediate excision repair of uracil and 5-FU through pathways generating different outcomes, the former increasing viability and suppressing mutations, the latter being cytotoxic and mutagenic.

### **Excision of AID generated uracil in DNA causes cytotoxicity**

To rationalize the differential response of Ung1 and Thp1 deficient cells to 5-FU exposure, we considered that, due to its unusually high affinity for AP-sites, (Hardeland et al., 2003), Thp1 initiated BER, unlike Ung1 initiated BER, will generate protected AP-sites, retarding their excision and further processing by the repair system. Hence, in the case of Ung1 saturation (5-FU exposure), Thp1 generated AP-sites would accumulate and trigger cell death and mutations. To test this hypothesis, we challenged cells by artificially generating G•U mismatches across the genome by overexpression of the human activation-induced cytosine deaminase (AID; Figure 3E, left panel). Ectopic AID expression, driven by the

inducible *nmt1*-promoter induced severe cell death in wild-type, Ung1, Thp1 and doubly deficient cells (Figure 3E, right panel). Yet, the *ung1Δthp1Δ* double mutants showed a more than 100-fold higher resistance to AID-expression than the wild-type cells, establishing that excision of AID-generated uracil, and thus the generation of AP-sites, is a major cause of the AID toxicity. Moreover, AID toxicity was less pronounced in Thp1 than in Ung1 deficient cells, indicating that repair of G•U mismatches by Thp1 is more disruptive than repair by Ung1. We therefore conclude that the repair of excessive uracil in DNA by UDGs generates toxic intermediates that are less efficiently processed when Thp1 rather than Ung1 is the initiating glycosylase.

### **Thp1 initiated BER is recombinogenic**

We reasoned that the slow dissociation of Thp1 from AP-sites is responsible for its toxic effects during 5-FU exposure and AID expression. The generation of long-lived labile AP-sites is likely to be accompanied by occasional DNA breakage, which would then trigger recombination events. To study a potential relationship between Thp1 initiated BER and recombination, we crossed a heteroallelic duplication of mutant alleles of the *ade6<sup>+</sup>* gene (Schuchert & Kohli, 1988) into our UDG proficient and deficient strain backgrounds. This system assays homologous recombination events (primarily gene conversion with and without associated crossover) that restore the intact *ade6<sup>+</sup>* allele and hence adenine prototrophy (*ade<sup>+</sup>*, Figure 4A). We first determined spontaneous mitotic recombination rates in vegetatively growing cells (Figure 4B). Both wild-type and *ung1Δ* cells exhibited similar rates of spontaneous recombination at this locus (overlapping 95% confidence intervals). Thp1 deficient cells, however, displayed significantly reduced mitotic recombination as *thp1Δ* single and *ung1Δthp1Δ* double mutants showed a more than 60% reduction of the wild-type rate. This illustrates that Thp1 but not Ung1 dependent excision of base lesions is responsible for a significant part of spontaneously occurring mitotic recombination at the *ade6* locus.

To corroborate the recombinogenic action of Thp1, we then determined the contribution of Thp1 to recombination under damage inducing conditions. For this purpose, we exposed wild-type and Thp1 deficient cells to a non-lethal dose of IR (X-rays, 100 Gy, >95% cell survival), which we expected to generate ROS mediated Thp1 relevant DNA base damage, such as oxidized pyrimidines (Figure 4E). Under these conditions, we measured a 2-fold

increase in recombination rates in wild-type cells that was significant while no increase was apparent in Thp1 deficient cells. Hence, both, spontaneous and IR induced mitotic recombination show a clear Thp1 dependency. We thus conclude that repair of base-damage induced by Thp1 is recombinogenic and can result in gross genomic instability.

### **Effect of Thp1 on global gene expression**

Considering the questionable performance of Thp1 in DNA repair, we wondered whether this UDG could have evolved for purposes other than the conventional repair of damaged DNA bases. In the style of mammalian TDG, which has similar properties and was eventually shown to play an important role in gene regulation (Cortázar et al., 2011; Cortellino et al., 2011), we tested the possibility of a Thp1 engagement in gene expression. To this end we analyzed genome-wide gene expression in wild-type and *thp1* $\Delta$  *h*<sup>+</sup> strains capable of mating type switching (Beach & Klar, 1984; Egel et al., 1984). We isolated nine spores of each genotype from a cross between wild-type and *thp1* $\Delta$  cells and extracted RNA from cultures expanded from these spores. We then pooled RNAs of three independent cultures for each genotype and used three of these pools for gene expression analysis on an Affymetrix GeneChip<sup>®</sup> *S. pombe* Tiling 1.0FR Array.

Comparing the mean log<sub>2</sub> gene expression values of wild-type and Thp1 deficient cells, we spotted only 3 mRNAs that were differentially expressed by more than 2-fold (log<sub>2</sub> fold change (log<sub>2</sub>FC) >1 or log<sub>2</sub><1), one of them derived from the *thp1*<sup>+</sup> gene itself, as expected (Figure 5A). Similarly, only minor expression differences were notable when 5' and 3' untranslated regions (5UTR, 3UTR, respectively) as well as long non-coding RNAs were analyzed separately (lncRNAs, Supplementary Figure 1A). Hence, under vegetative growth conditions, loss of Thp1 does not deregulate distinct patterns of gene expression. Yet, Thp1 deficient cells displayed a general suppression of transcript levels normalized to wild-type levels (log<sub>2</sub>FC) affecting all, mRNAs (Figure 5B), 5UTRs, 3UTRs and lncRNAs (Supplementary Figure 1B). 96% of all mRNAs with a log<sub>2</sub>FC higher than 0.3 or smaller than -0.3 (n= 471) were less expressed in Thp1 deficient cells than in wild-type cells (Figure 5C). The high proportion of genes showing lower transcription upon *thp1*<sup>+</sup> deletion did not depend on the applied threshold as it was also observed when all mRNAs were analyzed (Supplementary Figure 1C). Although the differences were mostly small and individually not statistically significant, they



nevertheless indicate a consistent trend towards lower mRNA expression across most of the transcriptome in *thp1Δ* cells.

Remarkably, principal component analysis of mRNA expression patterns not only clearly separated the wild-type from *thp1Δ* samples but also revealed that the WT triplicates were more closely related to each other than the *thp1Δ* triplicates (Figure 5D). Again, we observed this effect for all analyzed gene subsets (Supplementary Figure 2A). The higher variation among the *thp1Δ* triplicates became also visible by plotting the standard deviations (SDs) resulting from the wild-type and *thp1Δ* expression measurements (Figure 5E for mRNAs, Supplementary Figure 2B for 5UTR, 3UTR, lncRNAs). We considered the possibility that the higher SD in the *thp1Δ* samples is a technical artifact resulting from the generally lower expression levels in these cells; lower values are generally more variable than higher values. However, differences in expression were rather low and the expression data show that SDs are higher upon *thp1<sup>+</sup>* deletion irrespective of its upregulation (Supplementary Figure 3). Thus, the decreased and more variable expression observed in Thp1-deficient cells appear to be true phenotypes.

Considering a possible role of Thp1 in gene regulation, we then explored potential Thp1-chromatin associations within the *S. pombe* genome by chromatin immunoprecipitation and deep-sequencing (ChIP-seq). For this purpose we generated a strain expressing Thp1 with a C-terminal 13myc tag from its endogenous locus. Following confirmation of the activity of the myc-tagged Thp1 (Supplementary Figure 4), we performed duplicate ChIP experiments, which were then sequenced along with an input control. Sequence reads were mapped to 500 base pair windows spanning the complete genome. Comparing immunoprecipitated (IP) and input samples of the two biological replicates (A and B) revealed no sequence enrichments above twofold ( $\log_2FC > 1$ , Supplementary Figure 5). Nevertheless, setting the  $\log_2FC$  threshold at 0.3 (values before rdmup correction for PCR bias) identified 23 regions covering 41 genes that were reproducibly enriched (Supplementary Figure 6). 9 of the 23 regions located at or close to the three centromeres, a fact that also explains the high proportion for tRNA genes among the identified genes (38% compared to the expected 3.5%; Wood et al., 2002). As highly expressed genes, including tRNA genes may represent false positives in ChIP experiments (Teytelman et al., 2013), and the centromeric enrichments did not show as distinct peaks when compared to the input, we conclude Thp1 does not detectably associate with specific regions in the genome. Nevertheless, we cannot strictly rule out the possibility

that the ChIP protocol applied was not successful as we lack any positive control for Thp1 binding.

## Discussion

*S. pombe* has evolved with two active UDGs, *Ung1* and *Thp1*, that seem to operate side by side. Their loss of function phenotypes show that, while both are capable of and even partially redundant in excising uracil from DNA, they do it in different ways. Uracil repair by *Ung1* is efficient and productive, repair by *Thp1* is inefficient and generates cytotoxic, mutagenic and recombinogenic intermediates. Hence, whereas *Ung1* fulfills requirements of a robust DNA repair enzyme, *Thp1* does not and is therefore likely to have specialized functions beyond mutation avoidance, possibly in stabilizing gene expression.

In biochemical assays with cell-free extracts, *Ung1* accounts for all detectable uracil excision activity in wild-type cells, irrespective if the uracil is paired with adenine or guanine. Notwithstanding considerations of spatio-temporal separation, this suggests that *Ung1* provides the major activity for the repair of misincorporated dUMPs and deaminated cytosines. A similar prominent role has been reported on the grounds of biochemical evidence for UNG2 in mammalian cells (Doseth et al., 2011; Otterlei et al., 1999). *Thp1* activity on U•A and U•G substrates was not detectable, unless overexpressed, and appears to account for less than 5% of uracil excision activity in cell-free extracts. In contrast to this biochemical assessment, we found that inactivating *thp1*<sup>+</sup> increased the level of genomic uracil in an *Ung1* deficient background, showing that *Thp1* does contribute to uracil repair and can compensate for the loss of *Ung1* in living cells (Figure 2B). Together, these results suggest that *Ung1* operates as the prime UDG in wild-type cells while *Thp1* provides a backup activity that engages mainly upon the loss or saturation of *Ung1*. Such redundancy was also apparent as a significant and synergistic increase in spontaneous mutation rates in *ung1Δthp1Δ* double mutant cells (Figure 2C and 2D), showing mutation spectra consistent with a specific loss of repair of deaminated cytosines. Notably, however, the canavanine forward mutation assay revealed a trend for increased mutations in *thp1Δ* single mutant cells (Figure 2D and 3D). Consistent with broader substrate spectrum of *Thp1* (Borys-Brzywczy et al., 2005; Hardeland et al., 2003), this can be explained if certain base lesions are processed by *Thp1* but not by *Ung1*. The mutation spectra for the *ade6-M387* revertants in the *thp1Δ* background indicate that these lesions might generate C→G and C→A

transversions. Thus, although Thp1 has overlapping functions with Ung1 in uracil repair, it also contributes to the repair of other base lesions. Etheno DNA adducts, such as 3,N4-ethenocytosine, are candidates for such lesions, as they not only lead to C→T transitions, but also to G→T transversion mutations (Moriya et al., 1994) and they were suggested as a substrate for human TDG (Goto et al., 2014). Although the exact nature of these Thp1 relevant lesions remains to be clarified, this suggests that Mug proteins may indeed play a role in the defense against lipid-peroxidation.

Intriguingly, exposing cells to the uracil analog 5-FU revealed diametrically opposite features of Ung1 and Thp1 dependent uracil repair. Whereas Ung1 was protective against 5-FU mediated cytotoxicity (Figure 3B), Thp1 was detrimental, sensitizing wild-type as well as Ung1 deficient cells to the drug (Figure 3C). A cell sensitizing effect of Thp1 was also observed upon overexpression of AID, generating genomic uracil by cytosine deamination, thus establishing that toxicity is a general feature of Thp1 dependent uracil repair rather than a specific occurrence following 5-FU treatment. As previously noted for a 5-FU hyper-resistance observed upon *TDG* inactivation in mouse and human cells (Kunz et al., 2009), the Thp1 mediated 5-FU toxicity might be a consequence of the enzymatic properties associated with this subfamily of UDGs (Fitzgerald & Drohat, 2008; Hardeland et al., 2000; 2003; Jacobs & Schar, 2012; Waters & Swann, 1998; Waters et al., 1999). Due to its rate-limiting dissociation from AP-sites, Thp1 is likely to delay downstream processing of the cytotoxic and mutagenic repair intermediate. An accumulation of labile AP-sites in Thp1 proficient cells, particularly under conditions of stalled BER such as under 5-FU treatment, would thus compromise cell viability and increase the mutation rate (Figure 3). 5-FU induced Thp1 dependent mutations were detectable most clearly in Ung1 deficient cells, establishing that uracil repair by Ung1 is largely non-mutagenic whereas channeling repair into the Thp1 pathway is mutagenic. Together, these results show that although DNA uracil can be repaired by both UDGs, Ung1 or Thp1 initiated repair proceeds through distinct pathways producing qualitatively different outcomes.

Prolonged existence of AP-sites might not be restricted to cells facing large amounts of base lesions following exposure to DNA damaging agents, but could also occur in unchallenged cells, though at a much lower level. One indication that Thp1 contributes to the accumulation of spontaneously arising AP-sites is its effect on mitotic recombination; AP-

sites are well known to interfere with replication fork progression (Bao & Kow, 2009; Maga et al., 2009; Shibutani, 1997), which can occasionally initiate homologous recombination. We found that inactivation of *thp1<sup>+</sup>* significantly reduces the spontaneous mitotic recombination rate (Figure 4C). Thp1 may indeed do this by blocking downstream repair of its own base excision products and/or by binding to AP-sites generated by spontaneous base hydrolysis or by other DNA glycosylases. The latter is supported by Thp1's contribution to IR induced mitotic recombination. The increase in recombination observed by the non-lethal X-ray dose applied was fully dependent on the presence of functional Thp1. Assuming that oxidative lesions, including various substrates for other DNA glycosylases, are predominant under these conditions, this result strongly indicates that Thp1 can interfere with the repair of any AP-site generated.

We were surprised to find that *thp1Δung1Δ* double mutant cells show only a moderate (6.3-fold) increase in the C→T transition rate, this being assessed at an intragenic position of a transcribed gene. Provided that no other enzyme removes G•U mismatches as suggested by the lack of residual activity in extracts of the double mutant cells and by the absence of other UDG homologs in the fission yeast genome, the C→T transition rate equals the cytosine deamination frequency at this base pair. Assuming a G-C content of 36% (Wood et al., 2002) and a genome size of 13.8 Mb, the genome-wide rate of endogenous cytosine deamination can be calculated to 0.04 events per haploid *S. pombe* cell cycle, i.e. one event per 25 cell divisions.

The mammalian Thp1 ortholog has specific functions in the regulation of gene expression and the control of epigenetic chromatin states (Cortázar et al., 2011; Cortellino et al., 2011). We were not able to identify a specific role of fission yeast Thp1 in gene regulation; Thp1 deficient cells did not show distinct patterns of genomic deregulation. Interestingly though, we found a wide-spread suppression and increased variation of gene expression in Thp1 deficient cells. 96% of differentially expressed genes ( $|\log_2|$ normalized enrichment > 0.3) showed lower expression in *thp1Δ* than in wild-type cells, indicating that Thp1 contributes to gene activity in wild-type cells. Also, biological triplicates showed an increase variation of global expression in Thp1 deficient cells compared to wild-type cells, indicating a deficiency in stable regulation in these cells (Figure 5A-5C). On the basis of these observations, we

propose a function of Thp1 in maintaining gene activity by a mechanism that remains to be clarified. Given the absence of DNA methylation and hence a need for processing of demethylation intermediates in *S. pombe* (Antequera et al., 1984; Capuano et al., 2014; Wilkinson et al., 1995), Thp1 may operate at the chromatin level, through interactions with transcription factors or chromatin modifiers. Consistently, a TAP-tagging approach has identified a physical interaction of Thp1 with Cbf11 (Pancaldi et al., 2012), a transcription factor capable of activating gene transcription (Oravcová et al., 2013). Yet, the lack of specific Thp1 enrichment in our ChIP-seq experiment is inconsistent with a strong functional association of Thp1 with transcription factors

Our results provide insight into the distinct modes of operation used by the two apparently redundant UDGs of fission yeast. In uracil repair, Ung1 fulfills all expectations for an efficient and productive DNA repair enzyme, while Thp1 shows cytotoxic, mutagenic and recombinogenic activities that seem highly unsuitable for general genome maintenance. Such functional properties, however, may serve specific biological functions under specific circumstances or on yet unidentified biological relevant substrates. One aspect to be addressed will be the temporal separation of function of the two UDGs in the cell cycle as this would require an efficient, high fidelity enzyme during S-phase and a versatile counterpart for repair of uracil in non-replicating DNA, where the turnover of the enzyme might have to be coordinated with the recruitment of downstream acting repair proteins. Such temporal separation was proposed for UNG2 and TDG in mammalian cells as UNG2 expression is upregulated during S-phase while TDG is absent at the same time (Hagen et al., 2008; Hardeland et al., 2007). Forced TDG expression in S-phase even abrogates cell cycle progression. Also, UDGs might act in different biological contexts, an emerging concept is the targeted editing of genomes or chromatin structure by DNA base modification, excision and repair. To induce somatic hypermutation and class switch recombination, error-prone BER is coupled to targeted enzymatic deamination (Di Noia et al., 2007; Imai et al., 2003; Kavli et al., 2005; Nilsen et al., 2003; Rada et al., 2002; 2004). BER is also targeted to gene regulatory region to participate in active DNA demethylation initiated by enzymatic 5mC oxidation (He et al., 2011; Ito et al., 2011; Maiti & Drohat, 2011; Tahiliani et al., 2009) or possibly thymine oxidation (Pfaffeneder et al., 2014). These are vertebrate examples and it is currently unclear whether and for what purpose analogous mechanisms operate in fission

yeast. It is interesting though, that the phenotype of the Thp1 knockout includes some perturbation of gene expression and it will be of high interest to further identify potential Thp1 related gene regulatory mechanism.

Here, we presented evidence for Ung1 and Thp1 dependent repair giving rise to different outcomes. Repair by Ung1 appears fast and error-free, whereas Thp1 mediated repair can be mutagenic and recombinogenic, suggesting specialized, not yet identified, cellular functions. Addressing growth conditions other than vegetative growth might reveal specific functions of Thp1, e.g. in sporulation, pseudohyphal growth (Borup, 2006) or stationary phase cells.

## **Material and Methods**

### **Strains, growth conditions, plasmids, oligonucleotides**

All strains used in this study derive from wild-type strain 972 h<sup>-</sup> and are listed in Table 1. Gene disruption and tagging were done using PCR-generated fragments providing short homology for recombination as described (Bahler et al., 1998). Complete *thp1*<sup>+</sup> or *ung1*<sup>+</sup> open reading frames were replaced in strain PRS000d with the kanMX4 marker gene cassette (Wach et al., 1994) and the *ura4*<sup>+</sup> gene cassette (Grimm et al., 1988), respectively. The 13myc-tagged strain was obtained by inserting a 13myc tag and the kanMX marker cassette to the C-terminus of the endogenous *thp1*<sup>+</sup> open reading frame. Enhanced green fluorescent (EGFP) tagging of *thp1*<sup>+</sup> was obtained by first inserting a *ura4*<sup>+</sup> gene 3' of the *thp1*<sup>+</sup> open reading frame in the PRS000d strain that was subsequently replaced by the EGFP coding sequence.

Standard media and growth conditions were described before (Forsburg & Rhind, 2006). EMM-Can-G plates contain 3.75 g/l of glutamate as nitrogen source and 75 µg/ml of L-canavanine sulphate (Sigma). For inducible *nmt1* promoter-driven overexpression, cells initially grown under repressive conditions (EMM + 5 µg/ml thiamine) to 1x10<sup>7</sup> cells/ml were washed twice in water, diluted for induction with EMM lacking thiamine to a density of 5x10<sup>5</sup> cells/ml and grown at 30°C for 16 to 48 h (as indicated).

pPRS271 and pREP1-AID for Thp1 and AID overexpression, respectively, was constructed by sub-cloning the *thp1*<sup>+</sup> and human AID open reading frames as an NdeI-SalI PCR fragment into the matching sites of pREP1 (ARS1/LEU2 based episomal *S. pombe* vector for inducible *nmt1*

promoter-driven gene expression). For Thp1-13myc overexpression, a genomic Thp1-13myc was amplified, adding BglIII and Sall restriction sites, and inserted into pREP1 cut with BamHI and Sall.

Oligonucleotide sequences are available on request.

### **Cell sensitivity tests**

For 5-FU sensitivity tests, cells were grown to  $5 \times 10^6$  cells/ml and washed in water. Ten fold serial dilutions were spotted onto MMA plates supplemented with the indicated concentration of 5-FU and incubated for 6-12 days at 26°C. To test sensitivity to AID overexpression, cells transformed with pREP1-AID were grown under repressive conditions (EMM + 5 µg/ml thiamine) to a density of  $5 \times 10^6$  cells/ml, harvested and washed in water. Serial dilutions of cells in water were spotted onto EMM (inducible conditions) and incubated for 7 days at 30°C.

### **Cell-free extracts**

Cells were harvested at  $1 \times 10^7$  cells/ml and sequentially washed in water and twice in 10 ml of lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 20% glycerol, 0.1% Tween-20, 10 mM β-mercaptoethanol, 1 mM PMSF, 1x complete™ protease inhibitors (Roche, Switzerland)). Cells were resuspended in ice-cold lysis buffer and disrupted by adding glass beads and vigorous shaking for ten times 30 s in a Mini-Beadbeater (Biospec Products). After centrifugation (20800 g, 20 min, 4°C) the protein concentration of the supernatant was determined using a Bradford assay.

### **Western blotting**

50 µg of protein were separated on 10% SDS-polyacrylamide gels, transferred to a nitrocellulose membrane (Protran®, Schleicher & Schuell) and incubated with an affinity purified rabbit polyclonal anti-Thp1 antiserum or a mouse anti-AID antiserum according to standard protocols. The rabbit polyclonal antiserum was raised against purified recombinant Thp1 protein and subsequently affinity purified as Thp1p coated affinity beads (Affi-Gel 10 beads, Bio-Rad) according to standard procedures. This Thp1 antibody was diluted 1:500 in TBS-T containing 5% dry milk as blocking reagent.

### Base release assay

A 6His-tagged Thp1 was expressed in *E. coli* and purified as described (Hardeland et al., 2003). Substrate preparation and nicking assay were done as previously described (Hardeland et al., 2000). Briefly, 60 bp double-stranded oligonucleotide substrates containing A•U and G•U mismatches were prepared by annealing an unlabeled oligonucleotide (5'-TAGACATTGCCCTCGAGGTACCATGGATCCGATGTC(A/G)ACCTCAAACCTAGACGAATTCCG-3') to a 5'-fluorescein (F)-labeled lower oligonucleotide strand (5'-CGGAATTCGTCTAGGTTTGAGGTUGACATCGGATCCATGGTACCTCGAGGGCAATGTCTA-3'). 0.5 μM labeled and 1 μM unlabeled oligonucleotides were annealed in 10 mM Tris-HCl pH 8.0 and 50 mM NaCl by heating to 95 °C for 5 min and gradual cooling to 25 °C over 30 min. Nicking assays were performed in a total volume of 20 μl nicking buffer (50 mM Tris-HCl pH 8.0, 1 mM DTT, 0.1 mg/ml BSA, 1 mM EDTA). 2 pmol of recombinant Thp1 or 7 μg of cell free extract was incubated with 2 pmol of substrate. When indicated, 1 unit of uracil-DNA glycosylase inhibitor (Ugi, New England BioLabs) was included. Reactions containing cell free extracts or recombinant Thp1 were incubated for 30 min at 30°C or for 15 min at 37°C (3 h at 37°C for the Thp1-13 myc activity test), respectively. For AP-site cleavage, NaOH was added to a final concentration of 90 mM, followed by an incubation at 99°C for 10 min. DNA was ethanol-precipitated and resuspended in 10 μl of formamide gel loading buffer (90% formamide, 1 X TBE) for 5 min at 99°C. After cooling, samples were separated using denaturing polyacrylamide gel electrophoresis. Fluorescein-labeled DNA was visualized using the blue fluorescent mode of the Storm 860 (Molecular Dynamics).

### Spontaneous reversion rates and spectra at the *ade6-M387* locus

The *ade6-M387* allele has been described before (Schar & Kohli, 1993). For fluctuation tests, at least 24 colonies freshly grown on non-selective medium (YEA supplemented with adenine and uracil) were used to inoculate 4 ml of fresh non-selective medium. After 24 h of incubation at 30°C, cells were harvested, washed with water and resuspended in 3 ml of water. 1 ml of the cell suspension was plated onto five MMA plates supplemented with uracil. Adenine prototrophic colonies (revertants) were scored after 9-11 days at 30°C. In parallel, the number of viable cells was determined on non-selective YEA after 3 days at 30°C. Spontaneous reversion rates were determined by the method of the median (Lea &



Coulson, 1949). The 95% confidence interval for the median was calculated according to Nair (Nair, 1940). To determine the mutation spectra among the spontaneous *ade6-M387* revertants, one randomly chosen adenine prototrophic clone per culture was used for sequence analysis by colony PCR. Transversion and transition rates were calculated using the following formula: (transversions or transitions) x (reversion-rate) / (number of tested clones).

### **Rates of spontaneous forward mutation conferring canavanine resistance**

To measure rates of spontaneous forward mutation conferring canavanine resistance (Kaur et al., 1999), 30 independent cultures from 3 experiments were used for fluctuation analysis. Cells from single colonies, freshly grown on non-selective YEA plates, were grown in 5 ml YEL for 28 h at 30°C, harvested, washed in water, resuspended in 1.2 ml of water and plated onto 4 selective EMM-Can-G plates (0.2 ml/plate). Colonies were scored after 14-16 days at 30°C. In parallel, viable cells were determined on non-selective YEA plates after 3 days at 30°C. Mutation rates and the 95% confidence interval were calculated as described above. For mutation rates upon overexpression of *Thp1*, three experiments with four cultures each were performed.

For 5-FU-induced mutation rates, one culture per strain was split in two halves and grown to midlog phase in YEL. One half culture was supplemented with a final concentration of 10 mg/l 5-FU and both were incubated for 76 h at 30°C; the cells were washed and plated as described above. Viable cells were scored as above. Mutation rates were determined by the method of the median (Lea & Coulson, 1949) from 3 to 4 independent experiments. Inducibility of canavanine resistance was calculated as follows: [(induced mutations of mutant) x (spontaneous mutations of wild-type)] / [(spontaneous mutations of mutant) x (induced mutations of wild-type)].

