

TDG-mediated Active DNA Demethylation in the Genome-wide Control of Transcriptional Initiation and Elongation

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Simon Schwarz

aus

Buchs SG und Zürich ZH, Schweiz

Basel, 2021

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel
edoc.unibas.ch

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Primo Schär (Fakultätsverantwortlicher und Dissertationsleiter)

Prof. Dr. Renato Paro (Korreferent)

Basel, den 21.09.2021

Prof. Dr. Marcel Mayor
Dekan der Philosophisch-Naturwissenschaftlichen Fakultät

Content

Acknowledgements	6
Frequent Abbreviations.....	7
1 Summary	9
2 Introduction	13
2.1 Basics of Transcription	13
2.1.1 Transcriptional Initiation at Enhancers and Promoters.....	14
2.1.2 Elongation, Splicing and Termination	16
2.2 DNA Base Excision Repair and the Thymine DNA Glycosylase	18
2.2.1 General Mechanisms of Base Excision Repair	19
2.2.2 The Thymine DNA Glycosylase TDG.....	21
2.3 Epigenetic Modifications and their Regulation.....	25
2.3.1 DNA Methylation	25
2.3.2 Modification on Chromatin Proteins	29
2.4 Interplay between DNA Methylation, Histone Modification and Transcription	33
2.4.1 Epigenetic Rearrangements in Stem Cells	33
2.4.2 Epigenetic Control of Repetitive Elements	35
3 Aim of the Thesis.....	38
4 Results & Conclusions	39
4.1 PARP Inhibition Induces Cytotoxicity in mESCs by Activating Endogenous Retroviruses.....	39
4.1.1 Summary.....	39
4.1.2 Contribution.....	44
4.1.3 Manuscript, see appendix I.....	45
4.2 Covalent PARylation is an Integral Part of TDG-BER-mediated Active DNA Demethylation in Embryonic Stem Cells.....	46
4.2.1 Summary.....	46
4.2.2 Contribution.....	50
4.2.3 Manuscript, see appendix II.....	51

4.3	Active DNA demethylation Evicts H2A.Z to Facilitate Pause Release of RNA Polymerase 2	52
4.3.1	Contribution	52
4.3.2	Manuscript, see appendix III	52
4.4	Additional preliminary data: Intragenic DNA methylation, TDG and SETD2	53
4.4.1	H3K36me3 Participates in Recruitment of TDG to the Gene Body but not the Processing of oxmCs	54
4.4.2	TDG and SETD2 are Required for Transcriptional Efficacy.....	55
4.4.3	Loss of SETD2 Affects Recruitment of TDG to the Chromatin	56
4.4.4	TDG and SETD2 Double Knockout Affects Cell Viability.....	58
5	Conclusion and Outlook	60
6	References.....	68
7	Appendix	87
	I – PARP Inhibition Induces Cytotoxicity in mESCs by Activating Endogenous Retroviruses	
	II – PARP1 Promotes BERosome Disassembly To Facilitate DNA Demethylation	
	III - Active DNA demethylation Evicts H2A.Z to Facilitate Pause Release of RNA Polymerase 2	
	IV - Inducible TDG Knockout Models to Study Epigenetic Regulation	

Acknowledgements

My sincerest thanks go to Primo Schär for allowing me to undertake this endeavour. His patience and trust as well as his support throughout these years are invaluable. I really appreciate the fruitful exchanges and his unwavering motivation for curiosity-driven research and therefore being a great teacher. Despite his many office-related duties he remained approachable and involved.

I also want to thank David Schürmann for being my mentor and his constant support whenever I needed him. With his criticism that was always constructive, he helped me to find my way in the laboratory and showed a lot of patience in correcting and advising me on writing of this thesis as well as the manuscripts.

Special thanks go to Jianming Xu and Roland Steinacher for involving me in their projects and being challenging as well as nurturing counterparts in scientific discussions.

Uncountable thanks also go to the other past and present members of the lab who kept it running, who supported me and generally tried – and succeeded – in making this time a great and impressive experience.