### **Intra-chromosomal mitotic recombination**

One freshly grown red (adenine auxotrophic) colony was isolated from YEA plates of strains PRS807, PRS809, PRS811 and PRS813 and resuspended in water. 100 cells were plated onto two YEA plates and incubated for 4 days at 30°C. 24-28 randomly chosen colonies (excluding white colonies) were resuspended in a final volume of 0.7 ml. For IR induced recombination analyses, cells suspended in H<sub>2</sub>O were X-ray exposed (100Gy, 100kv, 0.2mm Al) before

plating to YEA. Two times 0.2 ml were spread onto EMM plates supplemented with uracil. Adenine prototrophs were scored after 3 days at 30°C. The number of viable cells in each colony was determined on non-selective YEA plates after 3 days at 30°C. Rates of spontaneous mitotic recombination were determined by the method of the median from 24-28 colonies per strain tested (Lea & Coulson, 1949). The 95% confidence interval was calculated according to Nair (Nair, 1940).

### **Fluorescence microscopy**

Cells were arrested by glucose starvation for 16 h in EMM containing 0.5% glucose. 500 µl of cells were harvested, washed in water, resuspended in 1 M sorbitol and stained with Hoechst 33342 dye (1 µg/ml, Sigma). Images of EGFP and Hoechst 33342 were analyzed using a LEICA fluorescence microscope equipped with a PL Fluorstar objective (X100 oil) and a LEICA DC200 digital camera.

### **DNA isolation in agarose plugs and PFGE**

DNA isolation in agarose plugs and PFGE were modified from (Baumann & Cech, 2000). Cells were grown in YEL to a density of  $1 \times 10^7$  cells/ml at 30°C. For 5-FU treatment, cells were grown in YEL to  $5 \times 10^6$  cells/ml before 5-FU was added to a final concentration of 10 mg/l for 48 h at 30°C. Cells were washed in water and resuspended in PRO buffer (1 M sorbitol, 25 mM EDTA, 20 mM Tris-HCl pH 8.0, 10 mM DTT) at a density of  $5.5 \times 10^8$  cells/ml.  $5.0 \times 10^8$  cells were treated with 1 mg/ml Zymolyase-20T (Amsbio) at 37°C for 60 min. Spheroplasts were collected and resuspended in 120 µl of TSE buffer (10 mM Tris-HCl pH 7.5, 900 mM sorbitol, 45 mM EDTA) before being mixed with 375 µl of 1.5% agarose (ultra pure L.M.P. agarose, GIBCO BRL) in TSE equilibrated at 43°C. Agarose plugs were poured, washed in PW1 (50 mM Tris-HCl pH 7.5, 250 mM EDTA, 1% SDS) at 50°C for 4 h and incubated twice in PW2 (10 mM Tris-HCl pH 9.0, 500 mM EDTA, 1% N-lauroyl sarcosine, 1 mg/ml proteinase K) at 50°C for 22 h. Finally, plugs were washed five times for 60 min in 5 ml T10XE (10 mM Tris-HCl pH 7.5, 10 mM EDTA) and kept in T10XE at 4°C until use.

For digestion, plugs were washed twice in water for 15 min and once in 5 ml Ung-digestion buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl) for 60 min at room temperature. Four quarters of the same plug were transferred to 4 ml Ung-digestion buffer supplemented with  $\text{MgCl}_2$  (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM  $\text{MgCl}_2$ ) containing either no enzyme(s),

13 units Ung uracil-DNA glycosylase (E. coli, New England BioLabs), 3.5 pmol APE1 (recombinant human AP-endonuclease) or both enzymes and incubated at 37°C for 4 h. The plugs were washed three times in 4 ml T10XE at 4°C, equilibrated two times for 1 h in 0.5 X TBE (pH 8.3, 45 mM Tris-HCL, 45 mM borate, 1 mM EDTA) and loaded onto 0.8% agarose gels (Chromosomal Grade Agarose, Bio-Rad) in 0.5 X TBE. Electrophoresis was performed in a CHEF DR III pulsed-field electrophoresis system (Bio-Rad) using 0.5 X TBE running buffer (72 h, 14°C, 2 V/cm, 1800 s switch time, included angle: 100°). DNA was stained with ethidium bromide (1 µg/ml) in 0.5 X TBE for 60 min and de-stained for 90 min in 0.5 X TBE.

### **Genome-wide expression profiling**

WT (975+) and *thp1Δ* (PRS555) strains were crossed and 9 WT and 9 *thp1Δ* strains of the mating type *h<sup>+</sup>* derived from single spores were selected. RNA isolation was done as previously described (Emmerth et al., 2010) with acidic (pH 4.3) phenol. RNAs from three samples were pooled and three such pools from each genotype were used for profiling on Affymetrix GeneChip® *S. pombe* Tiling 1.0FR Arrays. The arrays were processed, analyzed and annotated as previously described (Woolcock et al., 2012). Correlation and regression analyses were done in Prism. Principal component analysis was performed using Cluster 3.0 and the results visualized in TOPCAT.

### **Chromatin immunoprecipitation paired with next generation sequencing (ChIP-Seq)**

50 ml cultures were grown at 30°C to a density of  $5 \times 10^6$  cells/ml, switched to 18°C for 2 h and crosslinked at 18°C by adding 1% or 3% of freshly prepared para-formaldehyde for 30 or 15 min, respectively. Para-formaldehyde was quenched with 1/20 volume of 2.5 M glycine for 5 min at room temperature. Cells were harvested and washed in 20 ml of ice-cold PBS and 1 ml of IP buffer (50 mM Hepes/KOH, 140 mM NaCl, 1mM EDTA/NaOH pH 8.0, 1% Triton-X, 0.1% sodium deoxycholate). After resuspending the cells in 400 µl of IP+ buffer (IP buffer containing 1x proteinase inhibitors (Roche complete), 50 mM beta-glycerophosphate, 50 mM NaF, 1 mM Na<sub>3</sub>VO<sub>4</sub>, 1 mM PMSF), cells were lysed using 1 ml glass beads in a FastPrep®-24 (2x 30 sec at maximum speed). 400 µl of IP+ were added to the lysate before sonicating for 15 min (30 sec ON, 30 sec OFF, Bioruptor® Plus). Lysates were cleared by centrifugation (17000 g, 5 min) and the supernatants frozen at -80°C. 25 µl were saved as input controls.

For each sample 40  $\mu$ l of magnetic sheep anti mouse beads (M280, Invitrogen) were blocked in 600  $\mu$ l PBS containing 5% BSA for 30 min rolling at 4°C. The blocking solution was replaced by fresh one and 1.5  $\mu$ g of purified 9E10 mouse anti-myc antibody (Gasser laboratory, Friedrich Miescher Institut, Basel, Switzerland) added for 2 h rolling at 4°C. The beads were washed twice in IP buffer before being added to the defrosted cell extracts. After 2 h of rolling at 4°C the beads were successively washed in 1ml of IP+ buffer, IP+ buffer containing 500 mM NaCl and wash buffer (10 mM Tris-HCl, 250 mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA pH8.0) for 5 min rolling at 4°C. After a final wash in 1 ml TE pH8.0 for 1 min, the IP was eluted in 125  $\mu$ l TES (50 mM Tris-HCl pH8.0, 10 mM EDTA pH8.0, 1% SDS), shaking 1400 rpm at 65°C for 15 min. Crosslinks were reversed by the addition of 100  $\mu$ l TES and overnight incubation at 65°C. The samples were treated with 400  $\mu$ g Proteinase K for 2 h at 50°C. DNA was recovered by phenol:chloroform:isoamyl alcohol extraction and isopropanol precipitation and resuspended in 20  $\mu$ l of TE pH8.0. Samples of the same treatment were pooled and purified using Minelute columns (Qiagen) to remove the observed unwanted small fragments of DNA. At the end samples of the different crosslinking conditions were pooled at the same ratio for input and IP samples such as to obtain enough material for next generation sequencing.

IP and input DNA was send tot the Genome Technology Access Center (GTAC; <https://gtac.wustl.edu/index.php>) at the Washington University School of Medicine in St. Louis, Missouri, USA. The CHIP DNA was blunt-ended before adenosine was added to the 3' end and sequencing adaptors were ligated to the ends. The fragments were size selected to 200-600 base pairs, and underwent amplification for 15 cycles. The resulting libraries were sequenced using the Illumina HiSeq-2500 as single reads extending 50 bases and the raw data was demultiplexed. Reads were mapped to the *S. pombe* genome (Bsgenome.Spombe.Ensembl.ASM294v2.22, ref: <http://www.pombase.org>) using the Bowtie aligner (version 1.0.1) (Langmead et al., 2009) implemented in QuasR package (<http://www.bioconductor.org>, version 1.4.2, Rbowtie version 1.4.5) allowing up to 10 best alignment positions in the genome. The sample specific fragment sizes were estimated from cross correlation profiles of read density on both chromosomal strands using the Chipcor software (ref: <http://ccg.vital-it.ch/chipseq>). Reads were shifted by half of the fragment size (80bp) towards the middle of the fragment.

The *S. pombe* genome was tiled with 500bp non-overlapping windows and the ChIP enrichment was calculated for two independent biological replicates in each of these windows. ChIP enrichment was calculated as  $e = \log_2( (n_{fg} / N_{fg} * \min(N_{fg}, N_{bg}) + p) / (n_{bg} / N_{bg} * \min(N_{fg}, N_{bg}) + p) )$ , where  $n_{fg}$  and  $n_{bg}$  are the number of overlapping foreground and background (input chromatin) read alignments, respectively.  $N_{fg}$  and  $N_{bg}$  are the total number of aligned reads in foreground and background samples, and  $p$  is a pseudocount constant ( $p=8$ ) used to minimize the sampling noise for peaks with very low counts. In order to minimize the PCR bias we also calculated and compared the ChIP enrichments after removal of potential PCR duplicates using `remove duplicate` function in `samtools` (`samtools` version 0.1.19-44428cd, parameters: `rmdup -s`) (Li et al., 2009).

### Acknowledgements

We thank Marc Bühler for his valuable input and advice.

### References

- Akbari, M., Solvang-Garten, K., Hanssen-Bauer, A., Lieske, N. V., Pettersen, H. S., Pettersen, G. K., et al. (2010). Direct interaction between XRCC1 and UNG2 facilitates rapid repair of uracil in DNA by XRCC1 complexes. *DNA Repair*, 9(7), 785–795. doi:10.1016/j.dnarep.2010.04.002
- An, Q., Robins, P., Lindahl, T., & Barnes, D. E. (2007). 5-Fluorouracil incorporated into DNA is excised by the Smug1 DNA glycosylase to reduce drug cytotoxicity. *Cancer Research*, 67(3), 940–945. doi:10.1158/0008-5472.CAN-06-2960
- Andersen, S., Heine, T., Sneve, R., König, I., Krokan, H. E., Epe, B., & Nilsen, H. (2005). Incorporation of dUMP into DNA is a major source of spontaneous DNA damage, while excision of uracil is not required for cytotoxicity of fluoropyrimidines in mouse embryonic fibroblasts. *Carcinogenesis*, 26(3), 547–555. doi:10.1093/carcin/bgh347
- Antequera, F., Tamame, M., Villanueva, J. R., & Santos, T. (1984). DNA methylation in the fungi. *The Journal of Biological Chemistry*, 259(13), 8033–8036.
- Ardalan, B., Cooney, D. A., Jayaram, H. N., Carrico, C. K., Glazer, R. I., Macdonald, J., & Schein, P. S. (1980). Mechanisms of sensitivity and resistance of murine tumors to 5-fluorouracil. *Cancer Research*, 40(5), 1431–1437.
- Bahler, J., Wu, J. Q., Longtine, M. S., Shah, N. G., McKenzie, A., Steever, A. B., et al. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast*, 14(10), 943–951. doi:10.1002/(SICI)1097-0061(199807)14:10<943::AID-YEA292>3.0.CO;2-Y
- Bao, G., & Kow, Y. W. (2009). Effect of sequence context and direction of replication on AP site bypass in *Saccharomyces cerevisiae*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 669(1-2), 147–154. doi:10.1016/j.mrfmmm.2009.06.006
- Baumann, P., & Cech, T. R. (2000). Protection of telomeres by the Ku protein in fission yeast. *Molecular Biology of the Cell*, 11(10), 3265–3275.

- Beach, D. H., Klar, A. J. (1984). Rearrangements of the transposable mating-type cassettes of fission yeast. *The EMBO Journal*, 3(3), 603.
- Borup, M. T. (2006). Stress-Induced Switch to Pseudohyphal Growth in *S. Pombe*. *Cell Cycle*, 5(18), 2138–2145. doi:10.4161/cc.5.18.3206
- Borys-Brzywczy, E., Arczewska, K. D., Saparbaev, M., Hardeland, U., Schär, P., & Kuśmierk, J. T. (2005). Mismatch dependent uracil/thymine-DNA glycosylases excise exocyclic hydroxyethano and hydroxypropano cytosine adducts. *Acta Biochimica Polonica*, 52(1), 149–165. doi:055201149
- Brooks, S. C., Adhikary, S., Rubinson, E. H., & Eichman, B. F. (2013). Recent advances in the structural mechanisms of DNA glycosylases. *Biochimica Et Biophysica Acta (BBA) - Proteins and Proteomics*, 1834(1), 247–271. doi:10.1016/j.bbapap.2012.10.005
- Capuano, F., Mülleder, M., Kok, R., Blom, H. J., & Ralser, M. (2014). Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. *Analytical Chemistry*, 86(8), 3697–3702. doi:10.1021/ac500447w
- Chen, D., Lucey, M. J., Phoenix, F., Lopez-Garcia, J., Hart, S. M., Losson, R., et al. (2003). T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha. *The Journal of Biological Chemistry*, 278(40), 38586–38592. doi:10.1074/jbc.M304286200
- Copur, S., Aiba, K., Drake, J. C., Allegra, C. J., & Chu, E. (1995). Thymidylate synthase gene amplification in human colon cancer cell lines resistant to 5-fluorouracil. *Biochemical Pharmacology*, 49(10), 1419–1426.
- Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., Macdougall, E., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*. doi:10.1038/nature09672
- Cortellino, S., Turner, D., Masciullo, V., Schepis, F., Albino, D., Daniel, R., et al. (2003). The base excision repair enzyme MED1 mediates DNA damage response to antitumor drugs and is associated with mismatch repair system integrity. *Proceedings of the National Academy of Sciences of the United States of America*, 100(25), 15071–15076. doi:10.1073/pnas.2334585100
- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, 146(1), 67–79. doi:10.1016/j.cell.2011.06.020
- Dalton, S. R., & Bellacosa, A. (2012). DNA demethylation by TDG. *Epigenomics*, 4(4), 459–467. doi:10.2217/epi.12.36
- Di Noia, J. M., Rada, C., & Neuberger, M. S. (2006). SMUG1 is able to excise uracil from immunoglobulin genes: insight into mutation versus repair. *The EMBO Journal*, 25(3), 585–595. doi:10.1038/sj.emboj.7600939
- Di Noia, J. M., Williams, G. T., Chan, D. T. Y., Buerstedde, J.-M., Baldwin, G. S., & Neuberger, M. S. (2007). Dependence of antibody gene diversification on uracil excision. *The Journal of Experimental Medicine*, 204(13), 3209–3219. doi:10.1084/jem.20071768
- Dong, L., Mi, R., Glass, R. A., Barry, J. N., & Cao, W. (2008). Repair of deaminated base damage by *Schizosaccharomyces pombe* thymine DNA glycosylase. *DNA Repair*, 7(12), 1962–1972. doi:10.1016/j.dnarep.2008.08.006
- Doseth, B., Visnes, T., Wallenius, A., Ericsson, I., Sarno, A., Pettersen, H. S., et al. (2011). Uracil-DNA glycosylase in base excision repair and adaptive immunity: species differences between man and mouse. *The Journal of Biological Chemistry*, 286(19), 16669–16680. doi:10.1074/jbc.M111.230052

- Egel, R., Beach, D., & Klar, A. (1984). Genes required for initiation and resolution steps of mating-type switching in fission yeast. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 81(11), 3481–3485.
- Elder, R. T., Zhu, X., Priet, S., Chen, M., Yu, M., Navarro, J. M., et al. (2003). A fission yeast homologue of the human uracil-DNA-glycosylase and their roles in causing DNA damage after overexpression. *Biochemical and Biophysical Research Communications*, 306(3), 693–700.
- Emmerth, S., Schober, H., Gaidatzis, D., Roloff, T., Jacobeit, K., & Bühler, M. (2010). Nuclear retention of fission yeast dicer is a prerequisite for RNAi-mediated heterochromatin assembly. *Developmental Cell*, 18(1), 102–113. doi:10.1016/j.devcel.2009.11.011
- Fitzgerald, M. E., & Drohat, A. C. (2008). Coordinating the Initial Steps of Base Excision Repair APURINIC/APYRIMIDINIC ENDONUCLEASE 1 ACTIVELY STIMULATES THYMINE DNA GLYCOSYLASE BY DISRUPTING THE PRODUCT COMPLEX. *The Journal of Biological Chemistry*, 283(47), 32680–32690. doi:10.1074/jbc.M805504200
- Forsburg, S. L., & Rhind, N. (2006). Basic methods for fission yeast. *Yeast*, 23(3), 173–183. doi:10.1002/yea.1347
- Gallinari, P., & Jiricny, J. (1996). A new class of uracil-DNA glycosylases related to human thymine-DNA glycosylase. *Nature*, 383(6602), 735–738. doi:10.1038/383735a0
- Glazer, R. I., & Peale, A. L. (1979). The effect of 5-fluorouracil on the synthesis of nuclear RNA in L1210 cells in vitro. *Molecular Pharmacology*, 16(1), 270–277.
- Goto, M., Shinmura, K., Matsushima, Y., Ishino, K., Yamada, H., Totsuka, Y., et al. (2014). Human DNA glycosylase enzyme TDG repairs thymine mispaired with exocyclic etheno-DNA adducts. *Free Radical Biology and Medicine*. doi:10.1016/j.freeradbiomed.2014.07.044
- Goulian, M., Bleile, B., & Tseng, B. Y. (1980). The effect of methotrexate on levels of dUTP in animal cells. *The Journal of Biological Chemistry*, 255(22), 10630–10637.
- Grimm, C., Kohli J., Murray J. and Maundrell, K. (1988). Genetic Engineering of Schizosaccharomyces Pombe: a System for Gene Disruption and Replacement Using the Ura4 Gene as a Selectable Marker. *Molecular & General Genetics : MGG* 215 (1): 81–86.
- Grogan, B. C., Parker, J. B., Guminski, A. F., & Stivers, J. T. (2011). Effect of the thymidylate synthase inhibitors on dUTP and TTP pool levels and the activities of DNA repair glycosylases on uracil and 5-fluorouracil in DNA. *Biochemistry*, 50(5), 618–627. doi:10.1021/bi102046h
- Guo, J. U., Su, Y., Zhong, C., Ming, G.-L., & Song, H. (2011). Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell*. doi:10.1016/j.cell.2011.03.022
- Hagen, L., Kavli, B., Sousa, M. M. L., Torseth, K., Liabakk, N. B., Sundheim, O., et al. (2008). Cell cycle-specific UNG2 phosphorylations regulate protein turnover, activity and association with RPA. *The EMBO Journal*, 27(1), 51–61. doi:10.1038/sj.emboj.7601958
- Hardeland, U., Bentele, M., Jiricny, J., & Schar, P. (2000). Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *The Journal of Biological Chemistry*, 275(43), 33449–33456. doi:10.1074/jbc.M005095200
- Hardeland, U., Bentele, M., Jiricny, J., & Schar, P. (2003). The versatile thymine DNA-glycosylase: a comparative characterization of the human, Drosophila and fission yeast orthologs. *Nucleic Acids Research*, 31(9), 2261–2271.
- Hardeland, U., Kunz, C., Focke, F., Szadkowski, M., & Schar, P. (2007). Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2. *Nucleic Acids Research*, 35(11), 3859–3867.

doi:10.1093/nar/gkm337

- Hardeland, U., Steinacher, R., Jiricny, J., & Schar, P. (2002). Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *The EMBO Journal*, *21*(6), 1456–1464. doi:10.1093/emboj/21.6.1456
- Hashimoto, H., Liu, Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., et al. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research*, *40*(11), 4841–4849. doi:10.1093/nar/gks155
- He, Y.-F., Li, B.-Z., Li, Z., Liu, P., Wang, Y., Tang, Q., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science (New York, NY)*, *333*(6047), 1303–1307. doi:10.1126/science.1210944
- Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., & Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature*, *401*(6750), 301–304. doi:10.1038/45843
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, *35*(Web Server), W585–W587. doi:10.1093/nar/gkm259
- Horváth, A., Békési, A., Muha, V., Erdélyi, M., & Vértessy, B. G. (2013). Expanding the DNA alphabet in the fruit fly: Uracil enrichment in genomic DNA. *Fly*, *7*(1), 23. doi:10.4161/fly.23192
- Imai, K., Slupphaug, G., Lee, W.-I., Revy, P., Nonoyama, S., Catalan, N., et al. (2003). Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination. *Nature Immunology*, *4*(10), 1023–1028. doi:10.1038/ni974
- Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., et al. (2011). Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science (New York, NY)*, *333*(6047), 1300–1303. doi:10.1126/science.1210597
- Jacobs, A. L., & Schar, P. (2012). DNA glycosylases: in DNA repair and beyond. *Chromosoma*, *121*(1), 1–20. doi:10.1007/s00412-011-0347-4
- Jobert, L., Skjeldam, H. K., Dalhus, B., Galashevskaya, A., Vågbo, C. B., Bjørås, M., & Nilsen, H. (2013). The Human Base Excision Repair Enzyme SMUG1 Directly Interacts with DKC1 and Contributes to RNA Quality Control. *Molecular Cell*, *49*(2), 339–345. doi:10.1016/j.molcel.2012.11.010
- Kaur, B., Fraser, J. L., Freyer, G. A., Davey, S., & Doetsch, P. W. (1999). A Uve1p-mediated mismatch repair pathway in *Schizosaccharomyces pombe*. *Molecular and Cellular Biology*, *19*(7), 4703–4710.
- Kavli, B., Andersen, S., Otterlei, M., Liabakk, N. B., Imai, K., Fischer, A., et al. (2005). B cells from hyper-IgM patients carrying UNG mutations lack ability to remove uracil from ssDNA and have elevated genomic uracil. *The Journal of Experimental Medicine*, *201*(12), 2011–2021. doi:10.1084/jem.20050042
- Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., et al. (2002). 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. *Jbc.org*.
- Kemmerich, K., Dingler, F. A., Rada, C., & Neuberger, M. S. (2012). Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil- and UNG-backup uracil-excision activities and increases cancer predisposition of Ung<sup>-/-</sup>Msh2<sup>-/-</sup> mice. *Nucleic Acids Research*, *40*(13), 6016–6025. doi:10.1093/nar/gks259
- Kim, M.-S., Kondo, T., Takada, I., Youn, M.-Y., Yamamoto, Y., Takahashi, S., et al. (2009). DNA demethylation in hormone-induced transcriptional derepression. *Nature*, *461*(7266),



- 1007–1012. doi:10.1038/nature08456
- Kim, Y. J., Wilson, D. M. III (2012). Overview of Base Excision Repair Biochemistry. *Current Molecular Pharmacology*, 5(1), 3.
- Ko, R., & Bennett, S. E. (2005). Physical and functional interaction of human nuclear uracil-DNA glycosylase with proliferating cell nuclear antigen. *DNA Repair*, 4(12), 1421–1431. doi:10.1016/j.dnarep.2005.08.006
- Krijger, P. H., Langerak, P., van den Berk, P. C. M., & Jacobs, H. (2009). Dependence of nucleotide substitutions on Ung2, Msh2, and PCNA-Ub during somatic hypermutation. *The Journal of Experimental Medicine*, 206(12), 2603. doi:10.1084/jem.20091707
- Kunz, C., Focke, F., Saito, Y., Schuermann, D., Lettieri, T., Selfridge, J., et al. (2009). Base Excision by Thymine DNA Glycosylase Mediates DNA-Directed Cytotoxicity of 5-Fluorouracil. *PLoS Biology*, 7(4), e91. doi:10.1371/journal.pbio.1000091
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. doi:10.1186/gb-2009-10-3-r25
- Lea, D. E., & Coulson, C. A. (1949). The distribution of the numbers of mutants in bacterial populations. *Journal of Genetics*, 49(3), 264–285.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078–2079. doi:10.1093/bioinformatics/btp352
- Maga, G., van Loon, B., Crespan, E., Villani, G., & Hübscher, U. (2009). The Block of DNA Polymerase Strand Displacement Activity by an Abasic Site Can Be Rescued by the Concerted Action of DNA Polymerase and Flap Endonuclease 1. *The Journal of Biological Chemistry*, 284(21), 14267–14275. doi:10.1074/jbc.M900759200
- Maiti, A., & Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *The Journal of Biological Chemistry*, 286(41), 35334–35338. doi:10.1074/jbc.C111.284620
- Matuo, R., Sousa, F. G., Escargueil, A. E., Soares, D. G., Grivicich, I., Saffi, J., et al. (2010). DNA repair pathways involved in repair of lesions induced by 5-fluorouracil and its active metabolite FdUMP. *Biochemical Pharmacology*, 79(2), 147–153. doi:10.1016/j.bcp.2009.08.016
- Maudrell, K. (1990). nmt1 of fission yeast. A highly transcribed gene completely repressed by thiamine. *The Journal of Biological Chemistry*, 265(19), 10857–10864.
- Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., Macdougall, E., Hendrich, B., et al. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science (New York, NY)*, 297(5580), 403–405. doi:10.1126/science.1073354
- Missero, C., Pirro, M. T., Simeone, S., Pischetola, M., & Di Lauro, R. (2001). The DNA glycosylase T:G mismatch-specific thymine DNA glycosylase represses thyroid transcription factor-1-activated transcription. *The Journal of Biological Chemistry*, 276(36), 33569–33575. doi:10.1074/jbc.M104963200
- Mojardín, L., Botet, J., Quintales, L., Moreno, S., & Salas, M. (2013). New Insights into the RNA-Based Mechanism of Action of the Anticancer Drug 5'-Fluorouracil in Eukaryotic Cells. *PloS One*, 8(11), e78172. doi:10.1371/journal.pone.0078172
- Mokkapat, S. K., Fernández de Henestrosa, A. R., & Bhagwat, A. S. (2001). Escherichia coli DNA glycosylase Mug: a growth-regulated enzyme required for mutation avoidance in stationary-phase cells. *Molecular Microbiology*, 41(5), 1101–1111.
- Moriya, M., Zhang, W., Johnson, F., & Grollman, A. P. (1994). Mutagenic potency of exocyclic

- DNA adducts: marked differences between *Escherichia coli* and simian kidney cells. *Proceedings of the National Academy of Sciences of the United States of America*, 91(25), 11899–11903.
- Muha, V., Horváth, A., Békési, A., Pukáncsik, M., Hodoscsek, B., Merényi, G., et al. (2012). Uracil-containing DNA in *Drosophila*: stability, stage-specific accumulation, and developmental involvement. *Plos Genetics*, 8(6), e1002738. doi:10.1371/journal.pgen.1002738
- Nair, K. R. (1940, August 26). Table of Confidence Interval for the Median in Samples From Any Continuous Population. *Sankhyā: the Indian Journal of Statistics (1933-1960)*. Retrieved August 26, 2014, from <http://www.jstor.org/stable/40383959>
- Nilsen, H., Haushalter, K. A., Robins, P., Barnes, D. E., Verdine, G. L., & Lindahl, T. (2001). Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase. *The EMBO Journal*, 20(15), 4278–4286. doi:10.1093/emboj/20.15.4278
- Nilsen, H., Rosewell, I., Robins, P., Skjelbred, C. F., Andersen, S., SLUPPHAUG, G., et al. (2000). Uracil-DNA glycosylase (UNG)-deficient mice reveal a primary role of the enzyme during DNA replication. *Molecular Cell*, 5(6), 1059–1065.
- Nilsen, H., Stamp, G., Andersen, S., Hrivnak, G., Krokan, H. E., Lindahl, T., & Barnes, D. E. (2003). Gene-targeted mice lacking the Ung uracil-DNA glycosylase develop B-cell lymphomas. *Oncogene*, 22(35), 5381–5386. doi:10.1038/sj.onc.1206860
- Oravcová, M., Teska, M., Půta, F., Folk, P., & Převorovský, M. (2013). Fission Yeast CSL Proteins Function as Transcription Factors. *PloS One*, 8(3), e59435. doi:10.1371/journal.pone.0059435.t003
- Otterlei, M., Warbrick, E., Nagelhus, T., Haug, T., Slupphaug, G., Akbari, M., et al. (1999). Post-replicative base excision repair in replication foci. *The EMBO Journal*, 18(13), 3834–3844. doi:10.1093/emboj/18.13.3834
- Pancaldi, V., Sarac, O. S., Rallis, C., McLean, J. R., Prevorovsky, M., Gould, K., et al. (2012). Predicting the Fission Yeast Protein Interaction Network. *G3&#58; Genes/Genomes/Genetics*, 2(4), 453–467. doi:10.1534/g3.111.001560
- Peters, G. J., Backus, H. H. J., Freemantle, S., van Triest, B., Codacci-Pisanelli, G., Van der Wilt, C. L., et al. (2002). Induction of thymidylate synthase as a 5-fluorouracil resistance mechanism. *Biochimica Et Biophysica Acta*, 1587(2-3), 194–205.
- Petronzelli, F., Riccio, A., Markham, G. D., Seeholzer, S. H., Genuardi, M., Karbowski, M., et al. (2000a). Investigation of the substrate spectrum of the human mismatch-specific DNA N-glycosylase MED1 (MBD4): fundamental role of the catalytic domain. *Journal of Cellular Physiology*, 185(3), 473–480. doi:10.1002/1097-4652(200012)185:3<473::AID-JCP19>3.0.CO;2-#
- Petronzelli, F., Riccio, A., Markham, G. D., Seeholzer, S. H., Stoerker, J., Genuardi, M., et al. (2000b). Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA N-glycosylase. *The Journal of Biological Chemistry*, 275(42), 32422–32429. doi:10.1074/jbc.M004535200
- Pettersen, H. S., Sundheim, O., Gilljam, K. M., SLUPPHAUG, G., Krokan, H. E., & KAVLI, B. (2007). Uracil-DNA glycosylases SMUG1 and UNG2 coordinate the initial steps of base excision repair by distinct mechanisms. *Nucleic Acids Research*, 35(12), 3879–3892. doi:10.1093/nar/gkm372
- Pettersen, H. S., Visnes, T., Vågbø, C. B., Svaasand, E. K., Doseeth, B., Slupphaug, G., et al. (2011). UNG-initiated base excision repair is the major repair route for 5-fluorouracil in DNA, but 5-fluorouracil cytotoxicity depends mainly on RNA incorporation. *Nucleic Acids Research*, 39(19), 8430–8444. doi:10.1093/nar/gkr563

- Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S. K., Eisen, D., et al. (2014). Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nature Chemical Biology*, *10*(7), 574–581. doi:10.1038/nchembio.1532
- Převorovský, M., Groušl, T., Staňurová, J., Ryneš, J., Nellen, W., Půta, F., & Folk, P. (2009). Cbf11 and Cbf12, the fission yeast CSL proteins, play opposing roles in cell adhesion and coordination of cell and nuclear division. *Experimental Cell Research*, *315*(8), 1533–1547. doi:10.1016/j.yexcr.2008.12.001
- Rada, C., Di Noia, J. M., & Neuberger, M. S. (2004). Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation. *Molecular Cell*, *16*(2), 163–171. doi:10.1016/j.molcel.2004.10.011
- Rada, C., Williams, G. T., Nilsen, H., Barnes, D. E., Lindahl, T., & Neuberger, M. S. (2002). Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Current Biology*, *12*(20), 1748–1755.
- Rai, K., Huggins, I. J., James, S. R., Karpf, A. R., Jones, D. A., & Cairns, B. R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell*, *135*(7), 1201–1212. doi:10.1016/j.cell.2008.11.042
- Rogstad, D. K., Liu, P., Burdzy, A., Lin, S. S., & Sowers, L. C. (2002). Endogenous DNA lesions can inhibit the binding of the AP-1 (c-Jun) transcription factor. *Biochemistry*, *41*(25), 8093–8102.
- Schar, P., & Kohli, J. (1993). Marker effects of G to C transversions on intragenic recombination and mismatch repair in *Schizosaccharomyces pombe*. *Genetics*, *133*(4), 825–835.
- Schuchert, P., & Kohli, J. (1988). The Ade6-M26 Mutation of *Schizosaccharomyces Pombe* Increases the Frequency of Crossing over. *Genetics*, *119*(3), 507–515.
- Seiple, L. (2006). Linking uracil base excision repair and 5-fluorouracil toxicity in yeast. *Nucleic Acids Research*, *34*(1), 140–151. doi:10.1093/nar/gkj430
- Shibata, E., Dar, A., & Dutta, A. (2014). CRL4Cdt2 E3 Ubiquitin Ligase and PCNA Cooperate to Degrade Thymine DNA Glycosylase in S-phase. *The Journal of Biological Chemistry*. doi:10.1074/jbc.M114.574210
- Shibutani, S. (1997). Translesional Synthesis on DNA Templates Containing a Single Abasic Site. A MECHANISTIC STUDY OF THE “A RULE.” *The Journal of Biological Chemistry*, *272*(21), 13916–13922. doi:10.1074/jbc.272.21.13916
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, NY)*, *324*(5929), 930–935. doi:10.1126/science.1170116
- Teytelman, L., Thurtle, D. M., Rine, J., & van Oudenaarden, A. (2013). Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci US A*, *110*(46), 18602–18607. doi:10.1073/pnas.1316064110
- Um, S., Harbers, M., Benecke, A., Pierrat, B., Losson, R., & Chambon, P. (1998). Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *The Journal of Biological Chemistry*, *273*(33), 20728–20736.
- Verri, A., Mazzarello, P., Biamonti, G., Spadari, S., & Focher, F. (1990). The specific binding of nuclear protein(s) to the cAMP responsive element (CRE) sequence (TGACGTCA) is reduced by the misincorporation of U and increased by the deamination of C. *Nucleic Acids Research*, *18*(19), 5775.
- Wach, A., Brachat, A., Pöhlmann, R. and Philippsen, P. (1994). New Heterologous Modules for Classical or PCR-Based Gene Disruptions in *Saccharomyces Cerevisiae*. *Yeast* *10* (13):

1793–1808.

- Wang, Z., & Mosbaugh, D. W. (1989). Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. *The Journal of Biological Chemistry*, 264, 1163–1171.
- Waters, T. R., & Swann, P. F. (1998). Kinetics of the Action of Thymine DNA Glycosylase. *The Journal of Biological Chemistry*, 273(32), 20007–20014. doi:10.1074/jbc.273.32.20007
- Waters, T. R., Gallinari, P., Jiricny, J., & Swann, P. F. (1999). Human Thymine DNA Glycosylase Binds to Apurinic Sites in DNA but Is Displaced by Human Apurinic Endonuclease 1. *The Journal of Biological Chemistry*, 274(1), 67–74. doi:10.1074/jbc.274.1.67
- Wilkinson, C. R. M., Bartlett, R., Nurse, P., & Bird, A. P. (1995). The fission yeast gene pmt1+ encodes a DNA methyltransferase homologue. *Nucleic Acids Research*, 23(2), 203.
- Wood, V., Gwilliam, R., Rajandream, M. A., Lyne, M., Lyne, R., Stewart, A., et al. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature*, 415(6874), 871–880. doi:10.1038/nature724
- Woolcock, K. J., Stunnenberg, R., Gaidatzis, D., Hotz, H.-R., Emmerth, S., Barraud, P., & Bühler, M. (2012). RNAi keeps Atf1-bound stress response genes in check at nuclear pores. *Genes and Development*, 26(7), 683–692. doi:10.1101/gad.186866.112
- Zamir, L., Zaretsky, M., Fridman, Y., Ner-Gaon, H., Rubin, E., & Aharoni, A. (2012). Tight coevolution of proliferating cell nuclear antigen (PCNA)-partner interaction networks in fungi leads to interspecies network incompatibility. *Proc Natl Acad Sci US A*, 109(7), E406–E414. doi:10.1073/pnas.1108633109

### Figure legends

**Figure 1. Ung1 represents the main uracil excision activity in cell free extracts. (A)** DNA nicking assay. The 60 bp double-stranded DNA substrate containing an A•U base pair or a G•U mismatch is labeled with a 5' fluorescein (\*). Uracil excision results in a labeled 23-mer upon AP-site hydrolysis. 2 pmol of substrate were incubated with 2 pmol of recombinant Thp1 or 7 µg of cell-free extracts of the indicated strains. When indicated, 1 unit of uracil-DNA glycosylase inhibitor (Ugi) was added. For Thp1 overexpression, cells transformed with pPRS271 (*thp1<sup>+</sup>* controlled by the inducible *nmt1* promoter) or with the empty vector pREP1 were induced for 16 h. **(B)** Thp1 overexpression. Crude protein extracts of the indicated strains induced as in (A) were prepared and analyzed by Western blot using an affinity purified rabbit polyclonal anti-Thp1 antiserum. Size of Thp1 (arrow), unspecific bands (\*) and loaded amounts of protein extracts are indicated.

**Figure 2. Thp1 and Ung1 cooperate in uracil removal to prevent spontaneous mutations. (A)** Subcellular localization of Thp1 in wild-type cells. Cells expressing a C-terminally tagged, endogenous Thp1-EGFP fusion protein or the untagged Thp1 were arrested by glucose starvation for 16 h before imaging. **(B)** Uracil accumulation in *thp1Δung1Δ* strains. Chromosomal DNA was prepared in agarose plugs and treated with 13 units of *E. coli* Ung and 3.5 pmol of human AP-endonuclease (APE1) as indicated. DNA was separated by PFGE and stained with ethidium bromide. Intact *S. pombe* chromosomes (Chr) I, II and III and fragmented DNA (shaded bar) are indicated. **(C)** Spontaneous reversion rates and mutation spectra at the *ade6-M387* locus. Any base substitution at the *ade6-M387* locus lead to adenine prototrophy. Prototrophs were scored after 9 to 11 days of growth on selective medium and mutation rates were calculated from at least 24 independent cultures out of 3 experiments. Shown are the median and the 95% confidential interval. Transversion and transition rates were determined by DNA sequencing of the *ade6-M387* locus in one random clone of each culture from experiments shown in the graph. Significance of the transition rate in *thp1Δung1Δ* was confirmed by chi-square test. **(D)** Spontaneous forward mutation rates to canavanine resistance. Mutations at the arginine permease gene *can1<sup>+</sup>* abolish the uptake of L-canavanine, thereby allowing growth in presence of this toxic arginine analog. Resistant clones were scored after 14 to 16 days on EMM plates containing L-canavanine.

Mutation rates were determined using at least 30 independent cultures out of 3 experiments. Shown are the median and the 95% confidential interval.

**Figure 3. Thp1 deletion increases resistance to 5-FU and ectopic AID expression. (A)** 5-FU leads to AP-site accumulation in wild-type cells. Chromosomal DNA was prepared from wild-type cells grown for 48 h in YEL supplemented with 10 mg/l 5-FU. DNA was treated with 13 units *E. coli* Ung enzyme and 3.5 pmol recombinant APE1 as indicated. The digested DNA was separated by PFGE and stained with ethidium bromide. Fragmented DNA forms a smear (shaded bar) below intact chromosomes (Chr). **(B)** Cellular sensitivity to 5-FU. Serial dilutions of wild-type and mutant cells were spotted on MMA plates containing 5-FU at the indicated concentrations and incubated for 6 to 12 days at 26°C. **(C)** Effect of Thp1 overexpression on 5-FU sensitivity. Serial dilutions of WT and *thp1Δung1Δ* double mutant cells transformed with either control (pREP1) or Thp1 expression plasmids (PRS271) were spotted on plain EMM medium (induced expression) or EMM medium containing 5 μg/ml thiamine (repressed expression). 0.5 mg/l 5-FU was added when indicated. Cells were incubated for 8 to 12 days at 26°C. **(D)** 5-FU-induced forward mutations to canavanine resistance. Yeast cultures were divided in halves and incubated for 76 h at 30°C in the absence or presence of 10 mg/l 5-FU. Spontaneous and induced mutation rates were determined from 3 to 4 cultures per strain. Inducibility of mutations is given in the white boxes. **(E)** Expression of human AID in *S. pombe* cells. The human AID gene was cloned into the pREP1 vector carrying the inducible *nmt1* promoter (pREP1-AID). Cells were grown under inducing (+) or repressing (-) conditions for 16 h. Crude protein extracts (50 μg) were separated on SDS-PAGE and subjected to Western blot analysis using mouse anti-AID antiserum. The arrow indicates the AID-specific band. **(F)** Cytotoxicity of 5-FU treatment. Serial dilutions of cells transformed with pREP1 or pREP1-AID were spotted onto EMM medium (inducible condition) and incubated for 7 days at 30°C.

**Figure 4. Thp1 mediates spontaneous and X-ray-induced recombination. (A)** Construct for detection of intra-chromosomal recombination. Recombination substrate consisting of direct repeats of the *ade6-L469* and *ade6-M375* alleles, separated by a functional *ura4<sup>+</sup>* gene. Two types of gene conversion events are scored in this assay, deletion and reversion events, both conferring adenine prototrophy. **(B)** Rates of spontaneous mitotic

recombination. Recombination was scored by adenine prototrophy and recombination rates were calculated using at least 24 cultures per strain. The median and 95% confidential interval are shown. **(C)** X-ray induced mitotic recombination. Cells were exposed to a non-lethal X-ray dose (100 Gy) and recombination rates were determined as in (B). The median and 95% confidential interval are shown.

**Figure 5. Gene expression in WT and *thp1Δ* cells.** **(A)** Differential mRNA expression in WT and *thp1Δ*. The mean  $\log_2$  expression value from wild-type triplicates was correlated with that of *thp1Δ* triplicates. The diagonal (red line), the Spearman correlation coefficient ( $r$ ) and the point referring to the *thp1<sup>+</sup>* (arrow) gene are indicated. **(B)** Up- and down-regulation of gene expression in *thp1Δ* cells. mRNA expression values of *thp1Δ* were normalized to the wild-type ( $\log_2$  fold change,  $\log_2FC$ ) and the value corresponding to the *thp1<sup>+</sup>* gene was omitted. The red line represents the median value. **(C)** A threshold of  $|\log_2| > 0.3$  was applied to the normalized *thp1Δ* levels (N=471) and mRNAs were divided into up- and down-regulated genes. **(D)** Principal component analysis of single mRNA expression values (N=5022) of the wild-type and *thp1Δ* triplicates (A, B and C). **(E)** Comparison of standard deviations (SD) from wild-type and *thp1Δ* triplicates. SDs were calculated for each mRNA and *thp1Δ* SDs were plotted against WT SDs. The diagonal (red) and the Spearman correlation coefficient ( $r$ ) are indicated.

**Supplementary Figure 1 (related to Figure 5A-C). Expression of specific element classes in WT and *thp1Δ* cells.** 5' UTR, 5UTR; 3' UTR, 3UTR; long non-coding RNA, lncRNA. **(A)** Differential expression in WT versus *thp1Δ* cells. The diagonal (red) and the Spearman correlation coefficient ( $r$ ) are indicated **(B)** Up- and down-regulation of gene expression in *thp1Δ* cells. 5UTR, 3UTR and lncRNA expression values of *thp1Δ* were normalized to the wild-type ( $\log_2$  fold change,  $\log_2FC$ ) and the value corresponding to the *thp1<sup>+</sup>* gene was omitted. Red line, median value. **(C)** The  $\log_2FC$  was calculated for all mRNAs that were sorted as up- and down-regulated.

**Supplementary Figure 2 (related to Figure 5D to 5E). Variation between biological triplicates for 5UTRs, 3UTRs and lncRNAs.** **(A)** Principal component analysis of the six samples (3 wild-type and 3 *thp1Δ* samples). **(B)** Comparison of SDs from wild-type and *thp1Δ*

triplicates. SDs were calculated for 5UTR, 3UTR or lncRNA elements and *thp1Δ* SDs were plotted against WT SDs. The diagonal (red) and the Spearman correlation coefficient ( $r$ ) are indicated.

**Supplementary Figure 3 (related to Figure 5E). SD comparison of mRNAs between WT and *thp1Δ* samples.** All mRNAs were divided into relative enrichment above and below a  $\log_2FC$  of -0.3. WT SDs were correlated with *thp1Δ* SDs. The diagonal (red) and the Spearman correlation coefficient ( $r$ ) are indicated.

**Supplementary Figure 4. Activity of the Thp1-13myc.** Thp1-13 myc was overexpressed in cells from the pREP1-Thp1myc plasmid. The *nmt1* promoter was induced for 15 h and extracts were compared to those of cells carrying the pREP1 control plasmid. Base release assays were performed using a G•U mismatch and Ugi was added as indicated.

**Supplementary Figure 5. Thp1-13myc ChIP-seq.** ChIP-seq was performed in duplicates and reads were mapped to 500 base pair windows across the entire genome. **(A)** Correlation between input and IP samples for the duplicates (A and B) is shown in the left and middle panel. Correlation between the mean  $\log_2FC$  of the input and IP samples is shown in the right panel. The Spearman coefficient ( $r$ ) is indicated. **(B)** As in A, but the values were corrected for duplicates deriving from PCR amplification during the library preparation.

**Supplementary Figure 6. Table containing all Thp1-13myc associated sites with a  $\log_2FC$  above 0.3 in both duplicates.** The genomic regions of Thp1-13myc enrichment are shown. tRNAs are in grey. Chr, chromosome. Cen, centromere.



Figure 1

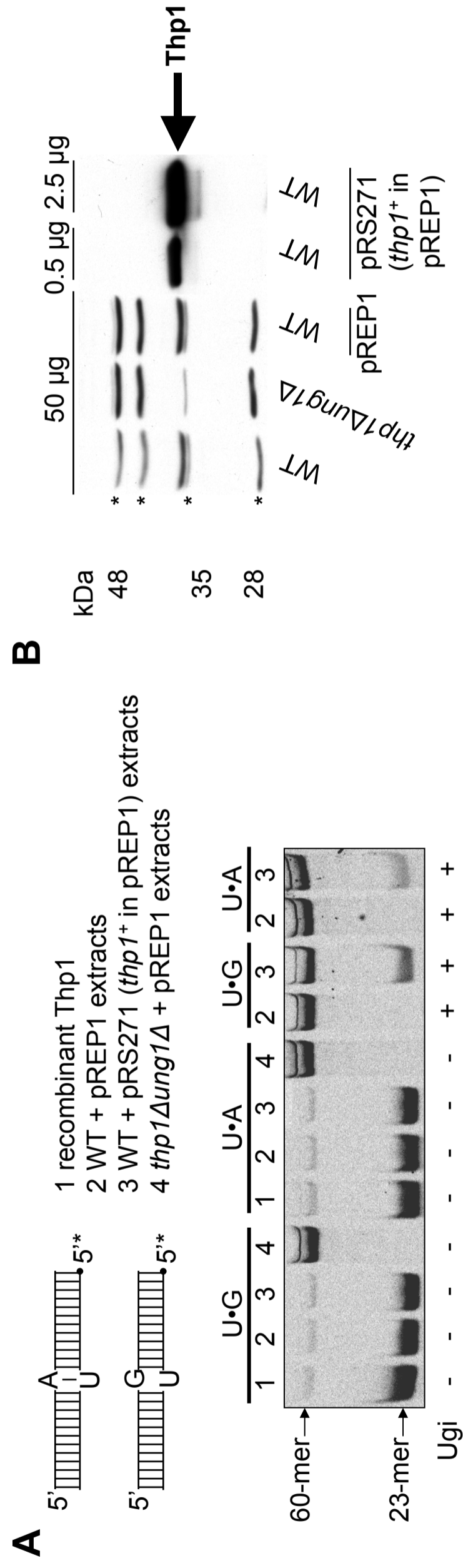
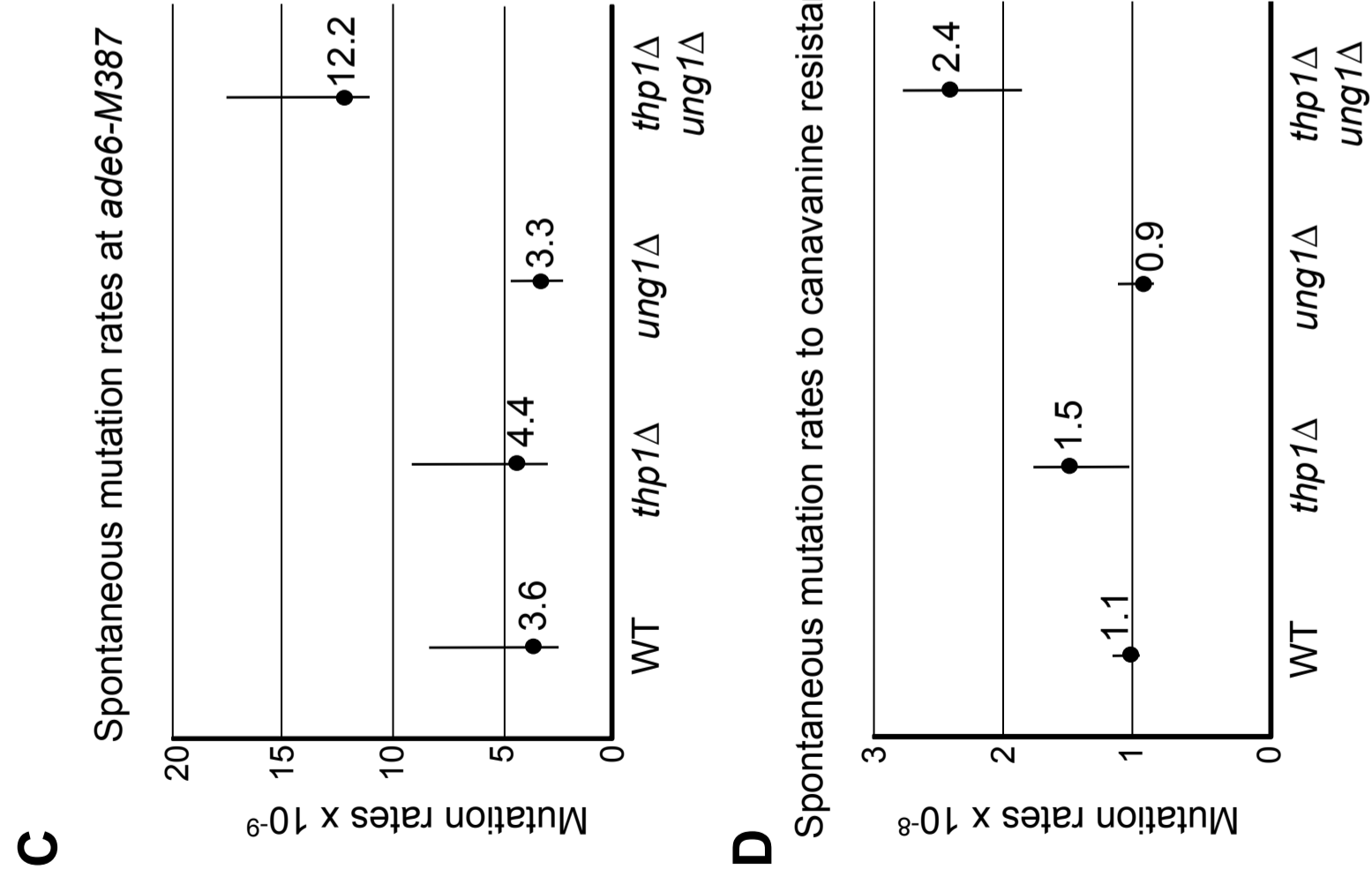
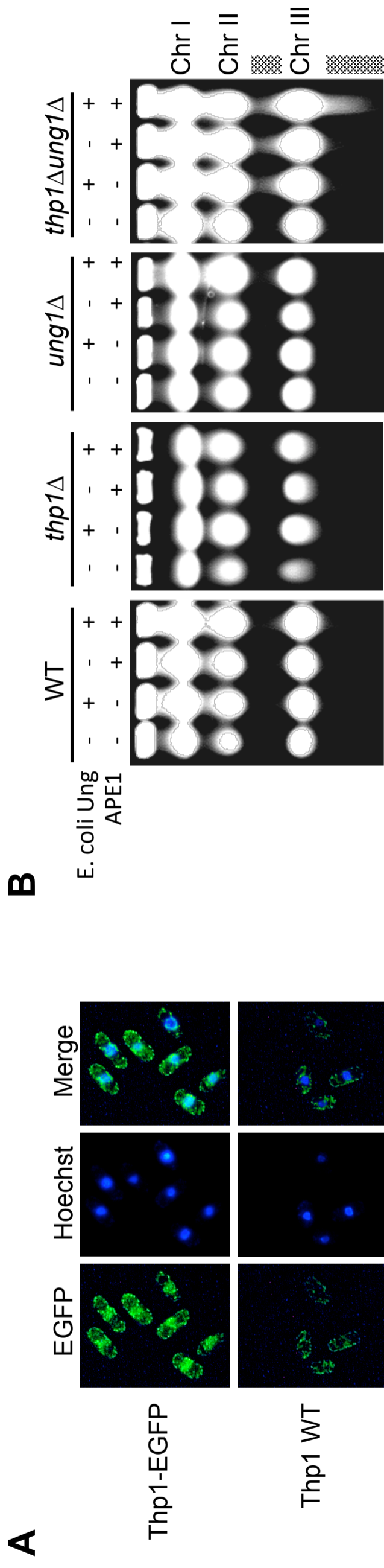