Certainly, I would also like to thank Renato Paro and Mihaela Zavolan for being in my committee and putting up with the many delays.

Increasing the scope, I want to thank the whole Mattenstrasse 28 as well as the remainder of the DBM locations for shaping an extremely enjoyable institute. Particular thanks go to the “crewmates” of the DBM PhD Club who, over these years, shared the motivation to increase the overall quality of being a PhD at this institute.

Great times come with great people and Friday’s “Bier um Vier” certainly brought many of them together. I thank all who participated and brought fun to these gatherings in- and outside of the institute, helping to find some energy after long days and -oh so often - before weekends of work.

The pandemic unfortunately restricted the size of these groups but also distilled an extremely important crew, of which I am very thankful to be part of. Cordially and very gratefully, I want to acknowledge the “Bubbles”, who were and are great friends and without whom, the struggle of the last 1.5 years would have become difficult to bear.

Likewise, I want to thank all my friends outside the institute, from the Rhine valley over Zürich to Basel, who shared their time with me. Be it in the park, at the Rhine, under water or climbing mountains, they managed to take my mind of the pressure and supported me in many ways.

Finally, I want to express my gratitude towards my family for being a safe harbour. Their encouragement and unquestioning support during this time were extremely valuable to me. Coming home and being around my kin, including the youngest ones, brings me a lot of joy.

Thank you all, very very much -
and stay curious!

Frequent Abbreviations

A Adenine

AP Apurinic/aprimidinic

APE1 AP endonuclease 1

BER Base excision repair

bp Base pair

C Cytosine

5caC 5-carboxylcytosine

CpG Cytosine - Phosphate - Guanine dinucleotide

CGi CpG island

CBP/p300 CREB-binding protein and homologue p300

DNA Deoxyribonucleic acid

DNMT DNA methyltransferase

dRP 5'-deoxyribose-phosphate

DSB Double-strand break

mESC Murine embryonic stem cell

5fC 5-formylcytosine

G Guanine

G•T Base (mis)pairing between Guanine and Thymine

GBM gene body methylation

H3K4/9/27 Histone 3 lysine 4/9/27

h hours

5hmC 5-hydroxymethylcytosine

5hmU 5-hydroxymethyluracil

IP Immunoprecipitation

KO knockout

LIG3 DNA ligase 3

LINE Long interspersed nuclear element

LTR Long terminal repeat

5mC 5-methylcytosine

min minutes

NER Nucleotide excision repair

oxmC oxidised methylcytosine (5hmC/5fC/5caC)

p53 (tumour) Protein 53

PAR/MAR Poly-ADP-ribose / Mono-ADP-ribose

PARP1 Poly-[ADP-ribose] polymerase 1

PI Pause index

POL β DNA polymerase β

RE Repetitive Element

RNA Ribonucleic acid

RNAP2 RNA-polymerase 2

SETD2 SET domain containing 2

SINE Short interspersed nuclear element

SUMO Small ubiquitin-like modifier

SSB Single-strand break

T Thymine

TE Transposable element

TDG Thymine DNA glycosylase

TET1-3 Ten-eleven-translocation protein 1-3

TSS Transcription start site

TES Transcription end site

U Uracil

WT Wild type

XRCC1 X-Ray Repair Cross Complementing protein 1

XPC Xeroderma Pigmentosum Group C Protein

1 Summary

The transcription of DNA into RNA is key to convert the genetic code - a linear sequence of the four “letters” of the genetic alphabet - into three-dimensional biological structures with functions that are central to create and maintain life. Differential regulation of transcription allows the use of a single genic locus in multiple ways (gene-variants), allowing cells to respond to different circumstances such as the surrounding environment, the stage in development of an organism or even the circadian clock. The possibility of this adaptive use of a gene vastly increases the functional spectrum contained in a single coding region, but also requires a complex regulatory system to read and convey the right information at the right time. One way of how this regulation is achieved, is being studied in the field of epigenetics. This research addresses covalent but reversible modifications to DNA or DNA-packaging proteins that are able to influence transcriptional levels. As these modifications can persist throughout multiple cell cycles and are actively passed on to daughter cells, epigenetic marks are being considered as heritable without actually altering the genetic code itself.