Figure 2

Mutation spectra at *ade6-M387*

Strain	Transversions (C→A, C→G)	Rate ( $10^{-9}$ )	Transitions (C→T)	Rate ( $10^{-9}$ )
WT	13/26	1.8	10/26	1.4
<i>thp1</i> $\Delta$	14/21	2.9	7/21	1.5
<i>ung1</i> $\Delta$	13/27	1.6	14/27	1.7
<i>thp1</i> $\Delta$ <i>ung1</i> $\Delta$	7/29	2.9	21/29	8.8*

Figure 3

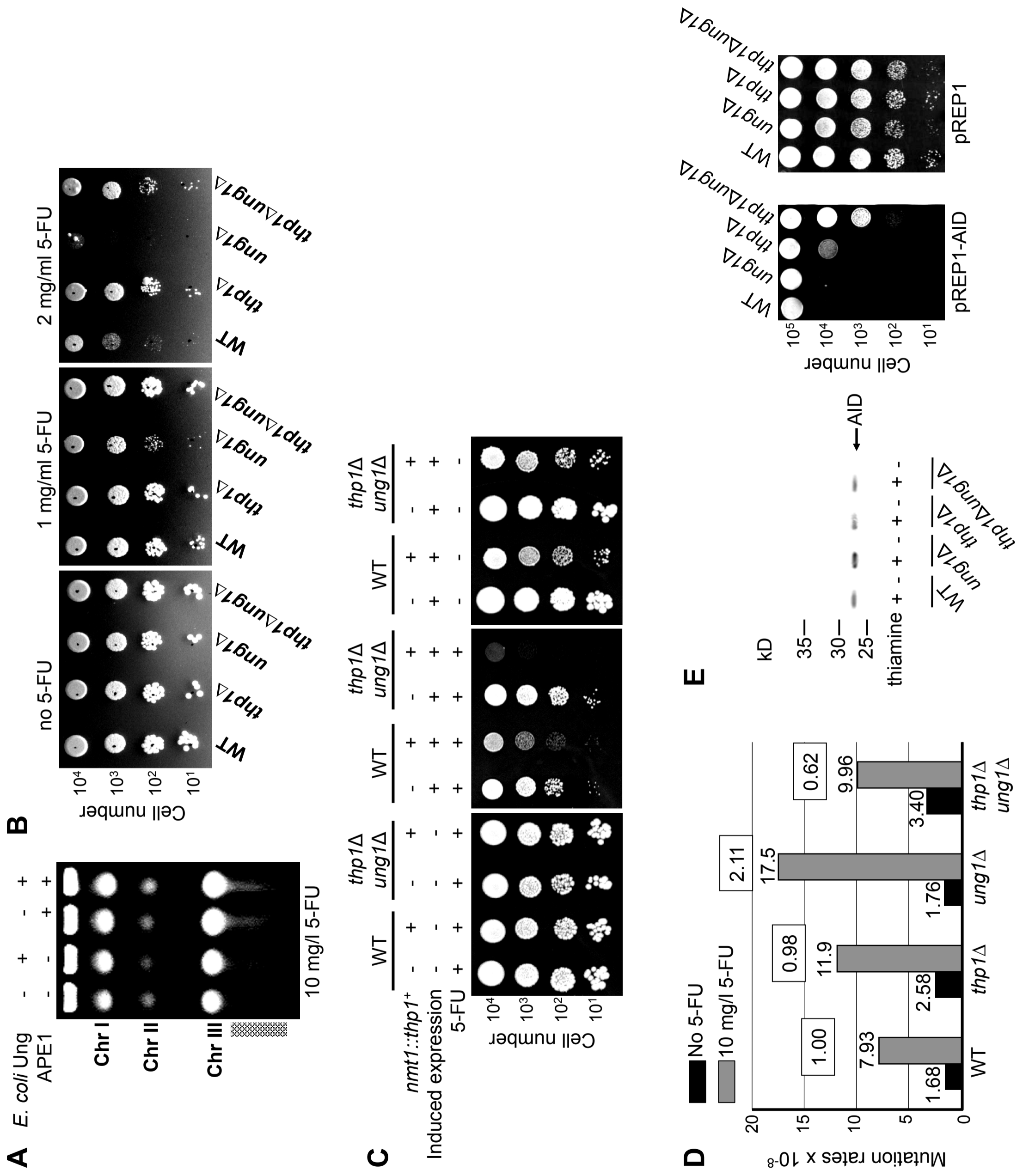


Figure 4

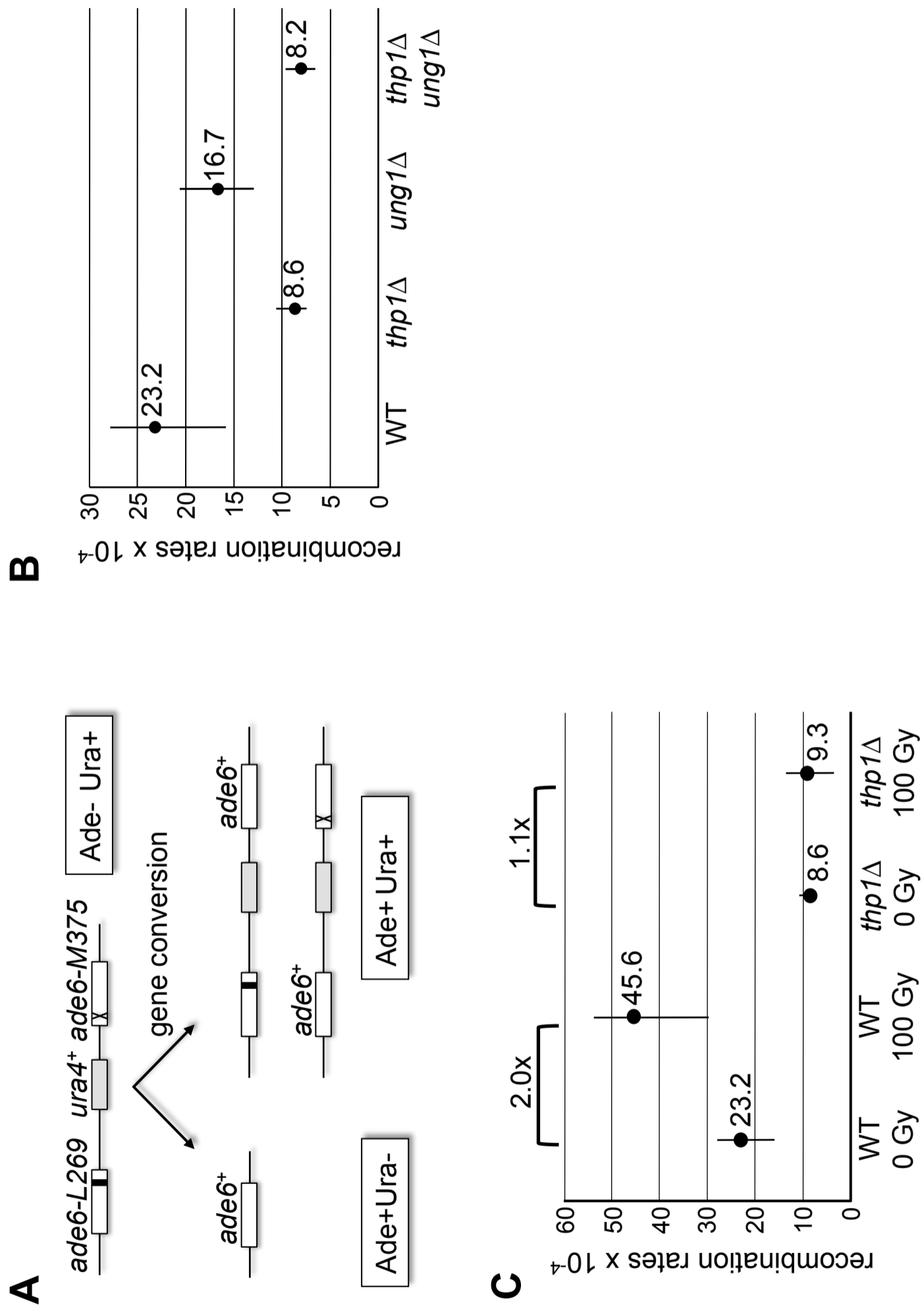
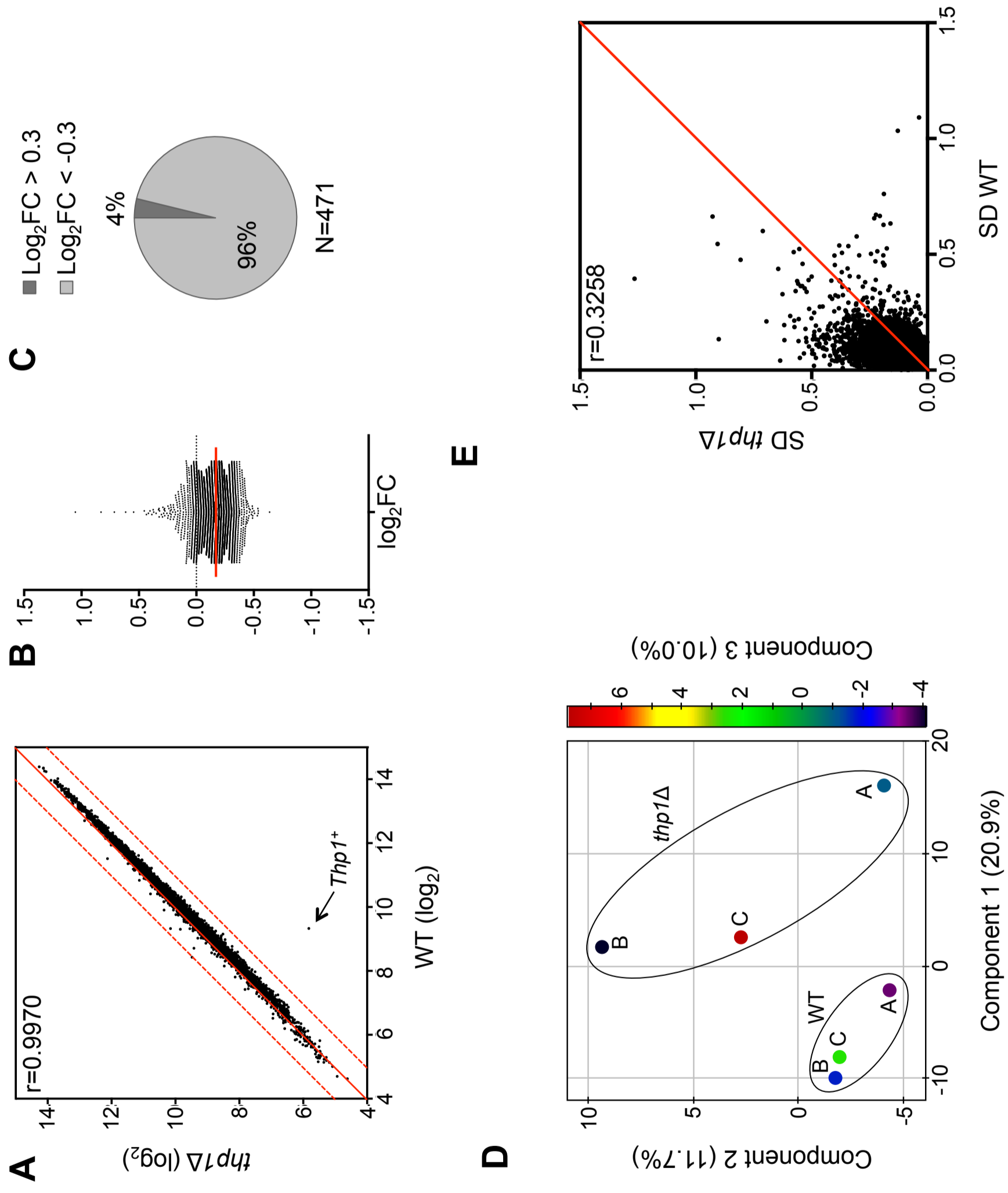
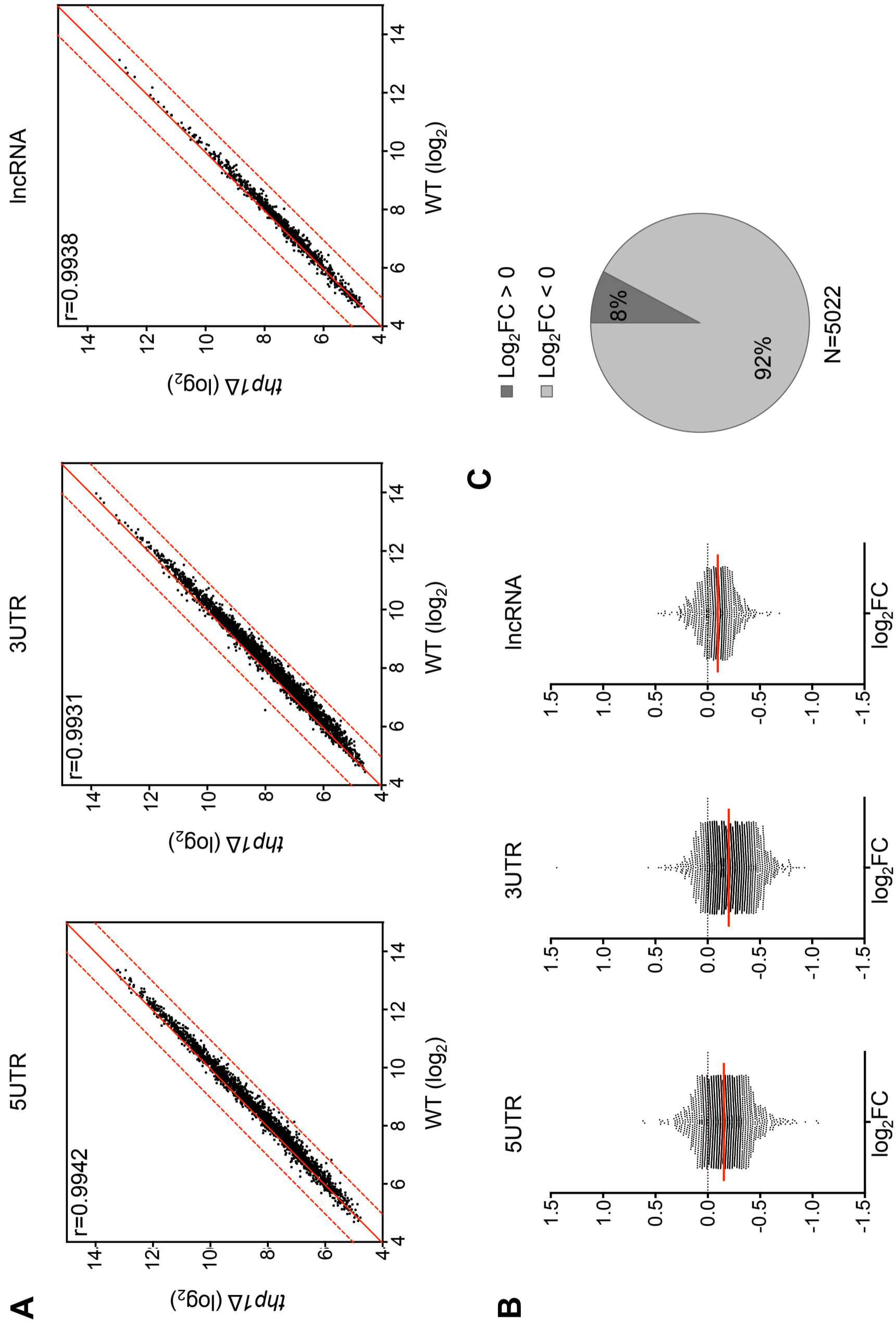


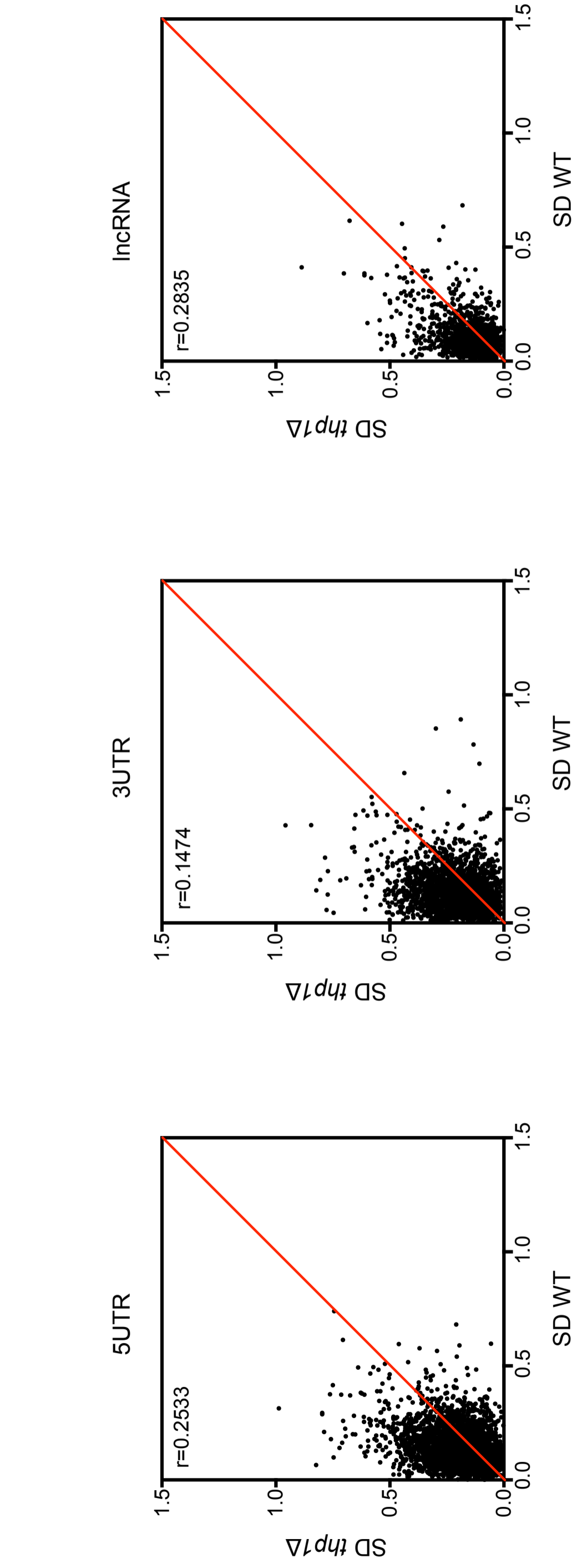
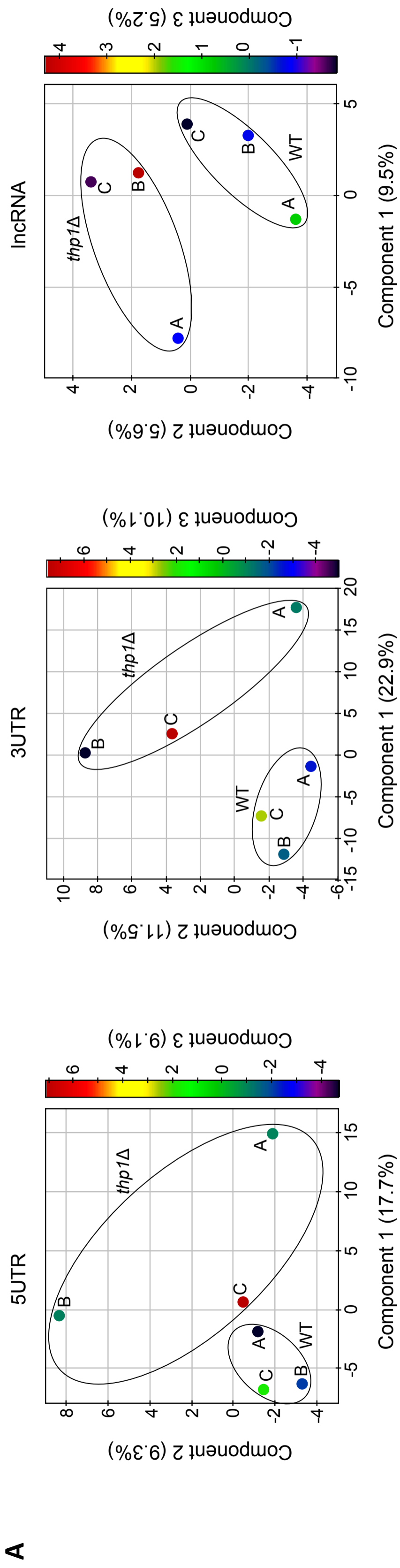
Figure 5



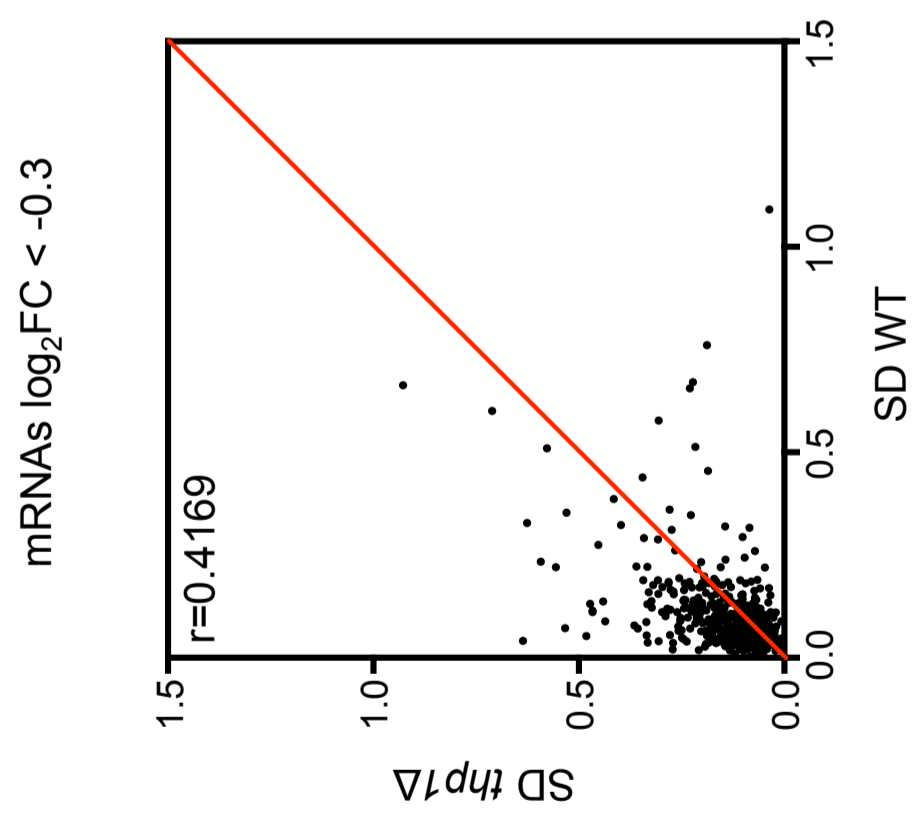
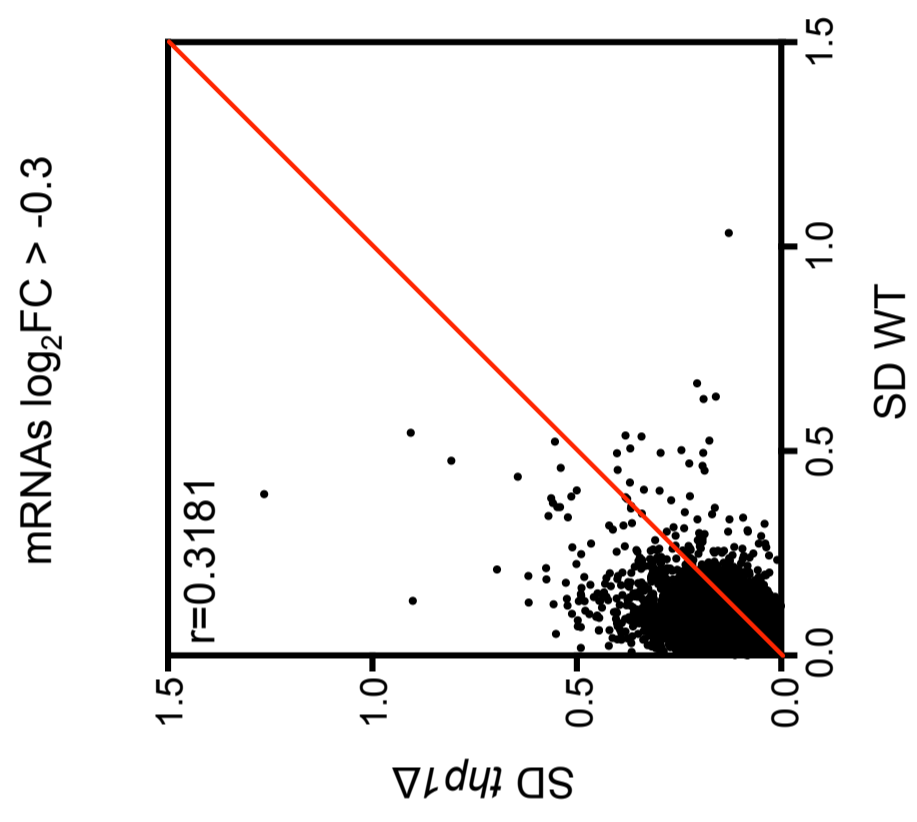
Supplementary Figure 1



## Supplementary Figure 2

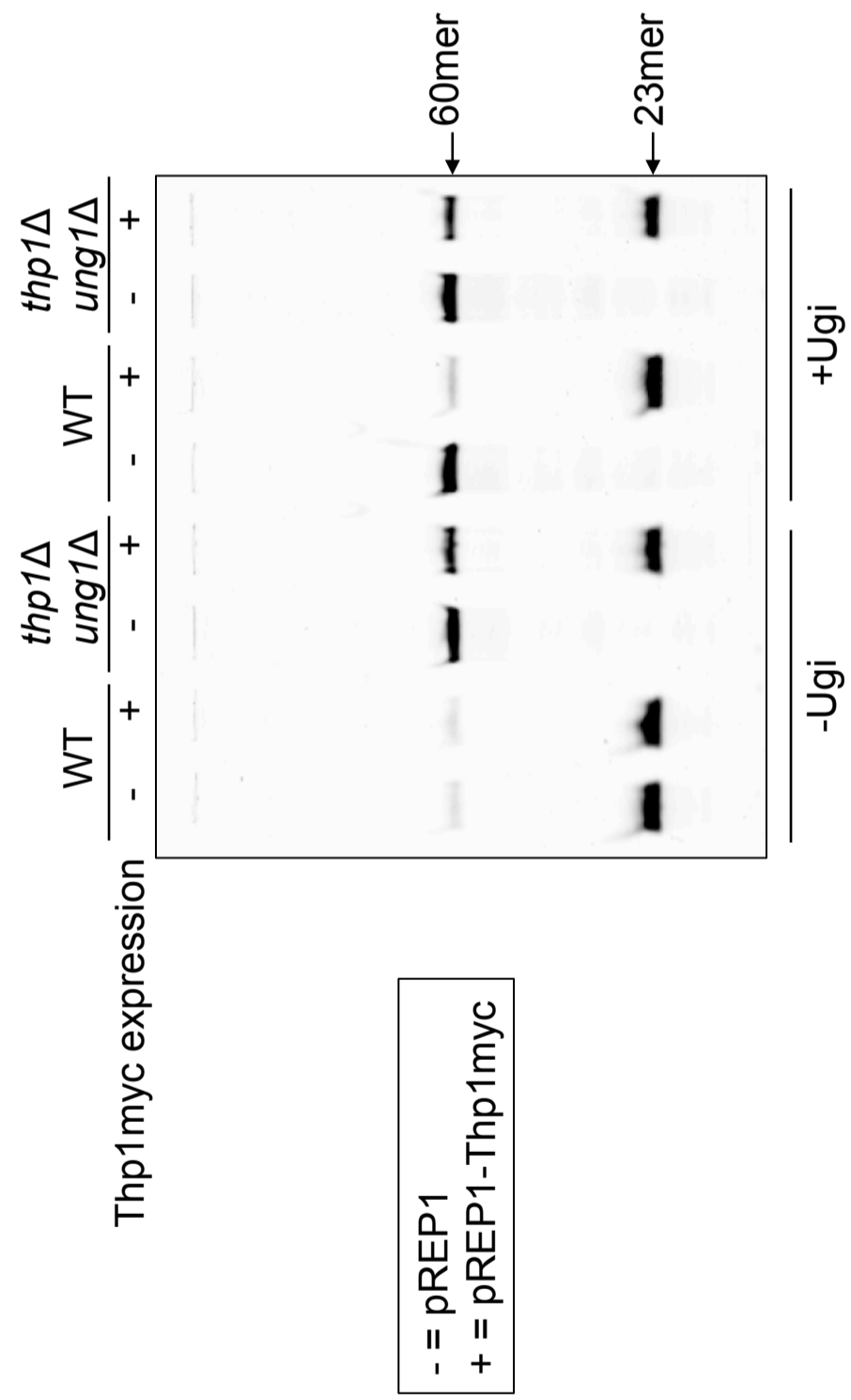


## Supplementary Figure 3

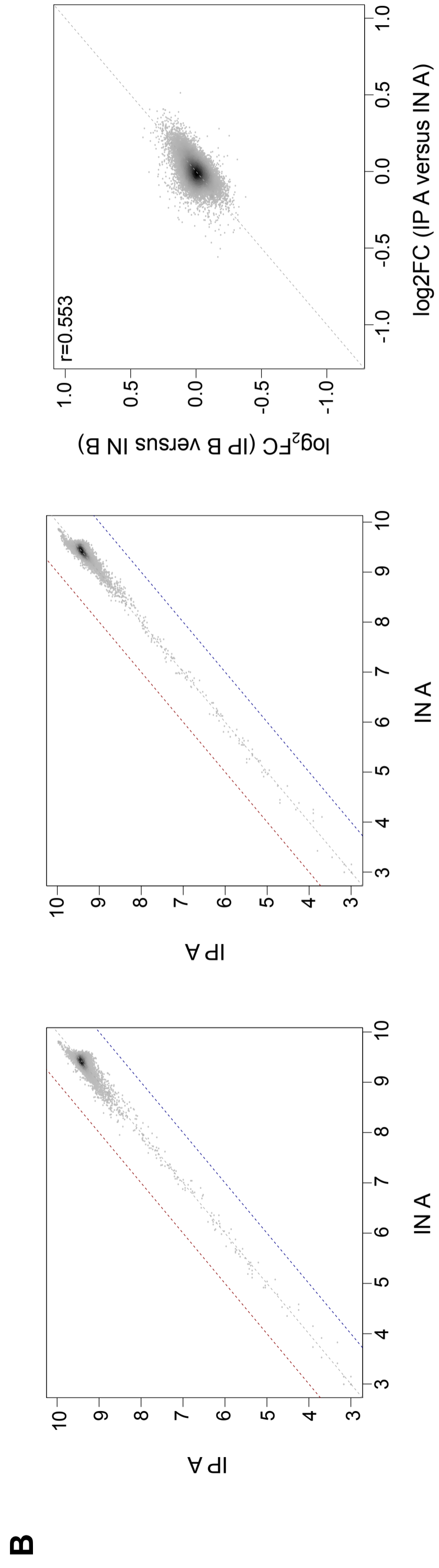
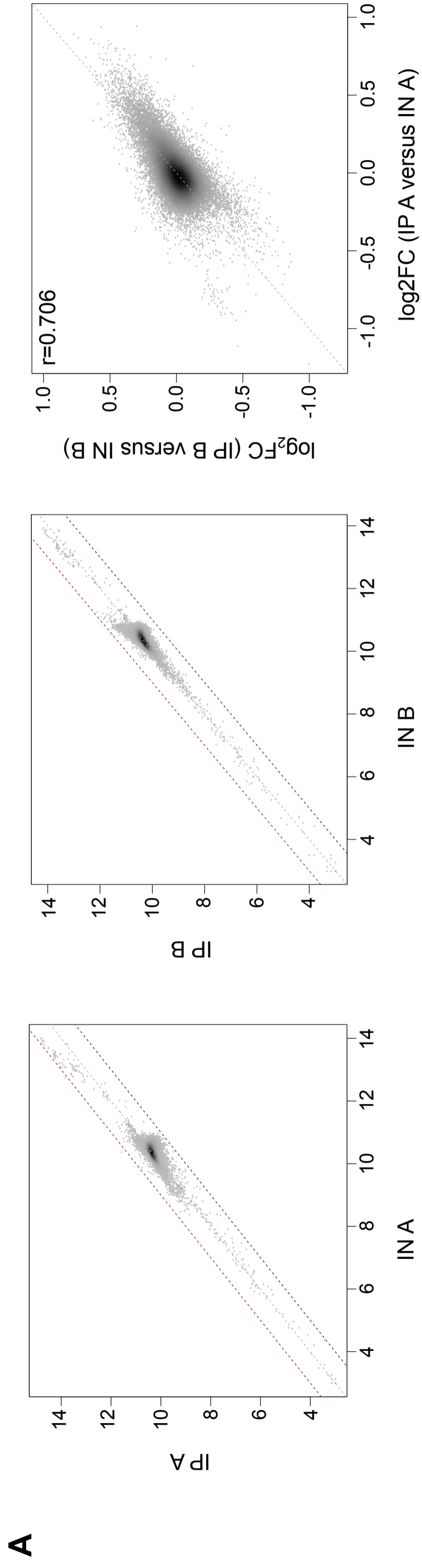




## Supplementary Figure 4



Supplementary Figure 5



## Supplementary Figure 6

Chr	Coordinates	Window width	Mean enrichment over window	Nearest gene (ID)	Nearest gene (description)	Located in centromer
1	339001 - 339500	500	0.626	SPAC12G12.05c	SAGA complex/ transcription initiation factor Taf9	no
2	527001 - 528000	1000	0.707	SPNCRNA.659 SPAC227.16c SPSNORNA.03 SPSNORNA.04 SPSNORNA.05 SPSNORNA.06	antisense RNA (predicted) GINS complex subunit Psf3 small nucleolar RNA Z7 small nucleolar RNA Z5 small nucleolar RNA Z4 small nucleolar RNA Z3	no
3	3765001 - 3765500	500	0.592	SPATRNAALA.04 SPATRNAGLU.03	tRNA Alanine tRNA Glutamic acid	yes
4	3776501 - 3777000	500	0.600	SPATRNAGLU.04 SPNCRNA.396	tRNA Glutamic acid non-coding RNA (predicted)	yes
5	4008001 - 4009500	1500	0.710	SPAC27E2.11c SPNCRNA.53 SPAC27E2.13	sequence orphan non-coding RNA, poly(A)-bearing (predicted) dubious	no
6	4433501 - 4434000	500	0.536	SPAC16.05c	transcription factor Sfp1 (predicted)	no
7	520001 - 520500	500	0.607	SPNCRNA.1366 SPRRNA.29	intergenic RNA (predicted) 5S rRNA	no
8	1308001 - 1309000	1000	0.934	SPSNORNA.21	small nucleolar RNA U14	no
9	1600001 - 1600500	500	0.569	SPBTRNAALA.08 SPBTRNAILE.05	tRNA Alanine tRNA Isoleucine	close to
10	1602001 - 1602500	500	0.647	SPBTRNAARG.06 SPBTRNAASP.05	tRNA arginine tRNA Asparagine	close to
11	1618001 - 1619000	1000	0.556	SPBTRNAALA.09 SPBTRNAILE.06	tRNA Alanine tRNA Isoleucine	yes
12	1629501 - 1630000	500	0.570	SPBTRNAALA.10	tRNA Alanine	yes
13	4037001 - 4037500	500	0.691	SPBC215.05	glycerol-3-phosphate dehydrogenase Gpd1	no
14	4442501 - 4443000	500	0.677	SPBC8E4.02c	sequence orphan	no
15	4446501 - 4447000	500	0.627	SPBP4G3.02	acid phosphatase Pho1	no
16	275001 - 275500	500	0.552	SPCC794.12c	malic enzyme, malate dehydrogenase (oxaloacetate decarboxylating), Mae2	no
17	1092501 - 1093000	500	0.596	SPCTRNAARG.12 SPCTRNAASP.06 SPCTRNAVAL.09	tRNA Arginine tRNA Asparagine tRNA Valine	yes
18	1096001 - 1096500	500	0.552	SPCTRNALEU.12	tRNA Leucine	yes
19	1106001 - 1106500	500	0.591	SPCTRNAARG.13 SPCTRNAASP.07 SPCTRNAVAL.10	tRNA Arginine tRNA Asparagine tRNA Valine	yes
20	1392501 - 1393000	500	0.552	SPCC1281.06c	acyl-coA desaturase (predicted)	no
21	1394001 - 1394500	500	0.626	SPCC1281.06c	"	no
22	1452001 - 1452500	500	0.706	SPCC11E10.01 SPNCRNA.1197	cystathionine beta-lyase (predicted) intergenic RNA (predicted)	no
23	2059001 - 2059500	500	0.556	SPCC1739.13	heat shock protein Ssa2	no

CEN 1: I:3753687-3789421

CEN 2: II:1602264-1644747

CEN 3: III:1070904-1137003

Table 1

Strain	Genotype	Figure	Source
972 h <sup>-</sup>	wild-type h <sup>-</sup>	1,2,3	(Leupold, 1954)
975 h <sup>+</sup>	wild-type h <sup>+</sup>	5, Supplementary 1-3	(Leupold, 1954)
PRS 000g	<i>leu1-32 h<sup>+</sup></i>	1,3,4	lab stock
PRS 801	<i>thp1::kanMX4 ung1::ura4 leu1-32 ura4-D18 h<sup>-</sup></i>	1,3,4, Supplementary 4	PRS000g x PRS571
PRS 571	<i>thp1::kanMX4 ung1::ura4 ura4-D18 h<sup>-</sup></i>	1,2,3	PRS556 x PRS603
PRS 827	<i>thp1<sup>+</sup>-EGFP h<sup>-</sup></i>	2	PRS823 x PRS556
PRS 555	<i>thp1::kanMX4 h<sup>-</sup></i>	2,3,5, Supplementary 1-3	975h <sup>+</sup> x PRS550
PRS 603	<i>ung1::ura4 ura4-D18 h<sup>-</sup></i>	2,3	this study
PRS 053	<i>ade6-M387 ura4-D18 h<sup>-</sup></i>	2	lab stock
PRS 563	<i>thp1::kanMX4 ade6-M387 ura4-D18 h<sup>-</sup></i>	2	this study
PRS 605	<i>ung1::ura4 ade6-M387 ura4-D18 h<sup>-</sup></i>	2	this study
PRS 575	<i>thp1::kanMX4 ung1::ura4 ade6-M387 ura4-D18 h<sup>-</sup></i>	2	this study
PRS 807	<i>ura4-D18 ade6-M375 h<sup>-</sup> int::pUC8/ura4/ade6-L469</i>	4	this study
PRS 809	<i>thp1::kanMX4 ura4-D18 ade6-M375 h<sup>-</sup> int::pUC8/ura4/ade6-L469</i>	4	this study
PRS 811	<i>ung1::ura4 ura4-D18 ade6-M375 h<sup>-</sup> int::pUC8/ura4/ade6-L469</i>	4	this study
PRS 813	<i>thp1::kanMX4 ung1::ura4 ura4-D18 ade6-M375 h<sup>-</sup> int::pUC8/ura4/ade6-L469</i>	4	this study
PRS 823	<i>ura4-D18 thp1<sup>+</sup>::3'-egfp<sup>+</sup> h<sup>-</sup></i>	2	this study
LV10	<i>rec12::ura4 ura4-D18 leu2-112 ade6-M26 h<sup>+</sup></i>	3	this study
LV22	<i>thp1::kanMX4 rec12::ura4 ura4-D18 leu2-112 ade6-M26 h<sup>-</sup></i>	3	this study
LV40	<i>thp1::kanMX4 ung1::ura4 rec12::ura4 ura4-D18 leu2-112 ade6-469 h<sup>+</sup></i>	3	this study
LV41	<i>ung1::ura4 rec12::ura4 ura4-D18 leu2-112 ade6-M26 h<sup>+</sup></i>	3	this study
PRS000f	<i>leu1-32 h<sup>-</sup></i>	Supplementary 4	lab stock
SP Thp1myc	<i>thp1+::13myc-kanMX3 h<sup>-</sup></i>	Supplementary 5-7	this study

## **Estrogen Receptor $\beta$ Regulates Epigenetic Patterns at Specific Genomic Loci through interaction with Thymine DNA Glycosylase**

**William Duong<sup>1</sup>, Claudia Krawczyk<sup>1</sup>, Nancy Bretschneider<sup>2</sup>, Christian Zinsler<sup>2</sup>, Primo Schär<sup>1</sup>, Joëlle Rüegg<sup>1,3</sup>§**

<sup>1</sup>Department of Biomedicine, Institute of Biochemistry and Genetics, University of Basel, Basel, Switzerland.

<sup>2</sup> Genomatix Software GmbH, Bayerstr. 85a, 80335 Munich, Germany.

<sup>3</sup> Department of Clinical Neurosciences, Karolinska Institutet, Stockholm, Sweden.

§Corresponding author, to whom reprint requests should be sent: Karolinska Institutet, CMM L8:00, 171 76 Stockholm, Sweden. Phone: +46-8-41776166; E-mail: joelle.ruegg@ki.se

**Running title:** ER $\beta$  regulates DNA Methylation at specific genomic loci

**Keywords:** Estrogen receptor  $\beta$ , Thymine DNA glycosylase, DNA methylation, Epigenetic regulation

## INTRODUCTION

DNA methylation and histone modifications are ways to encode epigenetic information and play a crucial role in regulating gene expression during embryonic development (1,2). Aberrant epigenetic patterns are found in various human diseases, including cancer, obesity, and psychiatric disorders (3).

The main epigenetic DNA modification is methylation at the fifth position of cytosine (5mC). The DNA methylation pattern is established and maintained by DNA methyltransferases (DNMTs) that transfer methyl groups from S-Adenosyl methionine (SAM) to cytosines, mainly at CpG dinucleotides (4). DNA methylation has long been considered as fairly stable and only removable by passive demethylation, i.e. by reduced DNMT activity during cell division. More recently, it has become clear that active DNA demethylation occurs during embryonic development, in primordial germ cells (5), and cell differentiation (6). Active demethylation is initiated by the oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) by TET (ten eleven translocation) proteins, a family of Fe(II)- and 2-oxoglutarate-dependent DNA dioxygenases (7). Genome-wide mapping of this mark revealed that 5hmC is mostly found in pluripotent cells and neurons, in bodies of transcribed genes and in gene-regulatory regions (promoters and transcriptional enhancers) (8) concomitant with the bivalent chromatin mark H3K4m3/H3K27m3 (9). Such regions are poised for activation or permanent silencing during lineage commitment and terminal cell differentiation (10). 5hmC can be further processed to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by the TETs. These modifications are recognised and excised by the DNA glycosylases Thymine DNA glycosylase (TDG) (11) and replaced by an unmethylated cytosine by the base excision repair (BER) (12). Evidence for active DNA demethylation by this mechanism stems from the findings that TDG deficiency is embryonic lethal in mice (13,14) and leads to changes in the distribution of cytosine modifications during stem cell differentiation (13,15,16), in particular in gene regulatory regions such as promoters and enhancers. Further, 5fC and 5caC accumulate in the absence of TDG in embryonic stem cells at promoter and enhancer regions (15,16).

An open question is how factors involved in regulation of DNA modifications are targeted to specific genomic loci. It has been suggested that locus specific *de novo* methylation is induced by proximal sequence elements coding for specific transcription factors (17-19) or by recruitment of DNMTs by non-coding RNAs (20-22). Nevertheless, the exact mechanism how DNA methylation is regulated at specific genomic regions is still not well understood.

Nuclear receptors (NRs) are inducible transcription factors that have been suggested to regulate epigenetic events, particularly histone modifications (23) but also DNA methylation (24-29). In our work (24), we could demonstrate that in cells lacking the NR estrogen receptor beta ( $ER\beta$ ), a single CpG in the promoter region of Glucose transporter 4 (Glut4) is hypermethylated. This hypermethylation correlated with changes in expression and inducibility of Glut4. Furthermore, we could recently show that expression of the dyslexia candidate gene *DYX1C1* is regulated by both  $ER\beta$  and DNA methylation at the same region (30). Thus  $ER\beta$  could be a transcription factor that is involved in the local regulation of DNA methylation.

$ER\beta$  is one of the two ER isoforms that mediate the physiological effects of estrogens, the female sex hormones. It is involved in development and functioning of the reproductive organs, but also of other tissues, e.g. the brain (31) and adipose tissue (32). It is mostly found in the cell nucleus where it, upon activation, binds to regulatory elements (estrogen response elements, EREs) on target genes. There are a number of co-activators that enhance ER transcriptional activity, including chromatin remodelling factors (33). The ERs are not only activated by endogenous hormones, but also by pharmaceuticals and food-derived compounds such as phytoestrogens, plant protection products, and plastizisers. Exposure to a number of these compounds induces epigenetic changes, particularly alteration of DNA methylation (34,35).

In this study, we set out to analyse the effect of  $ER\beta$ -deficiency on DNA methylation on a genome-wide level. Using reduced representation bisulfite sequencing (RRBS) in mouse embryonic fibroblasts (MEFs) derived from wildtype and  $ER\beta$  knock-out mice, we identified around 8000 differentially methylated positions (DMPs), of which around 6000 were hypomethylated and around 2000 were hypermethylated. Validation and further characterisation of selected DMPs showed that differences in methylation correlated with changes in expression of the nearest gene. Further, hyper- but not hypomethylation was reversible by re-introducing  $ER\beta$  into knock-out MEFs. We also show that  $ER\beta$  was recruited to hypermethylated genes in the wildtype cells and that re-introduction of  $ER\beta$  increased gene expression in the knock-out cells. On the other hand, hypomethylated DMRs occurred predominantly in genes that were not expressed in MEFs suggesting that misregulation of DNA methylation in the absence of  $ER\beta$  during cell differentiation had lead to erroneous passive demethylation. Finally, we show here that  $ER\beta$  interacts with TDG, that this interaction is functional, and that TDG is  $ER\beta$ -dependently recruited to hypermethylated

DMRs in MEFs. Thus we provide evidence that ER $\beta$  plays a role in regulating DNA methylation at specific genomic loci by targeting TDG to these regions.

## **MATERIALS AND METHODS**

### **Plasmids, and antibodies**

Mission shRNA against ER $\beta$  was obtained from Sigma-Aldrich. Expression plasmid for HA-tagged mER $\beta$  (pSSH25-mER $\beta$ ) was constructed by inserting PCR amplified mER $\beta$  cDNA into pSSH25 (36) using XhoI and BglII restriction sites. pSSH25-TDG for mammalian expression of HA-fused TDG(36), pPRS220 for yeast expression of Gal4-activation domain-fused TDG (37), 3xERE-luc (38) and pSG5hER $\beta$  (39) used in luciferase assays, have been published. pRL-TK for normalisation of luciferase activity was purchase from Promega. GST-ER $\beta$  was constructed by cloning cDNA encoding for human ER $\beta$  into pGEX-6P-3 (GE Healthcare) using BamHI and XhoI restriction sites. pACT2-ER $\beta$  was obtained by cloning cDNA encoding for human ER $\beta$  into pACT2 (Clontech) using SmaI and XhoI restriction sites. Antibodies: rat monoclonal anti-HA (3F10) Roche Applied Science, rabbit monoclonal anti-ER $\beta$  (05-824), rabbit polyclonal anti-H3K4m2 (07-030) and rabbit polyclonal anti-H3K27m3 (07-449) Millipore, rabbit polyclonal anti-H3K9m3 (060-050) Diagenode, mouse monoclonal anti-Hsp90 (F-8) Santa Cruz Biotechnology, Inc.

### **Cell culture and transfections**

Mouse embryonic fibroblasts (MEFs) from wildtype (wt) and ER $\beta$  knock-out ( $\beta$ erko) mice (24), and  $\beta$ erko MEFs complemented with ER $\beta$  ( $\beta$ erkohER $\beta$  (24)) as well as TDG  $-/-$  MEFs (36) were kept in high glucose DMEM supplemented with 10% FCS, 1 mM L-glutamine, 1 mM sodium pyruvate, and 1x non-essential amino acids and 5  $\mu$ g/ml blasticidine ( $\beta$ erkohER $\beta$ ). For stimulation with ER agonists, cells were put for at least 2 days into DMEM with 5% dextran-coated charcoal-treated serum and treated with 10 nM E2 or DPN.

Mouse embryonic stem cells (ESCs) were thawed on feeder cells in ESC medium (high glucose DMEM supplemented with 15% heat inactivated FCS, 1 mM sodium pyruvate, 1xnon-essential amino acids, and 0.1 mM  $\beta$ -mercaptoethanol) containing 1 U/ $\mu$ l LIF (Millipore). Upon feeder removal, cells were maintained in 2i medium (serum-free N2B27 (40) supplemented with 2i inhibitors (41), CHIR99021 (3 $\mu$ M), and PD0325901 (1  $\mu$ M), obtained from the Division of Signal Transduction Therapy, University of Dundee) containing



1 U/ $\mu$ l LIF. TDG knock-out ESCs complemented with TDG (-/-pTDG) or empty vector (-/-pvec) were maintained in 2i medium containing 1  $\mu$ g/ml puromycin.

For transfection with HA-ER $\beta$ , MEFs or ESCs were seeded onto 15 cm plates and transfected the following day using JetPRIME reagent (Polyplus transfection) according to the manufacturer's protocol. One day after transfection, cells were harvested for ChIP.

For transfection with shRNA constructs, ESCs were seeded into 6well dishes and transfected the following day using JetPRIME reagent according to the manufacturers protocol. 24 h after transfection, 1  $\mu$ g/ml puromycin was added to the medium. Transfected cells were selected for 4 days changing medium daily, and harvested for RNA extraction.

Transfection for luciferase assays were performed using JetPEI (Polyplus transfection) in 24-well plates with 50 ng pRL-TK, 50 ng pSG5-hER $\beta$ , 100 ng 3xERE-luc, and varying concentration of pSHH25-TDG per well, according to the manufacturer's protocol.

## **RRBS**

Library preparation for RRBS was carried out as described (42). In brief, genomic DNA derived from wt and  $\beta$ erko MEFs was isolated using the QIAamp DNA mini kit (Qiagen). Three batches of genomic DNA of each cell type were pooled, and 1  $\mu$ g was digested with 20 U MspI overnight. The reaction was stopped by addition of 1  $\mu$ l 0.5 M EDTA and purified using MinElute gel extraction kit (Qiagen). Subsequently, DNA fragments were end-repaired and A-tailed by incubation with 5 U Klenow fragment (New England Biolabs Inc.) and 0.5 mM dATP and 0.05 mM dGTP and dCTP at 30°C for 20 min followed by 20 min at 37°C. After purification using the MinElute gel extraction kit, methylated Illumina standard adapters were ligated to the fragments overnight at 16°C using 400 U T4 ligase (New England Biolabs Inc.). The fragments were then separated on a 3% Nusieve 3:1 agarose 0.5x TBE gel and 160-340 bp fragments were excised and purified using the MinElute gel extraction kit. The purified DNA was subjected to two rounds of bisulfite conversion using the EZ DNA Methylation kit (Zymo Research). Subsequently, the final library was prepared by 19 cycles of PCR amplification and purification using the MinElute gel extraction kit. Libraries were sequenced on the Illumina Genome Analyser IIx following the manufacturer's protocol.