Study of the DNA repair enzyme Thymine DNA glycosylase (TDG) revealed that DNA excision repair contributes to an epigenetic mechanism of transcriptional regulation, as TDG is able to indirectly remove methylated bases from the DNA (Jacobs and Schär, 2012; Schuermann et al., 2016). Beside G•T and other non-canonical DNA mispairings and base modifications, TDG can recognise and excise 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) from DNA, both of which are products of 5-methylcytosine (5mC) oxidation by ten-eleven translocation proteins (TETs, He et al., 2011; Ito et al., 2011). Methylation marks on DNA, placed on a C followed by a G (CpG), have been shown to affect transcription in many ways. The most direct and intuitive way is by the inhibition of transcription factor binding at gene promoters. Methylated cytosines thereby interfere with transcriptional initiation and the final expression level of a gene. TET and TDG-mediated active DNA demethylation can, in this case, remove these 5mCs and render the promoter more accessible to the transcription machinery and restore expression levels. This seemingly rather clear mechanistic concept of active DNA demethylation through oxidation and base excision could be shown on the biochemical level five years ago (Weber et al., 2016), but *in vivo* data supporting the whole process including TDG and the DNA repair part is rather scarce and/or convoluted. It is most likely attributable to the intrinsic dynamicity of this multistep process and the vast number of interactions of the active DNA demethylation machinery with other epigenetic regulators that remain to be untangled.

The Schär group and others reported that TDG depletion, unlike any other DNA glycosylase, leads to the death of mouse embryos around day E10.5 (Cortázar et al., 2011; Cortellino et al., 2011). This phenotype was not just accompanied by alterations in DNA methylation, but (even more so) by altered histone modifications, and consequential changes of RNA expression during differentiation. Other groups then described intermediates of active DNA demethylation being present outside of promoters and all over the

genome of embryonic stem cells (ESCs, Raiber et al., 2012; Shen et al., 2013), thereby increasing the complexity of the epigenetic landscape and its regulation by TDG (Kolendowski et al., 2018). A central but until now least-understood concept, before TDG can be fully established as an omni-present tool for transcriptional regulation, is the following: oxidation of 5mC through TET enzymes, followed by TDG-mediated excision and BER, provides an unmodified C that can again be methylated (Weber et al., 2016). This proposes the possibility that 5mC species can undergo multiple rounds of methylation and demethylation in the same cell, rather than being a stable mark (reviewed in Parry et al., 2020). This brings two further challenges into the research of TDG and active DNA demethylation: First, capturing of the effective methylation state is difficult/not meaningful as the dynamics of the (de-)methylation of a CpG have to be measured. And second, not just the 5mC species need to be investigated in their interaction with DNA-binding factors but also all of the constantly present DNA demethylation factors, too – as well as their interactors. A very recent study addressed the first aspect and measured methylation and oxidation rates of CpGs and could indeed show that active 5mC oxidation exceeds passive demethylation by far and that CpGs with different rates of oxidation, i.e. active DNA demethylation, associate with different genomic locations, transcriptional levels and histone modifications (Ginno et al., 2020). While this nicely showed the flexibility of TET-mediated 5mC oxidation, the further modulation of/by processing via TDG and BER to control transcription remains to be addressed.

Although TDG was described to be beneficial for transcriptional activation of certain genes, it was only rarely shown for multiple targets and often lacked the direct link to the actual presence and activity of TDG (Hassan et al., 2017, 2020; Hu et al., 2014; Léger et al., 2014). In order to contribute filling that lack of information, I studied TDG and its effect on transcriptional regulation in murine ESCs, which are a very dynamic experimental model with high activity of active DNA demethylation, and applied locus-specific as well as genome-wide experimental procedures.