## **Bioinformatic and statistical analyses**

Mapping of obtained sequences was performed using the Genomatix mining station (<http://www.genomatix.de/solutions/genomatix-mining-station.html>). Annotation and correlation analyses were carried out using the Genomatix Regionminer. Statistical analyses

were carried out using R (43) or GraphPad Prism. Statistical significance was assessed using the paired t-test. The level of significance was selected as  $p < 0.05$ .

#### **Bisulfite treatment and Pyrosequencing**

Genomic DNA (200-500 ng) was bisulfite treated and purified using the EZ DNA Methylation kit (Zymo Research). 1  $\mu$ l of the converted DNA was used for nested PCR amplification (for primer sequence see Suppl. Table S1) and the PCR product was sequenced by pyrosequencing in a Pyromark Q24 (Qiagen).

#### **Methylation sensitive restriction enzyme digest**

5  $\mu$ g genomic DNA was digested with 50 U HpaII or 100 U MspI overnight. Subsequently, enzymes were removed by digest with proteinase K for 30 min at 40°C and digested fragments were analysed by real time PCR (primers listed in Suppl. Table S1) using Rotor-Gene SYBR Green PCR Kit on a Rotor-Gene RG-3000 (Qiagen).

#### **RNA isolation, cDNA production, and real time PCR**

RNA was isolated using Tri (Sigma) according to the manufacturer's recommendations. 1  $\mu$ g of total RNA was treated with DNaseI (New England Biolabs Inc.) and reverse transcribed using random hexamer primers (Fermentas). 1  $\mu$ l of the resulting cDNA was used for real-time PCR using Rotor-Gene SYBR Green PCR Kit on a Rotor-Gene RG-3000 (Qiagen). Gene transcripts were normalized to the *GAPDH* RNA content (primers listed in Suppl. Table S1). All results are based on the  $\Delta\Delta$ CT method and represent the mean of at least 3 independent experiments.

#### **Chromatin Immunoprecipitation**

ChIP assays were performed as described (44) with minor modifications. Cells were grown to confluency on 15-cm dishes. Chromatin was cross-linked for 10 min with 1% formaldehyde (Pierce Biotechnologies Inc.) and the reaction was stopped by addition of 125 mM glycine for 10 min. After washing twice with cold PBS, cells were harvested in PBS by centrifugation at 4°C at 600g. Nuclei were isolated by sequential 5 min incubation on ice with 500  $\mu$ l cold Nucleus/Chromatin preparation (NCP) Buffer I (10 mM HEPES pH 6.5, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100) and twice cold NCP buffer II (10 mM HEPES pH 6.5, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl). Pelleted nuclei were lysed in 200-400  $\mu$ l lysis buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5% Triton X-100, 1% SDS, 1 mM PMSF, 1x Complete (Roche)) for 10 min followed by sonication for 15 cycles (30 s on, 30 s off,

power high) using a Bioruptor sonicator (Diagenode). After centrifugation at 4 °C at 14,000g for 10 min, chromatin concentration was estimated by absorbance at 260 nm. 100 µg (for HA, TDG and H3K9m3 ChIPs) or 50 µg (for H3K4m2 and H3K27m3 ChIPs) of chromatin were diluted 10 times in IP buffer I (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 150 mM NaCl, 0.1% Triton X-100, 1 mM PMSF, 1x Complete) for HA and TDG ChIPs or IP buffer II (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 1 mM PMSF, 1x Complete) for histone ChIPs. Diluted chromatin was pre-cleared at 4 °C for 1 h with 40 µl of a 50% slurry of magnetic Protein G beads (Invitrogen) preblocked with 1 mg/ml BSA and 1 mg/ml tRNA. Precleared chromatin was incubated with 2–5 µg of the respective antibody (Suppl. Table S2) overnight at 4 °C and immuno-complexes were precipitated with 40 µl of a 50% slurry of blocked Protein G beads at 4 °C for 2 h. Subsequently, beads were serially washed with 500 µl wash buffer I (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% Triton X-100), 500 µl wash buffer II (20 mM Tris-HCl pH 8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% Triton X-100), and 500 µl wash buffer III (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 250 mM LiCl, 1% sodium deoxycholate, 1% NP-40). For TDG ChIPs, beads were washed once with 500 µl wash buffer I and twice with 500 µl wash buffer II. After two additional washes with 500 µl TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA), complexes were eluted by incubating twice with 250 µl extraction buffer (1% SDS, 0.1 M NaHCO<sub>3</sub>) at 65 °C for 15 min. Crosslink was reversed by incubation at 65 °C for 4 h in the presence of 200 mM NaCl. Subsequently, proteins were removed by incubation with proteinase K (50 µg/ml) in the presence of 10 mM EDTA at 45 °C for 1 h, and DNA was purified by phenol/chloroform extraction and ethanol precipitation. The isolated DNA fragments were analysed by qPCR (primers listed in Suppl. Table S1) using Rotor-Gene SYBR Green PCR Kit on a Rotor-Gene RG-3000 (Qiagen).

### **Western blot and far western blot analyses**

The insoluble fraction of luciferase assay-lysates containing chromatin-bound proteins was used for protein expression analysis by immunoblotting using anti-TDG, anti-ER $\beta$ , and anti-Hsp90 antibodies.

For far western blot analysis, GST-tagged ER $\beta$  was expressed in *E. coli* BL21 at 15°C overnight upon induction with 0.2 mM IPTG. Cells were harvested and lysed in 2 ml GST buffer (10 mM Tris-HCl, pH 8; 50 mM NaCl; 5% Glycerol; 1 mM DTT; 0.1 mM PMSF; 1x complete (Roche)) by sonication in a Bioruptor® UDC-200 (Diagenode). After sonication, 1% Triton X-100 was added to the lysates, samples were gently mixed for 30 min at 4°C and

centrifuged at 12'000 x g for 10 min at 4°C. GST-ER $\beta$  was purified by affinity chromatography using glutathione sepharose (GE Healthcare) according to the manufacturer's protocol. Proteins were eluted by boiling in Laemmli buffer at 95°C for 5 min. Proteins were fractionated by SDS-PAGE and transferred to a nitrocellulose membrane. The proteins on the membrane were denatured by incubation with 6 M guanidine-HCl (GuHCl) in AC buffer (20 mM Tris-HCl, pH 7.6; 100 mM NaCl; 10% Glycerol; 0.1% Tween-20; 2% skim milk powder; 1 mM DTT; 0.5 mM EDTA) for 30 min at RT, and renatured by washing steps with 3 M GuHCl for 30 min at RT, 1 M GuHCl for 30 min at RT, 0.1 M GuHCl for 30 min at 4°C and AC buffer only for 1h at 4°C. Upon blocking with 5% skim milk powder in TBST for 1h at RT, the membrane was incubated at 4°C overnight with gentle shaking in protein-binding buffer (20 mM Tris-HCl, pH 8; 50 mM NaCl; 10% Glycerol; 0.1% Tween-20; 2% skim milk powder; 1 mM DTT) containing 7-9  $\mu$ g recombinant TDG (45). The membranes were washed thoroughly 3 times with TBS containing 0.2% NP-40. Subsequently, bound proteins were detected using antibodies against TDG and GST.

### **Yeast two hybrid analysis of ER $\beta$ -TDG interaction**

The Matchmaker<sup>TM</sup> yeast-two hybrid system (Clontech) was used. Bait and trait proteins were cloned into plasmids encoding the binding and activation domain of the Gal4 protein, respectively. The *Saccharomyces cerevisiae* strains AH109 (*MAT $\alpha$* , *trp1-901*, *leu2-3, 112*, *ura3-52*, *his3-200*, *gal4 $\Delta$* , *gal80 $\Delta$* , *LYS2::GAL1<sub>UAS</sub>-GAL1<sub>TATA</sub>-HIS3*, *MEL1*, *GAL2<sub>UAS</sub>-GAL2<sub>TATA</sub>-ADE2*, *URA3::MEL1<sub>UAS</sub>-MEL1<sub>TATA</sub>-lacZ*) and Y187 (*MAT $\alpha$* , *ura3-52*, *his3-200*, *ade2-101*, *trp1-901*, *leu2-3, 112*, *gal4 $\Delta$* , *mef*, *gal80 $\Delta$* , *URA3::GAL1<sub>UAS</sub>-GAL1<sub>TATA</sub>-lacZ*) were co-transformed with 50-500 ng of bait and trait plasmids according to the Clontech manual. For AH109, interactions were assessed by spotting serial dilutions of cells on selective medium (SC-LEU-TRP-ADE-HIS) and incubating them for 2 to 4 days at 30°C.  $\beta$ -Galactosidase activity was assayed using the Y187 strain (Clontech manual). Briefly, 10<sup>6</sup> cells were dropped on SC medium selecting for the plasmids (SC-LEU-TRP) and grown for 24 h at 30°C. Cells were transferred to filter paper (Filtrak, 80 g/m<sup>2</sup>) before snap-freezing in liquid nitrogen and subsequent thawing for cells lysis. The filter with the lysed cells was soaked with 2 ml of Z buffer (100 mM Na phosphate buffer pH 7.0, 10 mM KCl, 1 mM MgSO<sub>4</sub>, 33  $\mu$ M  $\beta$ -mercaptoethanol, 817  $\mu$ M X-Gal) and incubated at 30°C for up to 17 h.

### **Luciferase Assays**

Four hours after transfection, 10 nM E2 was added to the medium. The next day, luciferase reporter assays were performed using the Dual-Luciferase® Reporter Assay System (Promega) according to the manufacturer's protocol: cells were lysed in passive lysis buffer (Promega) and firefly and renilla luciferase activity measured in 96-well plates using a luminometer Centro LB 960 (Berthold technologies).

## RESULTS

### *ERβ deficiency leads to methylation changes in developmental genes.*

In order to identify genes that show DNA methylation changes in the absence of ERβ, reduced representation bisulfite sequencing (RRBS, (46)) was conducted using MEFs derived from ERβ<sup>+/+</sup> (wt) and ERβ<sup>-/-</sup> (βerko) mice (24). Sequencing resulted in roughly 40 million reads of which 40% were unambiguously mapped to the mouse genome. Around 3x10<sup>5</sup> CpGs were covered by the screen, which corresponds to 2.5% of all CpGs in the mouse genome. In both cell lines, most of the covered CpGs were unmethylated, and the rest mainly methylated (Figure 1A). The majority of positions were fully methylated or unmethylated in both cell types, whereas there was more variation between the cell lines at loci with intermediate methylation (20-80%) (Figure 1B). We chose to focus on CpGs which were covered by more than 4 reads in both cell types and were either unmethylated in wt and methylated in βerko cells or vice versa. Cytosines were considered methylated if more than 80% of the reads indicated methylation, and unmethylated if less than 20% of the reads indicated methylation. Using these criteria, 8071 differentially methylated CpGs were identified, 6016 of which were hypomethylated and 2055 were hypermethylated in βerko MEFs. Annotation of differentially methylated positions (DMPs) showed the expected enrichment for promoter- and intragenic regions compared to the whole genome (Figure 1C+D, Table 1). However, gene associated regions were more often found in hypo- than in hypermethylated loci (Table 1). Gene ontology (GO) analysis of genes containing DMPs showed enrichment for pathways involved in developmental processes for both hypo- and hypermethylated genes (Table 2).

### *Hypomethylated regions overlap with silenced transcriptional regulators.*

To further characterise the genomic regions containing DMPs, we compared our data obtained by RRBS with published datasets for enrichment of different chromatin marks. We found the largest significant overlaps for hypomethylated DMPs with histone 3 lysine 27 trimethylation (H3K27m3) and lysine 4 mono-methylation (H3K4m1) in MEFs (46) (Figure

1E+F). These marks are indicative for repressed promoter regions and poised transcriptional enhancers, respectively. Smaller, significant overlaps were found for histone 3 lysine 4 trimethylation (H3K4m3) in combination with H3K27m3 (bivalent chromatin), and H3K4m1 in combination with histone 3 lysine 27 acetylation (H3K27ac) (active enhancer regions). No significant overlap was found with hypermethylated positions. These results suggest that most of the hypomethylated DMPs lie in regions involved in transcriptional regulation, which are however inactive in wildtype MEFs. Indeed, literature-based analyses of associations with tissue and MeSH annotations showed that genes close to hypomethylated DMPs are overrepresented in neuronal tissue (Table 3) and in diseases of the nervous system (Table 4).

*Hyper- but not hypomethylation is complementable by re-introducing ER $\beta$*

For validation we selected 10 hypo- and 10 hypermethylated DMPs with different genomic position and histone methylation patterns (Table 5). We analysed DNA methylation by methylation-dependent restriction digest followed by real time PCR in wt and  $\beta$ erko MEFs as well as in  $\beta$ erko MEFs complemented with ER $\beta$  ( $\beta$ erkohER $\beta$ ) (Fig. 2A). We could confirm hypomethylation for all 10 hypomethylated DMPs in Table 3 and hypermethylation for 8 of the hypermethylated ones. Re-introduction of ER $\beta$  restored wt methylation only in a subset of hypermethylated DMPs (Fig. 2A, DMP hyper1-4). Coincidentally, these loci showed lower methylation in  $\beta$ erko MEFs compared to the other hypermethylation positions as well as the hypermethylated loci in wt cells.

Methylation patterns were validated by pyrosequencing. Promoter region of three genes are shown in Figure 2B: *Dyx1c1* found hypermethylated in wt MEFs, *HoxA9* found hypermethylated and complementable by re-expressing ER $\beta$ , and *Pitx1* found hypermethylated and not complementable. Further, we conducted chromatin immunoprecipitation (ChIP) to measure enrichment of the chromatin marks histone 3 lysine 4 dimethylation (H3K4m2) lysine 27 trimethylation (H3K27m3) and lysine 9 trimethylation (H3K9m3). Bisulfite-pyrosequencing confirmed DNA methylation patterns not only at the DMPs but also at neighbouring CpGs (Fig. 2B). The chromatin state at these regions correlated with DNA methylation in that *Dyx1c1* showed enrichment of both H3K4m2 and H3K27m3, reflecting a bivalent chromatin state, whereas for *HoxA9* and *Pitx1*, only enrichment for H3K4m2 was found in wt MEFs. This pattern was inverted in  $\beta$ erko MEFs, and complemented in  $\beta$ erkohER $\beta$  MEFs where DNA methylation was complementable (Fig. 2B).



*ER $\beta$  regulates hypermethylated targets in MEFs and hypomethylated targets in embryonic stem cells.*

Next, we analysed gene expression pattern of hyper- and hypomethylated targets. Gene expression of *Dyx1c1*, *HoxA9*, and *Pitx1* was inversely correlated with the DNA methylation pattern in MEFs: *Dyx1c1* showed higher expression in  $\beta$ erko and  $\beta$ erkohER $\beta$  cells, and *HoxA9* and *Pitx1* higher expression in wt cells (Figure 3A). As with DNA methylation, the gene expression pattern was reversed by re-introduction of ER $\beta$  for *HoxA9*. This correlation was confirmed for an additional hypomethylated gene, *HoxD9*, and two additional hypermethylated genes where re-introduction of ER $\beta$  complemented the methylation pattern, *HoxA10* and *Tnfaip2* (Figure 3A). No effect of the ER $\beta$  ligand DPN was found on the expression of the tested genes (Supplemental Figure or data not shown). Interestingly, while transcriptional activity was significantly lower for the hypomethylated compared to hypermethylated genes in wt MEFs (Figure 3B, left panel), their expression was comparable in undifferentiated mouse embryonic stem cells (ESCs) (Figure 3B, left panel). Concomitantly, DNA methylation was low at these DMPs in ESCs (Figure 3B, right panel). These results suggest that hypomethylated genes become repressed during differentiation into MEFs. This is corroborated by the findings that hypomethylated DMPs overlap with repressive chromatin marks in MEFs (Figure 1E+F), and associated genes are involved in embryonic development and neuronal functions, and are mostly found in neuronal tissues (Table 2-4).

In order to investigate direct involvement of ER $\beta$  in regulating these genes, ER $\beta$  recruitment to the regions around DMPs was assessed using ChIP in wt MEFs as well as in ESCs. As shown in Figure 3C, ER $\beta$  was enriched in wt MEFs at hypermethylated genes, but only if re-introduction of ER $\beta$  could reverse hypermethylation and increase gene expression. Recruitment to hypomethylated genes could only be detected in ESCs, suggesting that ER $\beta$  is involved in their regulation in these cells. To test this, we assessed their expression in the absence and presence of ER $\beta$  in ESCs using small hairpin (sh)RNA mediated knock-down of ER $\beta$ . As shown in Figure 3D knock-down of ER $\beta$  resulted in decreased expression of all of the tested genes, demonstrating that they are regulated by ER $\beta$ .

*ER $\beta$  interacts with thymine DNA glycosylase.*

Next, we investigated how ER $\beta$  can regulate DNA methylation at specific loci. We hypothesized that ER $\beta$  recruits a factor or factors involved in the regulation of DNA



methylation to these loci. It is known that ER $\alpha$  interacts with thymine DNA glycosylase (TDG) (47), a protein suggested to be involved in active DNA demethylation (13,15,16,48). Thus we investigated if ER $\beta$  can interact with TDG. To this end, we conducted far-western blot analyses, immobilizing GST-tagged ER $\beta$  to nitrocellulose membrane and probing this membrane with recombinant TDG. Subsequently, interaction could be visualized by detecting TDG on the membrane using an antibody against TDG. As seen in Figure 4A, TDG could be detected at the same height as GST-ER $\beta$  whereas it did not bind to the GST-tag alone. GST-tagged SUMO-1, which has been shown to interact with TDG (37), was used as positive control. Only unspecific bands were detected on a membrane that had not been probed with TDG (Figure 4A, right panel).

To corroborate an interaction between ER $\beta$  and TDG, yeast two hybrid assays were conducted in the *S. cerevisiae* strain AH109. This strain harbours the two Gal4-inducible reporter genes *HIS3* and *ADE2* and protein interactions can be assessed by growth on selective medium lacking adenine and histidine. As expression of TDG fused to the *GAL4* activation domain (AD) gave rise to a high degree of auto-activation (data not shown), we addressed potential interactions in strains expressing ER $\beta$  fused to the *GAL4* AD and TDG fused to the *GAL4* DNA binding domain (BD). As shown in Figure 4B, co-expression of ER $\beta$  and TDG enabled growth on selective medium. In contrast, little or no growth was detected when either factor was combined with the corresponding vector control, indicative of specific interactions between ER $\beta$  and TDG. This result was confirmed using another strain (Y187) in which the *lacZ* gene, coding for the  $\beta$ -Galactosidase, serves as the reporter gene to detect protein interactions (Figure 3C). In addition to full length ER $\beta$ , isolated domains of the receptor were tested in order to delineate the domain(s) responsible for interaction with TDG. We tested the AB, DEF, and CDEF domains of ER $\beta$ , however, no reporter activity above background could be detected for the interaction of TDG with any of these constructs. We thus were able to prove an interaction between ER $\beta$  and TDG using two independent methods, far-western blot and yeast two hybrid analyses.

#### *TDG enhances transcriptional activity of ER $\beta$ and is ER $\beta$ -dependently recruited to identified DMPs*

In order to test if the interaction with TDG has an effect on ER $\beta$  function, we conducted reporter gene assays, measuring ER $\beta$  transcriptional activity. To this end, TDG  $-/-$  MEFs were transfected with plasmids encoding for ER $\beta$ , a luciferase reporter gene driven by 3 EREs (38),

and different amounts of TDG. Cells were stimulated with E2 and harvested the next day for measurement of luciferase activity. As shown in Figure 5A, transfection of ER $\beta$  plasmid enhanced transcription of the reporter gene, and E2 treatment lead to a further increase of luciferase activity. Co-transfection with TDG vector enhanced transcriptional activity of ER $\beta$  additionally. This increase was observed both in the absence (maximal 2-fold) as well as in the presence (maximal 2.5-fold) of ligand.

We also investigated if TDG is recruited to ER $\beta$ -regulated genes in MEFs, and if this recruitment is dependent on the presence of ER $\beta$ . Using ChIP assays, we found TDG recruitment to these genes in wt and  $\beta$ koER $\beta$  MEFs (Figure 5B). However, for HoxA10 and Tnfaip2, recruitment was abolished in  $\beta$ ko MEFs, indicating ER $\beta$ -dependent recruitment of TDG to these loci.

#### *TDG regulates differentially methylated genes in ESCs.*

The suggested role of TDG in DNA demethylation is to process 5-carboxylcytosine (5caC) and 5-formylcytosine (5fC) (11). Indeed, TDG deficient ESCs show accumulation of 5fC and 5caC at gene regulatory elements (15,16). Thus, if ER $\beta$  can recruit TDG to certain genomic regions in order to regulate DNA methylation, lack of ER $\beta$  would result in less TDG recruitment and hence, accumulation of 5fC and 5caC at these loci. Comparison of genome-wide 5fC data (16) with the RRBS data revealed an overlap of hyper- and hypomethylated DMRs with 5fC in wt ESCs of 10% and 12.5%, respectively. However, in TDG deficient cells, this overlap increased to 32% and 46%, respectively (Figure 6A+B). In other words, almost half of the hypermethylated DMRs identified in our screen overlapped with regions where TDG processes 5fC. Thus, we tested if TDG also transcriptionally regulates the differentially methylated targets in ESCs. Gene expression was analysed in TDG deficient ESCs and cells complemented with TDG. A clear down-regulation was observed in ESCs lacking TDG (Figure 6C), demonstrating that they are regulated by TDG.

## **DISCUSSION**

There is accumulating evidence that exposure to compounds interfering with the estrogen system induce epigenetic changes, particularly alterations in DNA methylation. Several studies suggest that nuclear receptors including the ERs are directly involved in regulating DNA methylation (24-29). In this study, we systematically addressed the role of ER $\beta$  in regulating DNA methylation at specific loci. To this end, we carried out RRBS comparing ER $\beta$  proficient and deficient MEFs. Using this method, we limited the screen to CpG-rich

regions; further, we considered only DMPs that were fully methylated (80-100%) in one cell type and unmethylated (0-20%) in the other. Despite these stringent limitations, we could identify over 8000 DMPs, one-third of which was hyper- and two-thirds hypomethylated. For selected differentially methylated genes, we could show that their expression was correlated with their DNA methylation state, and that they were bound and transcriptionally regulated by ER $\beta$  when they were actively transcribed. Further, we demonstrated that ER $\beta$  interacts physically and functionally with TDG, and that TDG is recruited to DMPs and involved in the regulation of the associated genes.

Based on our results, we propose the following model (Figure 7): ER $\beta$  binds to regulatory regions of target genes and recruits TDG to these places. This interaction enhances gene expression on one hand and prevents DNA methylation on the other hand. The latter is achieved by the interplay with the TET proteins that oxidise stochastically methylated CpGs to 5fC and 5caC, which in turn can be processed by TDG and the base excision repair (BER) to unmethylated C. This mechanism can explain the observed hypermethylated loci where re-introduction of ER $\beta$  resulted in wildtype methylation and gene expression patterns. Indeed, these loci were hyper- but not fully methylated in  $\beta$ erko MEFs, supporting the notion that the difference arises from a lack of reverting stochastic methylation events rather than complete switch-off of these genes in  $\beta$ erko cells. Hypomethylated genes, on the other hand, were inactivated in MEFs but transcribed and regulated by ER $\beta$  and TDG in ESCs. It has been shown that when differentiation is initiated, dynamic de- and remethylation processes occur prior to establishing DNA methylation pattern defining the expression in the differentiated cell. If TDG is absent during this process, 5fC and 5caC are accumulated (15,16). As these marks are not recognized by the maintenance DNMT, this accumulation leads to passive demethylation during cell division. Thus, at genes that become silenced during differentiation to MEFs, lack of ER $\beta$ , and hence diminished TDG recruitment, leads to erroneous passive demethylation, resulting in hypomethylation in the differentiated cells. Indeed, we found a remarkable overlap between hypomethylated sites from our data set and 5fC found in TDG deficient ESCs (16). At this point, we do not have an explanation for the occurrence of hypermethylated positions whose methylation pattern is not revertible by re-expression of ER $\beta$  other than clonal differences between wt and  $\beta$ erko cells.

Notably, as in our previous study (49), we could not find any effect of estrogen on ER $\beta$ 's function at the investigated loci, neither on its transcriptional activity nor on its effects on

DNA methylation and interaction with TDG. In contrast, TDG was shown to interact with ER $\alpha$ , and enhance its transcriptional activity, in the presence of ligand (47). The interaction between ER $\alpha$  and TDG is mediated by SRC-1, and overexpression of this co-factor in COS1 cells resulted in enhancement of ER $\alpha$ 's ligand-independent transcriptional activity by TDG (50). Thus, availability of other factors involved in ER-TDG interaction in the different cell systems could explain the difference between the two receptor isoforms. On the other hand, ligand-independent function of ER $\beta$  has been shown previously in different contexts, e.g. when modulating ER $\alpha$ -induced gene expression in breast cancer cells (51). Additionally, we have found that ER $\beta$  is tightly bound to the chromatin in extracts of different cell types (unpublished observation) even in the absence of ligand. Further, we could neither identify any classical EREs in the regions that ER $\beta$  bound to, nor were EREs found enriched around differentially methylated sites. We therefore suggest that the function ER $\beta$  exerts at the identified targets differs from the classical ligand-induced ER signalling pathway. How ER $\beta$  binds to these loci and how its recruitment is regulated has to be determined in future studies.

Although the factors that regulate changes in DNA methylation patterns during cell differentiation are identified, it is still enigmatic how they are recruited to specific genomic loci. R loop formation at CGIs has been shown to exclude DNMT3a and DNMT3b, hence preventing methylation of these structures (52). On the other hand, it has been suggested that locus specific *de novo* methylation is induced by recruitment of DNMTs by non-coding RNAs (20-22) or by proximal sequence elements coding for specific transcription factors (17). The involvement of transcription factors in regulating DNA methylation patterns has also been shown at distal regulatory regions with low methylation (LMRs) (18,19). The data presented here further supports the notion that transcription factor can target DNA-methylation and -demethylation events, and provide a mechanism underlying this role. We suggest that interaction with factors regulating DNA methylation patterns is not limited to ER $\beta$  but could be general principle applying for many sequence-specific transcription factors. Further, we do not exclude that ER $\beta$  can also interact with proteins of the TET family or excludes DNMT-binding to certain loci. Such mechanisms could explain the fact that TDG binding was not ER $\beta$ -dependent at the HoxA9 gene, whose expression and methylation state was regulated by ER $\beta$ .

In conclusion, the data presented here suggest that ER $\beta$  can regulate methylation patterns at specific genomic loci by interaction with TDG. This implies an important regulatory function of ER $\beta$  during cell differentiation and could be part of the mechanism underlying epigenetic alterations observed after exposure to compounds disrupting ER function in early development. Further, it supports a general concept in which transcription factors regulate the DNA methylation state at specific regions in the genome.

**FIGURE LEGENDS****Figure 1: ER $\beta$ -deficiency leads to altered DNA methylation patterns.**

A: Histogram showing the distribution of methylation at the sequenced cytosines in wt and  $\beta$ erko MEFs.

B: Scatterplot of % methylation in wt vs  $\beta$ erko MEFs at cytosines covered in both cell types.

C: Pie chart presenting the genomic distribution of hypo- and hypermethylated positions. A position was considered hypermethylated if more than 80% of the reads indicated methylation and hypomethylated if less than 20% indicated methylation in  $\beta$ erko MEFs.

D: Enrichment ( $\log_2$  ratios of observed over random) of hypo- and hypermethylated positions at different genomic features.

E: Comparison of regions identified by RRBS with datasets for histone modifications in MEFs (46) using GenomeInspector (Genomatix). Hypo and hyper refer to the hypo- and hypermethylated regions, respectively. OD, odd ratios.

F: Enrichment ( $\log_2$  ratios of observed over random) of histone modifications at hypo- and hypermethylated positions.

**Figure 2: Hyper- but not hypomethylation is reversible by re-introduction of ER $\beta$  into  $\beta$ erko MEFs.**

A: DNA methylation analysis of 10 hypo- and 8 hypermethylated positions. DNA methylation was assessed by methylation specific enzymatic digest followed by qPCR. Positions with gene names in brackets were chosen for further analysis.

B: DNA methylation (left panel) and histone modifications (right panel) of differentially methylated genes in wt,  $\beta$ erko, and  $\beta$ erkohER $\beta$  MEFs. DNA methylation was assessed by pyrosequencing of bisulfite-treated DNA. \*\*\* indicates significant differences ( $p < 0.005$ ) for wt vs.  $\beta$ erko and  $\beta$ erkohER $\beta$  (Dyx1c1 and Pitx1) or  $\beta$ erko vs. wt and  $\beta$ erkohER $\beta$  (HoxA9). Histone modifications were analysed using ChIP followed by qPCR and normalised to *HPRT* (H3K4m2 and H3K27m3) or *GAPDH* promoter (H3K9m3) (means + sd;  $n \geq 3$ ).

**Figure 3: ER $\beta$ -dependent transcription of differentially methylated genes in MEFs and ESCs.**

A: Gene expression analysis of hypomethylated (Dyx1c1, HoxD9), hypermethylated complementable (HoxA9, HoxA10, and Tnfaip2) and hypermethylated non-complementable

genes in wt,  $\beta$ erko, and  $\beta$ erkohER $\beta$  MEFs. Gene expression was analysed by RT-qPCR (mean+sd;  $n \geq 3$ ).

B: Gene expression (left panel) and DNA methylation (right panel) of differentially methylated genes in wt MEFs and ESCs. Gene expression was analysed by RT-qPCR, DNA methylation by methylation specific enzymatic digest followed by q PCR. In MEFs, *Dyx1c1* expression was significantly lower than *HoxA9* (\*), *HoxA10* (\*\*), *Tnfaip2* (\*\*) and *Pitx1* (\*\*) expression, *HoxD9* levels were significantly lower than *HoxA10* (\*\*), *Tnfaip2* (\*\*) and *Pitx1* (\*\*) levels.

C: ER $\beta$  recruitment to differentially methylated genes in wt MEFs and ESCs. HA-tagged ER $\beta$  was precipitated and differentially methylated regions analysed by qRT-PCR. \* indicates significant ( $p < 0.05$ ) enrichment compared to binding at an unrelated, heterochromatic region on chromosome 2 (13) (means + sd;  $n \geq 3$ ).

D: ER $\beta$ -dependent expression of differentially methylated genes in ESCs. Gene expression was assessed by qRT-PCR 4 days after transfection with plasmid encoding for shRNA against ER $\beta$  or non-targeting control (means + sd;  $n \geq 3$ ). All the genes showed significantly decreased expression compared to shcontrol (\*\*  $p < 0.01$ , \*\*\* $p < 0.005$ ).

#### **Figure 4: ER $\beta$ interacts directly with TDG.**

A: Interaction of ER $\beta$  and TDG on far western blots. GST-tagged ER $\beta$  (lanes 1+2), GST (lane 3) and GST-tagged SUMO-1 as positive control (lane 4) were immobilised on a membrane and probed with recombinant TDG. Proteins were detected using antibody against GST (left panel) or TDG (middle panel). The right panel shows a membrane not probed with recombinant TDG. Asterisks mark unspecific bands.

B: Interaction between ER $\beta$  and TDG in yeast two hybrid assays. ER $\beta$  fused to the activation domain (AD) and TDG fused to the binding domain (BD) of *GAL4* were expressed in the yeast strain AH109. Serial dilutions of these cells were spotted on control (SC-LEU-TRP, left panel) and selective medium (SC-LEU-TRP-HIS-ADE, right panel) to monitor activity of the reporter genes *ADE2* and *HIS3*. As a positive control, murine p53 fused to *GAL4* BD was used in combination with SV40 large T-antigen fused to *GAL4* AD. The Gal4 BD and/or GAL4 AD alone served as negative controls (---).

C: Domain mapping for ER $\beta$  using yeast two hybrid assays. Activity was tested in the yeast strain Y187 using *lacZ* as a reporter gene. Activity of the *lacZ*-encoded  $\beta$ -Galactosidase leads to cleavage of X-gal and concomitant accumulation of a blue product (5,5'-dibromo-4,4'-dichloro-indico). In addition to constructs as in B, individual ER $\beta$  domains (AB, CDEF, DEF)

were fused to the *GAL4* AD and used for transformation of Y187 cells.  $10^6$  cells were dropped onto SC plates lacking leucine and tryptophan. After 24 h of growth, cells were lysed and incubated with X-Gal for up to 17 h to monitor appearance of blue color.

**Figure 5: ER $\beta$ -TDG interaction affects gene regulation.**

A: TDG increases ER $\beta$  transcriptional activity in reporter assays. Tdg<sup>-/-</sup> MEFs were transfected with plasmids encoding for the reporter gene 3xERE-luc, TK-renilla, ER $\beta$  and TDG in different concentrations, and treated with 10 nM E2. After 16h, firefly luciferase activity was measured and normalised against renilla luciferase activity (means + sd; n  $\geq$  3). TDG co-expression increased luciferase activity significantly (\*\*p<0.01, p\*\*\*<0.005). Expression of ER $\beta$  and TDG was confirmed by western blot analysis.

B: TDG is recruited to ER $\beta$ -regulated genes in MEFs. TDG recruitment to indicated genes in wt,  $\beta$ ko, and  $\beta$ ko/ER $\beta$  MEFs was analysed by ChIP-qPCR. \* indicates significant (p<0.05) enrichment compared to binding at an unrelated, heterochromatic region on chromosome 2 (13) (means + sd; n  $\geq$  3).

**Figure 6:**

A: Venn diagram showing the overlap between regions identified by RRBS with datasets for 5fC enrichment in deficient (TDG -/-) or proficient (TDG fl/fl) ESCs (16) using GenomeInspector (Genomatix). Hypo and hyper refer to the hypo- and hypermethylated regions identified in this study.

B: Enrichment (log2 ratios of observed over random) of 5fC in deficient (TDG -/-) or proficient (TDG fl/fl) ESCs at differently methylated positions.

C: TDG-dependency of differentially methylated genes in ESCs. Gene expression in TDG deficient (-/- pvector) and TDG complemented (-/- pTDG) cells was assessed by qRT-PCR (means + sd; n =3). \* indicates significant (p<0.05) increase in TDG proficient vs. deficient cells.

**Figure 7: Model of the role of ER $\beta$  in regulation of DNA methylation.**

Hypermethylated, complementable genes: ER $\beta$  binds to regulatory regions of target genes and recruits TDG to these places. This interaction enhances gene expression on one hand and prevents DNA methylation on the other hand. The latter is achieved by the interplay with the



TET proteins that oxidise stochastically methylated CpGs to 5fC and 5caC, which in turn can be processed by TDG and BER to unmethylated C.

Hypomethylated genes: ER $\beta$  binds to regulatory regions of target genes and recruits TDG to these places in ESCs. When cells start to differentiate, dynamic de- and remethylation processes take place in order to establish a stable DNA methylation pattern in committed cells. The absence of TDG during this process leads to an accumulation of 5fC and 5caC. These marks are not recognized by the maintenance DNMT, hence accumulation of these modifications leads to passive demethylation during cell division. Thus, at genes that become silenced during differentiation to MEFs, lack of ER $\beta$ , and thus diminished TDG recruitment, leads to erroneous passive demethylation, resulting in hypomethylation in the differentiated cells.

### **FUNDING**

This work was supported by the Swiss National Science Foundation, the University of Basel Research Fund, and the Swedish Research Council Formas.

### **ACKNOWLEDGMENTS**

## REFERENCES

1. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev*, **16**, 6-21.
2. Chen, T. and Dent, S.Y. (2014) Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nature reviews. Genetics*, **15**, 93-106.
3. Chen, M. and Zhang, L. (2011) Epigenetic mechanisms in developmental programming of adult disease. *Drug discovery today*, **16**, 1007-1018.
4. Patra, S.K., Patra, A., Rizzi, F., Ghosh, T.C. and Bettuzzi, S. (2008) Demethylation of (Cytosine-5-C-methyl) DNA and regulation of transcription in the epigenetic pathways of cancer development. *Cancer Metastasis Rev*, **27**, 315-334.
5. Saitou, M., Kagiwada, S. and Kurimoto, K. (2012) Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development*, **139**, 15-31.
6. Coskun, V., Tsoa, R. and Sun, Y.E. (2012) Epigenetic regulation of stem cells differentiating along the neural lineage. *Current opinion in neurobiology*, **22**, 762-767.
7. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. and Song, H. (2011) Emerging roles of TET proteins and 5-hydroxymethylcytosines in active DNA demethylation and beyond. *Cell cycle*, **10**, 2662-2668.
8. Shen, L. and Zhang, Y. (2013) 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. *Current opinion in cell biology*, **25**, 289-296.
9. Gao, F., Xia, Y., Wang, J., Luo, H., Gao, Z., Han, X., Zhang, J., Huang, X., Yao, Y., Lu, H. *et al.* (2013) Integrated detection of both 5-mC and 5-hmC by high-throughput tag sequencing technology highlights methylation reprogramming of bivalent genes during cellular differentiation. *Epigenetics : official journal of the DNA Methylation Society*, **8**.
10. Bernstein, B.E., Meissner, A. and Lander, E.S. (2007) The mammalian epigenome. *Cell*, **128**, 669-681.
11. Maiti, A. and Drohat, A.C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem*, **286**, 35334-35338.
12. Jacobs, A.L. and Schar, P. (2012) DNA glycosylases: in DNA repair and beyond. *Chromosoma*, **121**, 1-20.
13. Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F. *et al.* (2011) Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*, **470**, 419-423.
14. Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D. *et al.* (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, **146**, 67-79.
15. Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.L., Zhang, K. and Zhang, Y. (2013) Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*, **153**, 692-706.
16. Song, C.X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H. *et al.* (2013) Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell*, **153**, 678-691.
17. Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F. and Schubeler, D. (2011) Identification of genetic elements that autonomously determine DNA methylation states. *Nature genetics*, **43**, 1091-1097.
18. Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L. and Schubeler, D. (2013) Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS genetics*, **9**, e1003994.

19. Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D. *et al.* (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, **480**, 490-495.
20. Schmitz, K.M., Mayer, C., Postepska, A. and Grummt, I. (2010) Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev*, **24**, 2264-2269.
21. Zhang, Z., Tang, H., Wang, Z., Zhang, B., Liu, W., Lu, H., Xiao, L., Liu, X., Wang, R., Li, X. *et al.* (2011) MiR-185 targets the DNA methyltransferases 1 and regulates global DNA methylation in human glioma. *Molecular cancer*, **10**, 124.
22. Guseva, N., Mondal, T. and Kanduri, C. (2012) Antisense noncoding RNA promoter regulates the timing of de novo methylation of an imprinting control region. *Dev Biol*, **361**, 403-411.
23. Martens, J.H., Rao, N.A. and Stunnenberg, H.G. (2011) Genome-wide interplay of nuclear receptors with the epigenome. *Biochim Biophys Acta*, **1812**, 818-823.
24. Ruegg, J., Cai, W., Karimi, M., Kiss, N.B., Swedenborg, E., Larsson, C., Ekstrom, T.J. and Pongratz, I. (2011) Epigenetic regulation of glucose transporter 4 by estrogen receptor beta. *Mol Endocrinol*, **25**, 2017-2028.
25. Thomassin, H., Flavin, M., Espinas, M.L. and Grange, T. (2001) Glucocorticoid-induced DNA demethylation and gene memory during development. *The EMBO journal*, **20**, 1974-1983.
26. Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G. *et al.* (2008) Cyclical DNA methylation of a transcriptionally active promoter. *Nature*, **452**, 45-50.
27. Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F. and Reid, G. (2008) Transient cyclical methylation of promoter DNA. *Nature*, **452**, 112-115.
28. Kim, M.S., Kondo, T., Takada, I., Youn, M.Y., Yamamoto, Y., Takahashi, S., Matsumoto, T., Fujiyama, S., Shirode, Y., Yamaoka, I. *et al.* (2009) DNA demethylation in hormone-induced transcriptional derepression. *Nature*, **461**, 1007-1012.
29. Marques, M., Laflamme, L. and Gaudreau, L. (2013) Estrogen receptor alpha can selectively repress dioxin receptor-mediated gene expression by targeting DNA methylation. *Nucleic acids research*.
30. Tammimies, K., Tapia-Paez, I., Ruegg, J., Rosin, G., Kere, J., Gustafsson, J.A. and Nalvarte, I. (2012) The rs3743205 SNP Is Important for the Regulation of the Dyslexia Candidate Gene DYX1C1 by Estrogen Receptor beta and DNA Methylation. *Mol Endocrinol*.
31. Zhao, C., Dahlman-Wright, K. and Gustafsson, J.A. (2008) Estrogen receptor beta: an overview and update. *Nucl Recept Signal*, **6**, e003.
32. Barros, R.P. and Gustafsson, J.A. (2011) Estrogen receptors and the metabolic network. *Cell metabolism*, **14**, 289-299.
33. Metivier, R., Penot, G., Hubner, M.R., Reid, G., Brand, H., Kos, M. and Gannon, F. (2003) Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell*, **115**, 751-763.
34. Ho, S.M., Johnson, A., Tarapore, P., Janakiram, V., Zhang, X. and Leung, Y.K. (2012) Environmental epigenetics and its implication on disease risk and health outcomes. *ILAR journal / National Research Council, Institute of Laboratory Animal Resources*, **53**, 289-305.
35. Bernal, A.J. and Jirtle, R.L. (2010) Epigenomic disruption: the effects of early developmental exposures. *Birth defects research. Part A, Clinical and molecular teratology*, **88**, 938-944.

36. Kunz, C., Focke, F., Saito, Y., Schuermann, D., Lettieri, T., Selfridge, J. and Schar, P. (2009) Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil. *PLoS Biol*, **7**, e91.
37. Hardeland, U., Steinacher, R., Jiricny, J. and Schar, P. (2002) Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *The EMBO journal*, **21**, 1456-1464.
38. Legler, J., van den Brink, C.E., Brouwer, A., Murk, A.J., van der Saag, P.T., Vethaak, A.D. and van der Burg, B. (1999) Development of a stably transfected estrogen receptor-mediated luciferase reporter gene assay in the human T47D breast cancer cell line. *Toxicol Sci*, **48**, 55-66.
39. Delaunay, F., Pettersson, K., Tujague, M. and Gustafsson, J.A. (2000) Functional differences between the amino-terminal domains of estrogen receptors alpha and beta. *Mol Pharmacol*, **58**, 584-590.
40. Ying, Q.L. and Smith, A.G. (2003) Defined conditions for neural commitment and differentiation. *Methods in enzymology*, **365**, 327-341.
41. Ying, Q.L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008) The ground state of embryonic stem cell self-renewal. *Nature*, **453**, 519-523.
42. Gu, H., Smith, Z.D., Bock, C., Boyle, P., Gnirke, A. and Meissner, A. (2011) Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature protocols*, **6**, 468-481.
43. Team, R.c. (2014). R Foundation for Statistical Computing.
44. Ruegg, J., Swedenborg, E., Wahlstrom, D., Escande, A., Balaguer, P., Pettersson, K. and Pongratz, I. (2008) The transcription factor aryl hydrocarbon receptor nuclear translocator functions as an estrogen receptor beta-selective coactivator, and its recruitment to alternative pathways mediates antiestrogenic effects of dioxin. *Mol Endocrinol*, **22**, 304-316.
45. Steinacher, R. and Schar, P. (2005) Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Current biology : CB*, **15**, 616-623.
46. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B. *et al.* (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454**, 766-770.
47. Chen, D., Lucey, M.J., Phoenix, F., Lopez-Garcia, J., Hart, S.M., Losson, R., Buluwela, L., Coombes, R.C., Chambon, P., Schar, P. *et al.* (2003) T:G mismatch-specific thymine-DNA glycosylase potentiates transcription of estrogen-regulated genes through direct interaction with estrogen receptor alpha. *J Biol Chem*, **278**, 38586-38592.
48. He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L. *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303-1307.
49. Ruegg, J., Cai, W., Karimi, M., Kiss, N.B., Swedenborg, E., Larsson, C., Ekstrom, T.J. and Pongratz, I. Epigenetic Regulation of Glucose Transporter 4 by Estrogen Receptor beta. *Mol Endocrinol*, **25**, 2017-2028.
50. Lucey, M.J., Chen, D., Lopez-Garcia, J., Hart, S.M., Phoenix, F., Al-Jehani, R., Alao, J.P., White, R., Kindle, K.B., Losson, R. *et al.* (2005) T:G mismatch-specific thymine-DNA glycosylase (TDG) as a coregulator of transcription interacts with SRC1 family members through a novel tyrosine repeat motif. *Nucleic acids research*, **33**, 6393-6404.

51. Chang, E.C., Frasor, J., Komm, B. and Katzenellenbogen, B.S. (2006) Impact of estrogen receptor beta on gene networks regulated by estrogen receptor alpha in breast cancer cells. *Endocrinology*, **147**, 4831-4842.
52. Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I. and Chedin, F. (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular cell*, **45**, 814-825.

Figure 1

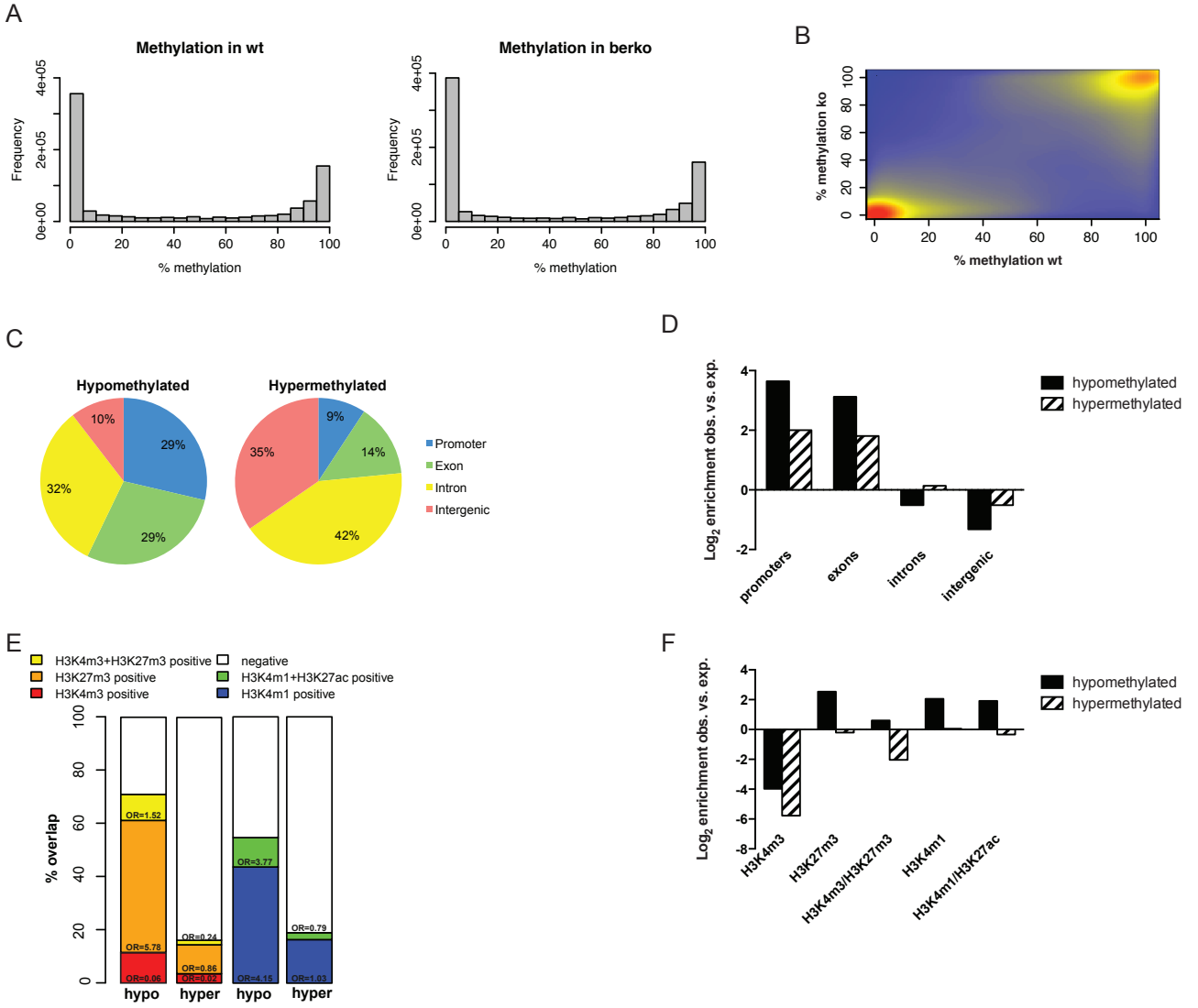
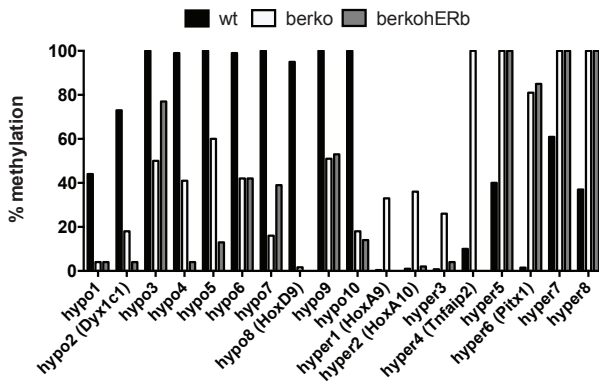