In a collaborative effort, I could show for the first time that TDG (and TET) is essential in the mediation of cytotoxicity upon a stress signal that is not caused by base-analogues as substrates for TDG (Kunz et al., 2009), but by the interference of the BER pathway downstream of TDG (see 4.1/appendix I). We observed that treatment of naïve ESCs, cultivated in 2i-medium, are extremely sensitive to an inhibitor of the DNA repair protein Poly-[ADP-ribose] polymerase 1 (PARP1) called Talazoparib (Tal). PARP-inhibitors are usually used in treatment of cancer cells that are defective in homologous repair (HR). There, inhibition of repair leads to an accumulation of SSBs, which, for example by collapse of replication forks, result in an overload of toxic DSBs that kill the HR-deficient cells. It should therefore not affect healthy (somatic) cells that are HR-proficient. The toxic effect of Tal in ESCs, which are highly proficient in HR, was therefore surprising and demanded further examination. We could show that the Tal-mediated cell death is not based on the increased generation of DSBs, but on the transmission of the damage signal via the damage sensor and

transcription factor, (tumour) protein p53. Depletion of TDG did not reduce the level of DSBs, but drastically reduced expression of all examined p53 target genes that were upregulated upon the treatment with Tal. Among these genes were classical mediators of apoptosis, like *Fas* (Fas Cell Surface Death Receptor), *Alox5* (Arachidonate 5-Lipoxygenase) or *Tap1* (Tocopherol-associated protein 1), indicating that TET/TDG-mediated active DNA demethylation is necessary to transmit a systemic stress response that includes at least 1500 genes. Furthermore, did we observe an increased transcription of several thousand repetitive elements upon the treatment with Tal that was also drastically reduced after the depletion of TET, TDG or p53. The expression of these elements, notably caused an upregulation of genes involved in necroptosis and an interferon-like response, very likely contributing to the observed cell death. Reduction of these transcriptional responses in absence of TDG additionally correlated with a reduced amount of generated SSBs, hinting towards a link of TDG-mediated excision of 5fC and 5caC and expression.

Together, I could clearly underline the essential role of TDG activity as a general factor in the toolbox of transcriptional regulation, by showing that the Tal-mediated activation of these p53-targets is depending on the presence of TDG and correlates with increased SSBs in TDG-proficient cells.

In the Tal-mediated stress response, we saw that the recruitment of p53 to its targets genes is slightly impaired upon the deletion of TDG but that the chromatin accessibility at these promoters, which is usually well correlated with transcriptional initiation, is not altered. Based on this and the fact that oxmCs as well as TETs and TDG are frequently found after the transcription start site (TSS) as well as throughout the whole gene body, I aimed to address transcriptional elongation in dependence of active DNA demethylation. Performing a precision run-on sequencing (PRO-seq), I observed that transcriptional initiation in TDG depleted ESCs is not significantly changed, but the pause release of RNA polymerase II (RNAP2) is clearly reduced (see 4.3/appendix III). Interestingly, this was true for the stress induction by Tal but also in unchallenged TDG KO ESCs and indicates that TDG and active DNA demethylation are involved in another step of transcriptional control, in addition to promoter demethylation. To figure out what mechanism leads from active DNA demethylation to pause release of RNAP2, I analysed the histone variant H2A.Z. H2A.Z was described as an important factor in facilitating transcriptional initiation by its presence upstream of the TSS, but at the same time hindering productive transcriptional elongation after the TSS (Giaino et al., 2019; Mylonas et al., 2020). By analysing published datasets of (acetylated) H2A.Z, I could correlate the abundance of this histone variant with sites of active DNA demethylation, SSBs and the release of paused RNAP2. In combination with a dataset that was previously generated in our laboratory, I could show that H2A.Z levels are strongly increased in ESCs depleted of TDG and that genes with an increased level of H2A.Z at their TSS also display a reduced release of paused RNAP2. This implies a role of TDG in the eviction of H2A.Z during RNAP2 pause release. Furthermore, did I observe that not only transcriptional elongation of genes is hindered in TDG null ESCs, but also that transcription at enhancers is impaired. Although this was

recently published (Kolendowski et al., 2018), neither the connection of transcriptional elongation and enhancer activity, nor the crucial abundance of H2A.Z at enhancers (Brunelle et al., 2015), was investigated in dependence of TDG. These observations, however, need further verification but are well supported by current literature and provide a solid basis on the description of how TDG-mediated active DNA demethylation affects transcriptional elongation via H2A.Z (Schaukowitch et al., 2014; Williams et al., 2015).