Figure 2

A



B

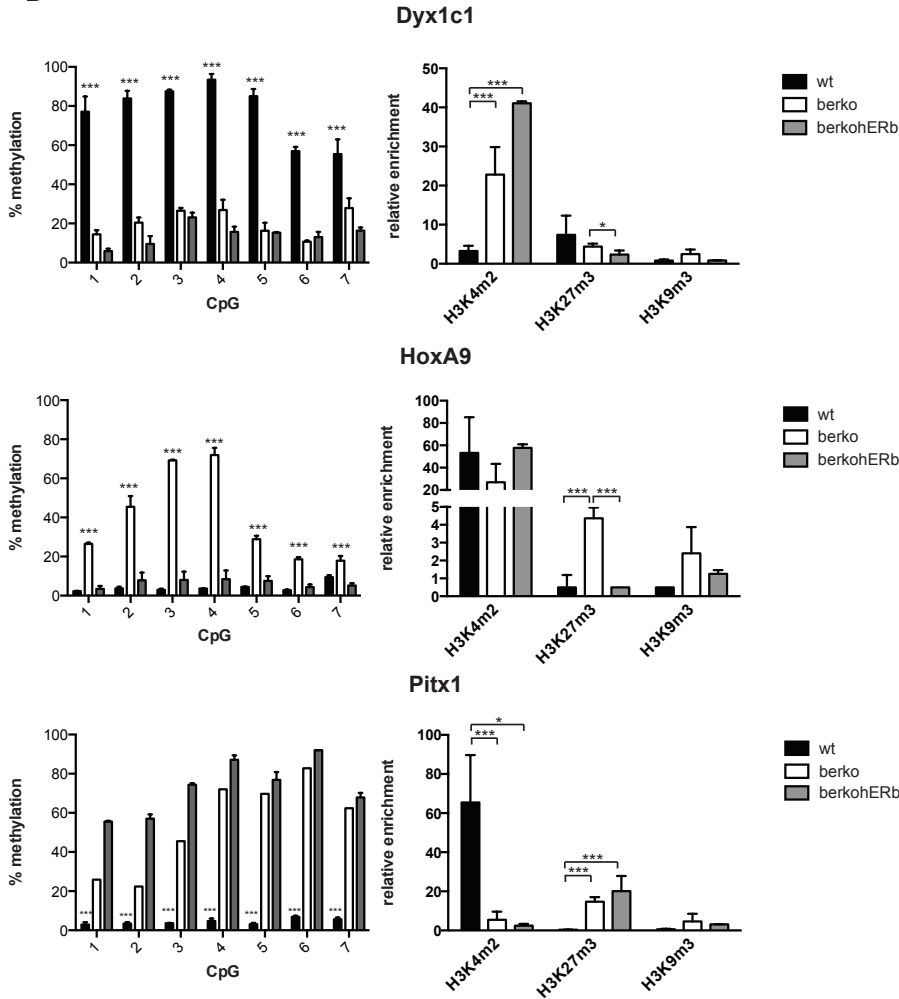


Figure 3

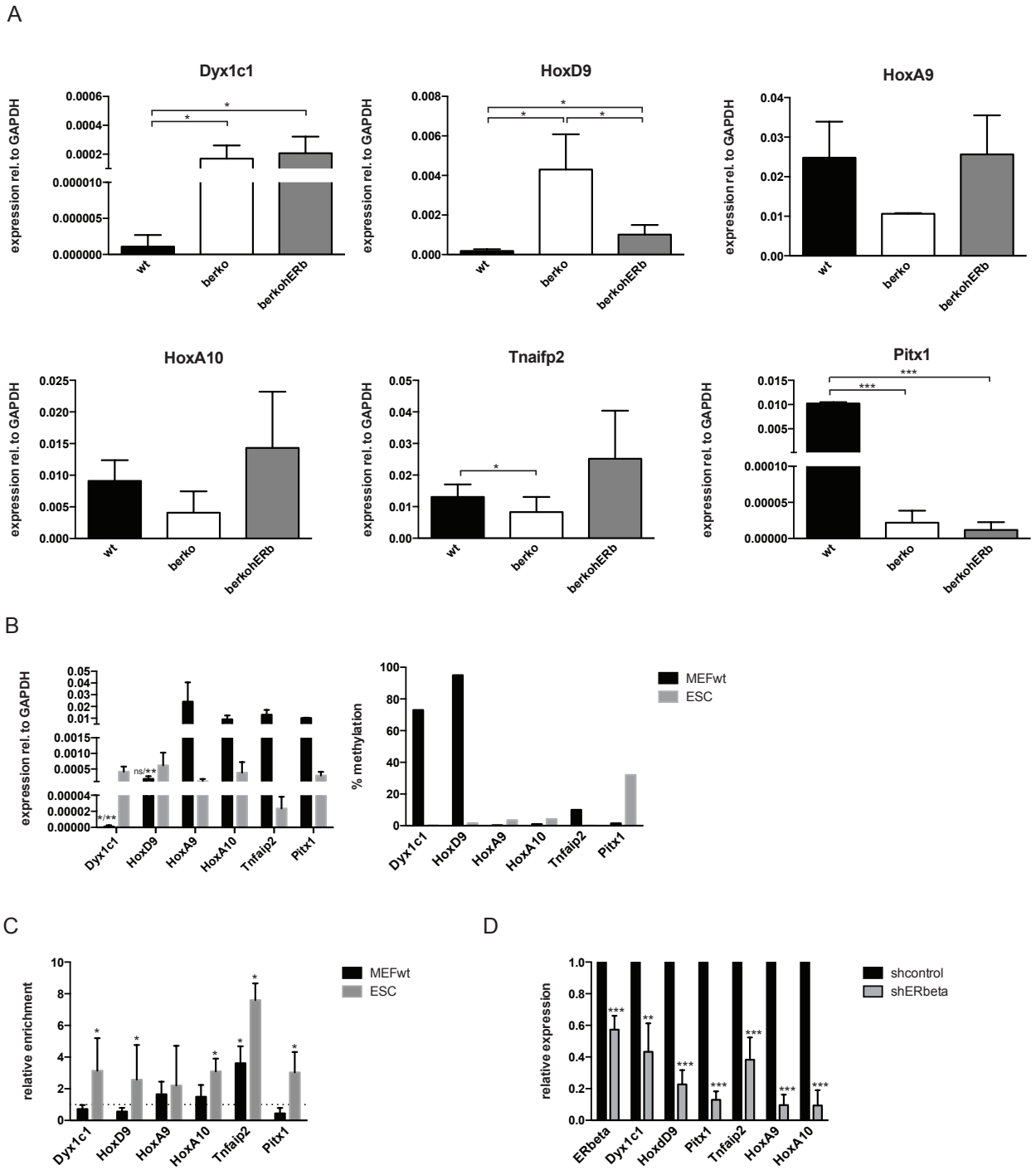




Figure 4

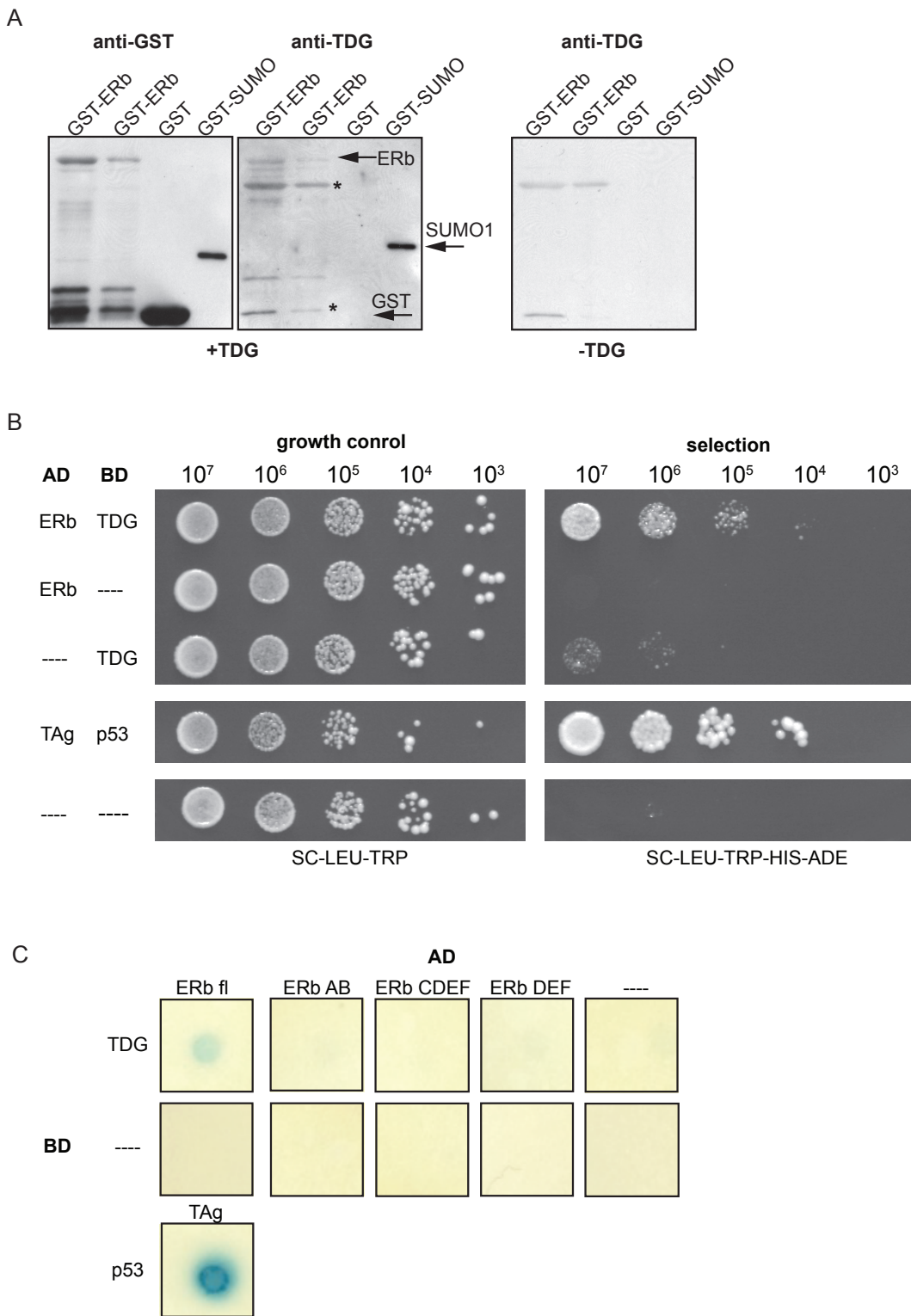


Figure 5

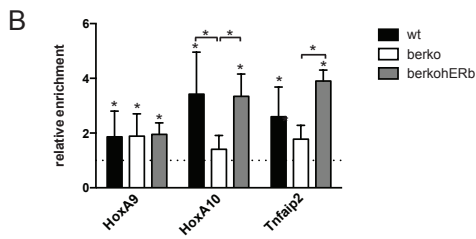
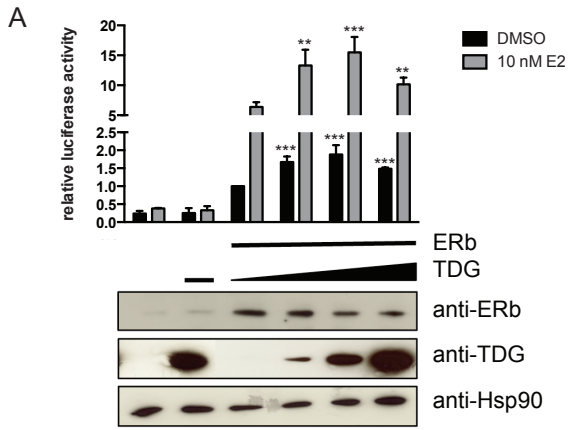


Figure 6

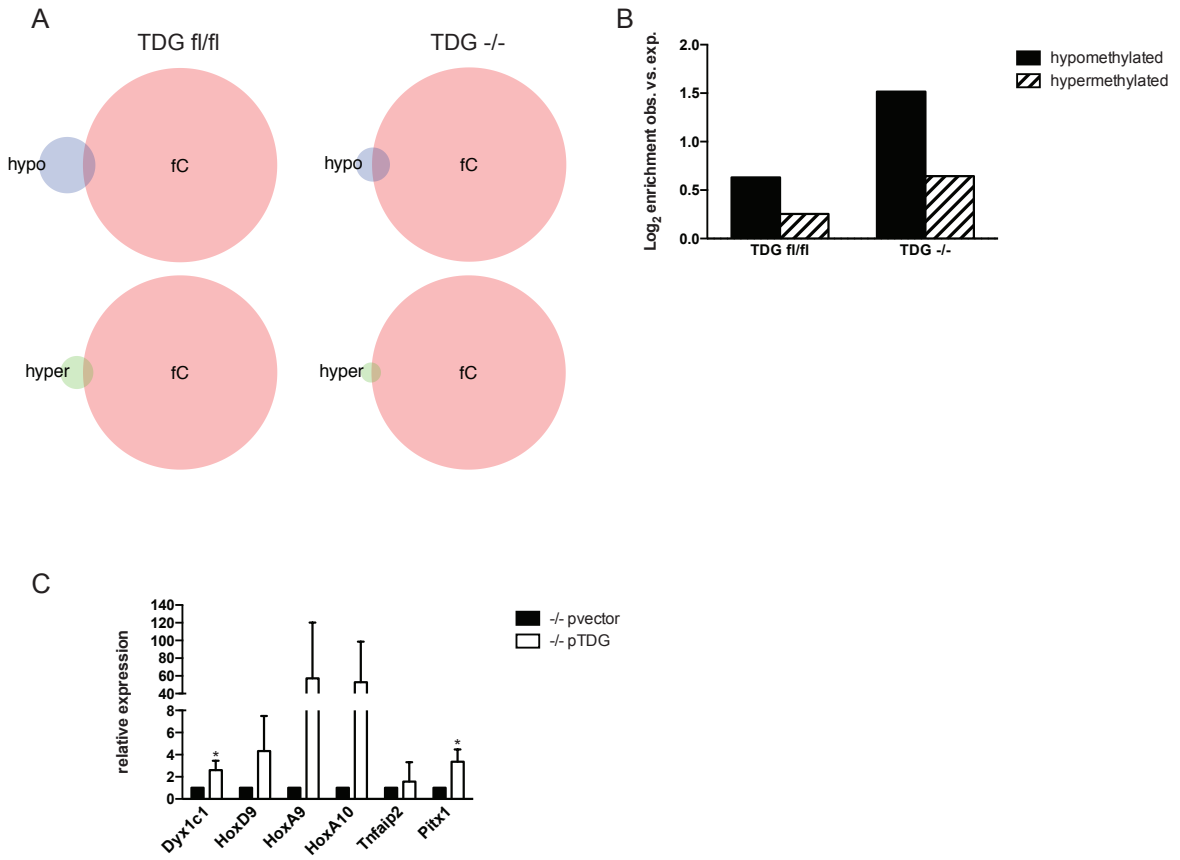
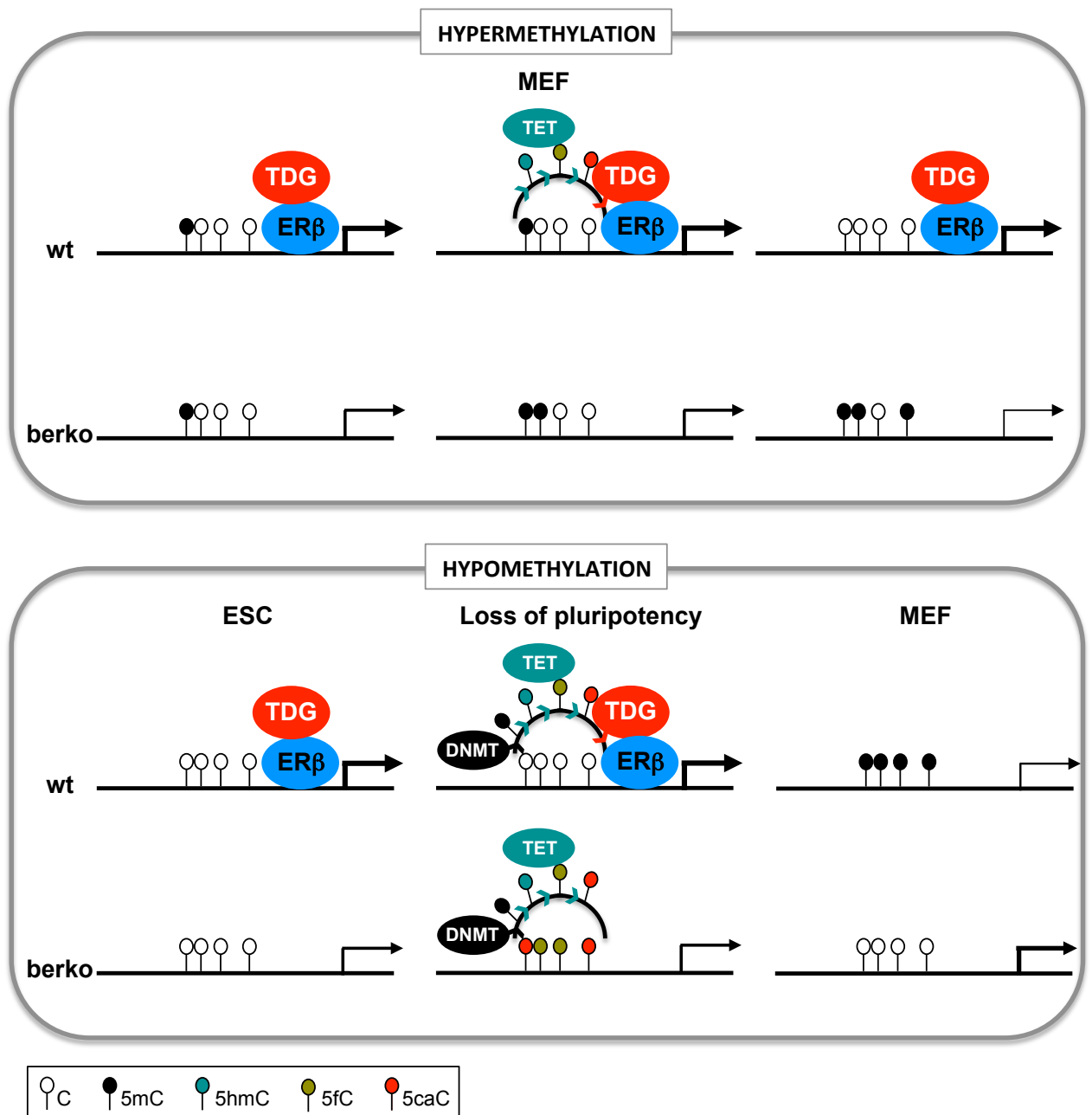


Figure 7



**Table 1: Genomic distribution of hyper- and hypomethylated differentially methylated positions (DMPs).**

	% in Genome	% in Hypermethylated	Enrichment in Hypermethylated	% in Hypomethylated	Enrichment in Hypomethylated
<b>Exonic</b>	5.5	19.1	3.5	48	8.7
<b>Intronic</b>	37.4	38.2	1	27	0.7
<b>Intergenic</b>	57.1	42.6	0.7	25	0.4
<b>Promoter</b>	2.6	10	3.8	30.8	11.8

**Table 2: Gene ontology (GO) enrichment analysis of genes whose promoter is either hyper- or hypomethylated.** Included genes show more than 30% difference in methylation between wt and  $\beta$ erko MEFs, and at least 30% of all promoter CpGs are covered by more than 4 reads in both cell types.**Hypermethylated genes**

GO-Term	P-value	# Genes (63 in total)
<b>Embryonic morphogenesis</b>	4.91E-04	5
<b>Embryonic appendage morphogenesis</b>	6.39E-04	3
<b>Embryonic limb morphogenesis</b>	6.39E-04	3
<b>Regulation of transcription, DNA-dependent</b>	7.62E-04	8
<b>Regulation of RNA metabolic process</b>	8.50E-04	8
<b>Transcription, DNA-dependent</b>	1.01E-03	8
<b>RNA biosynthetic process</b>	1.03E-03	8
<b>Appendage morphogenesis</b>	1.12E-03	3
<b>Limb morphogenesis</b>	1.12E-03	3
<b>Embryonic development</b>	1.19E-03	6

**Hypomethylated genes**

GO-Term	P-value	# Genes (223 in total)
<b>Developmental process</b>	5.29E-10	52
<b>Multicellular organismal development</b>	2.70E-09	48
<b>Anatomical structure development</b>	4.73E-09	44
<b>System development</b>	1.93E-08	41
<b>Regulation of transcription from RNA polymerase II promoter</b>	6.70E-08	20
<b>Regulation of transcription, DNA-dependent</b>	7.84E-08	27
<b>Regulation of RNA metabolic process</b>	1.10E-07	27
<b>Positive regulation of cellular biosynthetic process</b>	1.42E-07	20
<b>Transcription from RNA polymerase II promoter</b>	1.66E-07	20
<b>Transcription, DNA-dependent</b>	1.88E-07	27

**Table 3: Top ten tissues where hyper- and hypomethylated genes are enriched.**

Included genes show more than 30% difference in methylation between wt and  $\beta$ erko MEFs, and at least 30% of all promoter CpGs are covered by more than 4 reads in both cell types.

**Hypermethylated genes**

Tissue	P-value	# Genes (63 in total)
FORELIMB	7.52E-05	4
CYTOPLASMIC GRANULES	1.71E-04	3
PELVIS	6.73E-04	2
WING	8.42E-04	3
GRANULATION TISSUE	9.86E-04	3
LIMB BUDS	1.13E-03	4
TENDONS	4.96E-03	3
EXTREMITIES	5.17E-03	5
SYMPATHETIC FIBERS, POSTGANGLIONIC	5.19E-03	1
GINGIVAL CREVICULAR FLUID	5.78E-03	2

#### Hypomethylated genes

Tissue	P-value	# Genes (223 in total)
NERVE FIBERS	5.47E-04	6
SENSORY RECEPTOR CELLS	9.44E-04	9
SPINAL CORD	1.13E-03	14
ENDOCARDIUM	1.18E-03	4
NEURAL CREST	1.30E-03	9
EXTREMITIES	1.38E-03	12
MESODERM	1.88E-03	13
PERIPHERAL NERVES	2.09E-03	8
MESENCHYMAL STEM CELLS	2.28E-03	7
INTERNEURONS	2.30E-03	9

**Table 4: Top ten MeSH annotations for hyper- and hypomethylated genes.** Included genes show more than 30% difference in methylation between wt and  $\beta$ erko MEFs, and at least 30% of all promoter CpGs are covered by more than 4 reads in both cell types.

#### Hypermethylated genes

Disease	P-value	# Genes (63 in total)
Complex Regional Pain Syndromes	3.67E-05	4
Neuroaspergillosis	5.02E-05	2
Remission, Spontaneous	9.25E-05	6
Aphakia	1.44E-04	3
Neoplasm, Residual	2.59E-04	6
Clubfoot	3.92E-04	3
Hypocalcemia	5.08E-04	4
Humeral Fractures	5.93E-04	2
Encephalomyelitis	7.36E-04	4
Arm Injuries	8.50E-04	3

#### Hypomethylated genes

Disease	P-value	# Genes (223 in total)
Neurologic Manifestations	4.91E-05	59
Neoplasms, Nerve Tissue	8.69E-05	67
Neoplasms, Neuroepithelial	8.93E-05	56
Neuroectodermal Tumors	9.51E-05	66
Neuroectodermal Tumors	9.82E-05	66
Neurologic Manifestations	1.20E-04	63

<b>Neoplasms, Ductal, Lobular, and Medullary</b>	1.35E-04	29
<b>Neuroectodermal Tumors, Primitive</b>	2.50E-04	40
<b>Syncope, Vasovagal</b>	2.55E-04	4
<b>Urogenital Neoplasms</b>	2.69E-04	61

**Table 5: Features of validated differentially methylated positions (DMPs).** Genomic location, presence of a CpG island (CGI), and comparison with datasets for histone modifications enriched at promoter or enhancer regions in MEFs and embryonic stem cells (ESCs) using GenomeInspector (Genomatix). Hypo and hyper refer to hypo- and hypermethylation, respectively.

DMP	Genomic location	CGI	Promoter MEFs	Promoter ESCs	Enhancer MEFs	Enhancer ESCs
<b>Hypo1</b>	Intergenic	No	H3K4m3/ H3K27m3	H3K4m3	H3K4m1	
<b>Hypo2 (Dyx1c1)</b>	Promoter region	Yes	H3K4m3	H3K4m3		
<b>Hypo3</b>	Intragenic (intron)	No			H3K4m1	
<b>Hypo4</b>	Intragenic (intron)	No			H3K4m1	
<b>Hypo5</b>	Intragenic (intron)	No	H3K4m3		H3K4m1/ H3K27ac	H3K4m1
<b>Hypo6</b>	Promoter region	No	H3K27m3	H3K27m3	H3K4m1	H3K4m1/ H3K27ac
<b>Hypo7</b>	Intragenic (intron)	Yes	H3K4m3/ H3K27m3	H3K27m3		H3K4m1
<b>Hypo8 (HoxD9)</b>	Promoter region	Yes	H3K4m3/ H3K27m3	H3K4m3/ H3K27m3	H3K4m1	
<b>Hypo9</b>	Promoter region	Yes	H3K4m3/ H3K27m3	H3K4m3	H3K4m1/ H3K27ac	
<b>Hypo10</b>	Promoter region	Yes	H3K4m3/ H3K27m3	H3K4m3/ H3K27m3	H3K4m1/ H3K27ac	H3K4m1
<b>Hyper1 (HoxA9)</b>	Promoter region	No	H3K4m3	H3K4m3/ H3K27m3		
<b>Hyper2 (HoxA10)</b>	Promoter region	No	H3K4m3/ H3K27m3	H3K27m3		H3K4m1
<b>Hyper3</b>	Promoter region	No	H3K4m3/ H3K27m3	H3K27m3		H3K4m1
<b>Hyper4 (Tnfrsf2)</b>	Promoter region	Yes	H3K4m3	H3K4m3	H3K4m1	H3K4m1
<b>Hyper5</b>	Intragenic (intron)	No				
<b>Hyper6 (Pitx1)</b>	Promoter region	Yes	H3K4m3/ H3K27m3	H3K4m3/ H3K27m3		H3K4m1
<b>Hyper7</b>	Intragenic (intron)	No				
<b>Hyper8</b>	Intergenic	Yes	H3K4m3	H3K4m3	H3K4m1	

# Claudia Krawczyk

**Date of Birth** 28<sup>th</sup> May 1981

**Nationality** German

## Education

---

- 12/2009 – 10/2014** **Ph.D. in Molecular Biology (Genetics)**, Department of Biomedicine, University of Basel  
Thesis: "Novel Functional Aspects of Topoisomerase Top1 and DNA Glycosylase Thp1 in the Maintenance of Genetic and Epigenetic Stability in Yeast"
- 10/2006 - 09/2009** **M.Sc. in Molecular Biology**, Department of Biomedicine, University of Basel  
Thesis: "Dissection of Specific Functions of the two DNA Helicases Mph1 and Sgs1 in the Stabilization of Stalled Replication Forks"
- 09/2002 - 07/2006** **B.Sc. in Molecular Biology**, Fachhochschule Gelsenkirchen, Germany (Practical part of the Bachelor thesis at the Institute for Cell Biology, University Hospital Essen, Germany)  
Thesis: Subcellular localization of the potential tumour suppressor HNF4 $\alpha$
- 08/2000 - 07/2002** **Apprenticeship as a bank clerk** ("Bankkauffrau"), Deutsche Bank AG, Essen, Germany, 2-year apprenticeship in all areas of banking
- 06/2000** **Abitur** (German university entrance qualification), Haranni-Gymnasium, Herne, Germany

## Publications

---

**Krawczyk, C.**, Dion, V., Schär, P. and Fritsch, O. (2014); Reversible Top1 cleavage complexes are stabilized strand-specifically at the ribosomal replication fork barrier and contribute to ribosomal DNA stability. *Nucleic Acids Res.* 42, 4985-4995

**Krawczyk, C.\***, Bentele, M.\*, Fritsch, O., Latypov, V., Woolcock, K., Ivanek, R., Noreen, F., Bühler, M., Schär, P; Uracil Repair Causes DNA Glycosylase-Dependent Genome Instability; Manuscript in preparation (\* contributed equally to the work)

Duong, W., **Krawczyk, C.**, Bretschneider, N., Zinsler, C., Schär, P., Rüegg, J.; Estrogen Receptor  $\beta$  Regulates Epigenetic Patterns at Specific Genomic Loci through Interaction with Thymine DNA Glycosylase; Manuscript in preparation