Last but not least, I could contribute to the description of a novel step in the processing of DNA lesions by BER. Combination of the genome-wide data set of SSBs in murine ESCs, with newly obtained biochemical data in our laboratory, allowed to show that modification of BER-enzymes by PARP1 facilitates the initial generation of SSBs by TDG and the abasic site endonuclease APE1, by increasing the dissociation of BER factors after processing of their substrate. This adds a new level of PARP1-mediated regulation of BER in addition to its function in general SSB repair, as was previously assumed (see 4.2/appendix II, Fisher et al., 2007; Hanzlikova et al., 2016). Focussing the analysis on genomic locations that are not responding to the stress signalling evoked by Tal, I could show that PARP-inhibition leads to a reduction of SSBs mediated by active DNA demethylation. These sites that undergo targeted DNA demethylation are contrasted by sites with increased genomic instability (e.g., simple repeats), which gain SSBs upon inhibition by Tal because of PARP1 failing to repair spontaneously occurring damage. This data contributes to the understanding of random vs. instructed DNA damage in the genome and thereby again fortifies the role of active DNA demethylation as a targeted process in ESCs that is constantly ongoing.

Taken together I could underline that TDG is not just an accessory protein to TETs for the excision of oxmCs but that on top of this role, it possesses its own essential functions necessary for transcriptional regulation and contributes to the epigenetic landscape in a fundamental way.

2 Introduction

2.1 Basics of Transcription

Transcription is, along with translation and DNA replication, a fundamental genetic process that allows biological entities to regulate metabolism, homeostasis and finally their reproduction – therefore also being a prerequisite in the definition of life (e.g. Tirard et al., 2010; Vitas and Dobovišek, 2019). In addition to the combinatorial variety of DNA sequences, the possibility of temporal and cell-specific regulation of transcription is not only key to the generation of the variety in life forms, but also to the understanding of how the human achieves its complexity with just about ~23'000 genes, considering that the simple nematode *Caenorhabditis elegans* is built by a similar number of genes (The *C. elegans* Sequencing Consortium, 1998).

Transcription comprises the reading of DNA, the primary molecule of information containing the blueprint of an organism, to create a specific copy in form of ribonucleic acid (RNA) (see Figure 1A and B). So-called messenger RNAs (mRNAs) serve as the information source for translation, which finally produces proteins with catalytic and structural functions required for virtually all processes in life.

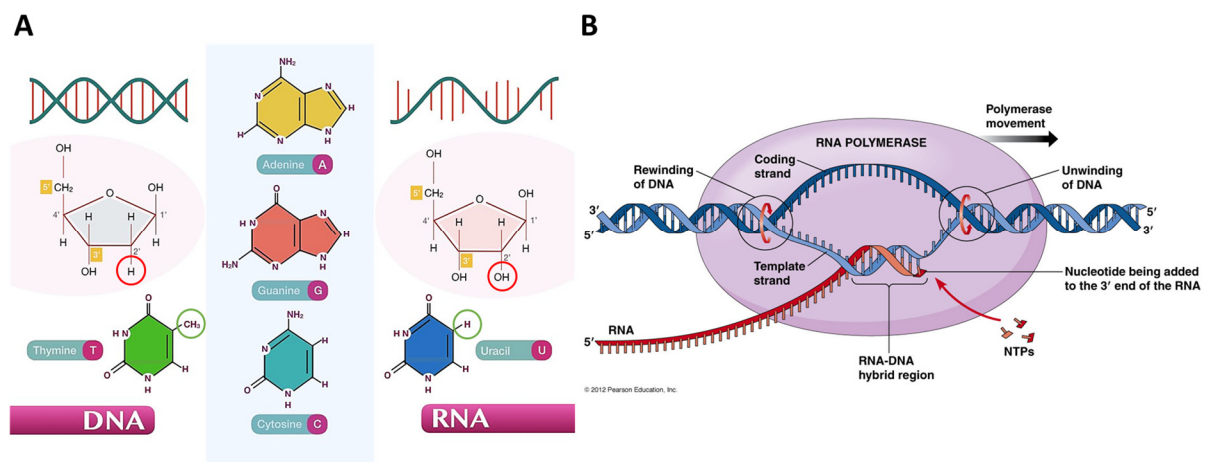


Figure 1 Basics of transcription

A Differences in the composition of DNA and RNA are found in the backbone, where DNA lacks an oxygen atom (red circles) and the use of an alternative base called uracil, lacking a methyl group (green circles) when compared to thymine that is used in DNA. From <https://geneticeducation.co.in/dna-vs-rna-differences-and-similarities> (Last Accession 18.02.2021) **B** Scheme of transcription by the RNA polymerase 2, generating a single-stranded mRNA molecule from the double-stranded template DNA. From https://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-21/21_09.jpg (Last Accession 22.02.2021)

Aside from mRNAs, produced by the RNA polymerase 2 (RNAP2), there is a wide variety of RNAs that are not translated into proteins and therefore called non-coding RNAs (ncRNAs). RNAP1 & 3 are responsible for the transcription of the majority of RNAs, including non-coding ribosomal (rRNA) and transfer RNAs (tRNA), which are essential components of the gene expression system constituting around 80-85% and 10-15% of the total RNA mass in a cell, respectively (Viktorovskaya and Schneider, 2015). Other types of ncRNAs have been found to participate in the regulation of transcription in various ways, for example by increasing the

recruitment of the Mediator complex to enhancers, in the degradation of mRNA via RNA interference or even as ribozymes with enzymatic activity to catalyse splicing (reviewed in Cech and Steitz, 2014).

With all its regulatory components, transcription facilitates differential “reading” of the ~23’000 genes in a human cell and the creation of an even higher number of mRNA variants and hence, gene variants, which are eventually being translated into cell type-specific proteins. Transcription is therefore the first step to reach a cellular “toolbox” of proteins and regulatory RNAs that are prerequisite for the generation of complex gene expression patterns in tissues and whole organisms (Pal et al., 2011).

2.1.1 Transcriptional Initiation at Enhancers and Promoters

The transcription of a gene generally depends on the cooperation of four functional components, that are 1) the instructive and regulatory DNA elements called promoter and enhancers, 2) the RNA polymerase II, 3) general transcription factors (GTFs) and 4) gene-specific transcription factors (TFs) (including activators and repressors). GTFs such as Transcription Factors A-H (TFIIA-H) are essential to establish the transcription pre-initiation complex (PIC) and mediate binding of RNAP2 to the promoter sequence, while gene-specific TFs selectively integrate regulatory signals from enhancers and/or other signalling molecules to individual genes (Orphanides et al., 1996). Core promoters, representing the DNA binding sites for GTFs, are usually found within -500 bp to +30 bp of the transcription start site (TSS) of a gene, where the synthesis of mRNA starts (Cooper, 2005; Haberle and Stark, 2018). The functional promoter of a gene is, however, considered to span hundreds of base pairs (up to several kbs) upstream to the TSS and can comprise regions with additional regulatory functions, including CpG-islands and binding sites of gene-specific TFs to further modulate the transcription initiation frequency.

While a core promoter together with a PIC is sufficient to reach basal levels of transcription, sometimes producing incomplete/abortive transcripts, it also requires the assistance of enhancers that can be found kilobases to megabases away from the actual gene promoter. Enhancer-promoter interactions promote the loading of new PICs to the promoter and enforce transcriptional elongation by the recruitment of co-transcription factors like p300 and the Mediator complex to release paused RNAP2 into productive transcription (Field and Adelman, 2020). The currently accepted sequence of events of transcription initiation starts with the removal of nucleosomes at the TSS by chromatin remodellers (also see chapter 2.3.2), guided by gene-specific TFs and other recruiting signals. This allows for the binding of GTFs, which in turn recruit RNAP2 to the core promoter (and the enhancer, Figure 2A). The assembly of the PIC positions the RNAP2 on the DNA and prepares the RNAP2 holoenzyme for activation by phosphorylation at the Serine-5 position in the amino-acid heptades of the C-terminal domain (CTD). The initiated RNAP2 will then start RNA synthesis, but will be paused in the region 3’ of the TSS through the association with the negative elongation factor (NELF) or DRB-sensitivity-inducing factor (DSIF), which inhibit productive elongation (Schaukowitch et al., 2014). This promoter-proximal pausing is not static but includes a steady-state of

constant initiation, pausing and backtracking of RNAP2 which creates short, 30-50 bp long, RNA molecules. The short RNAs can form RNA-DNA hybrids generating a ternary structure of the hybrid and the displaced ssDNA, called R-loop. The function of these promoter proximal R-loops is not very well understood, but it has been proposed, for instance, that they protect nearby CpG-islands from *de novo* methylation by DNMT3s via reinforcement of H3K4 trimethylation (also see chapter 2.3.2, Ginno et al., 2012) or maintain an increased accessibility by promoting antisense transcription (Tan-Wong et al., 2019). Short RNAs are also produced at enhancer elements, but their function is not very well described, either. While initially thought to be only by-products of high RNAP2 occupancy at these regions, current evidence hints towards a coordinated process, specific to active enhancers, mediating release of paused RNAP2 via the interference with NELFs (Arnold et al., 2020; Schaukowitch et al., 2014).

The transition of paused to elongating RNAP2 occurs through the interaction with co-transcription factors like Bromo-domain Containing 4 (BRD4), the super elongation complex (SEC) or the Mediator complex. These activate the positive transcription elongation factor-b complex (P-TEFb), including the cyclin-dependent kinase 9 (CDK9), to phosphorylate pausing factors and the CTD of RNAP2 at Serine-2. These activation events cause the dissociation of the pausing factors and promote productive elongation of the transcript (Figure 2B, reviewed in Haberle and Stark, 2018; Jonkers and Lis, 2015). Transcriptional elongation then occurs in bursts, whereby promoter size and sequence context define the burst size, i.e. how many elongation complexes are set free, and the enhancer affects the burst frequency (Larsson et al., 2018).

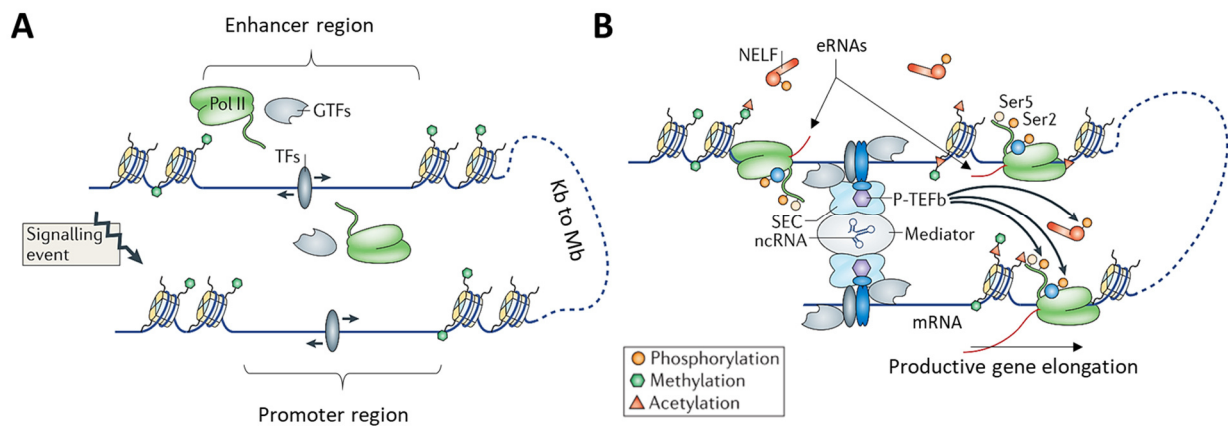


Figure 2 Transcription initiation and release of paused RNAP2

A Assembly of the PIC by binding of transcription factors to promoter and/or enhancer elements to allow for binding of RNAP2. **B** Paused RNAP2 is activated by P-TEFb in association with transcription co-factors super elongation complex (SEC) and Mediator. Serine 2 (Ser2) of RNAP2 becomes phosphorylated and so are NELFs to trigger their eviction. Enhancer RNAs (eRNAs) are produced, correlating with transcriptional activity of the gene. Adapted from Jonkers and Lis, 2015

2.1.2 Elongation, Splicing and Termination

Once the RNAP2 is released from the paused state, transcription does not proceed unhindered but may encounter obstacles that trigger a variety of processes. A major obstacle for progressive transcription is the abundant presence of histone proteins, around which the DNA is wound and the RNAP2 has to pass through. Processive RNAP2 is sufficient to remove H2A-H2B histone dimers to allow passage around the histone core. The speed of this process is however limited and for higher transcription efficiency, RNAP2 can associate with chromatin remodellers that completely or partially destabilize and evict the histone proteins to facilitate passage (see also chapter 2.3.2, Kulaeva et al., 2013; Venkatesh and Workman, 2015). These evicted histones have to be restored again to maintain chromatin conformation. The entire process is not fully understood but involves histone chaperones like Facilitator of Transcription (FACT) or Histone Cell Cycle Regulation Defective Homolog A (HIRA) that replace histones H2A-H2B or H3.3, respectively, and the remodeller Chromodomain Helicase DNA Binding Protein 1 (CHD1) that controls spacing of the histones (Figure 3A, step a/c).

In regions of high transcription, histones are very distinctly modified. Trimethylation at the H3 Lysine 36 residue (H3K36me3) is deposited through the interaction of the histone methyl transferase SET Domain Containing 2 (SETD2) with the phosphorylated CTD domain of RNAP2. This histone mark prevents the inclusion of acetylated histones and removes already present acetylation marks (Figure 3A, step b/d) that could otherwise increase histone exchange and therefore hamper the transcriptional process. H3K36me3 is also shown to functionally interact with FACT, affecting the mRNA splicing machinery, and to reduce initiation of cryptic transcripts, possibly by recruiting DNMT3s and thereby setting DNA methylation marks (reviewed in Daniel and Strahl, 2017). This exclusive association of H3K36me3 with active transcription makes it a landmark for transcription and its short-term presence after transcriptional inhibition could even be seen as “memory” of past transcriptional activity.

The obstacles an elongating RNAP2 can encounter are not evenly distributed throughout the gene body and can affect transcriptional progression, represented by increased detection frequency (i.e. decreased progression) of RNAP2, for example by ChIP-seq (Figure 3B, Hodges et al., 2009; Veloso et al., 2014). The combination of chromatin modifications and their impact on chromatin structure and RNAP2 behaviour also affects the essential co-transcriptional process of splicing (Figure 3C). mRNA splicing excises the non-coding parts of the primary transcript, the so-called introns, to finally make up the mature mRNA with a defined coding sequence (CDS) for the protein to be translated. In general, fast transcription speed correlates with the omission of splice sites, while spatial hindrance of elongation associates with the use of the splice site. Notably, RNAP2 generally slows down when traversing from an intronic sequence into an exon, presumably to facilitate co-transcriptional splicing. Splice sites can intrinsically be “weak” or “strong” by their sequence context, whereas this characteristic can be further modulated by local chromatin

features, such as DNA methylation, to lower or increase the recognition of the splice site by the splicing machinery (Figure 3D, Tellier et al., 2020). Splicing also allows for the alternative use of exons that can majorly change the structure as well as the function of the protein that will be produced (Birzele et al., 2008). In addition to the transcriptional control by regulatory DNA elements, which mostly affects the number of transcripts produced, splicing adds another layer of fine-tuning to transcription, allowing for generation of multiple protein variants with distinct functions that can be produced from one single gene.

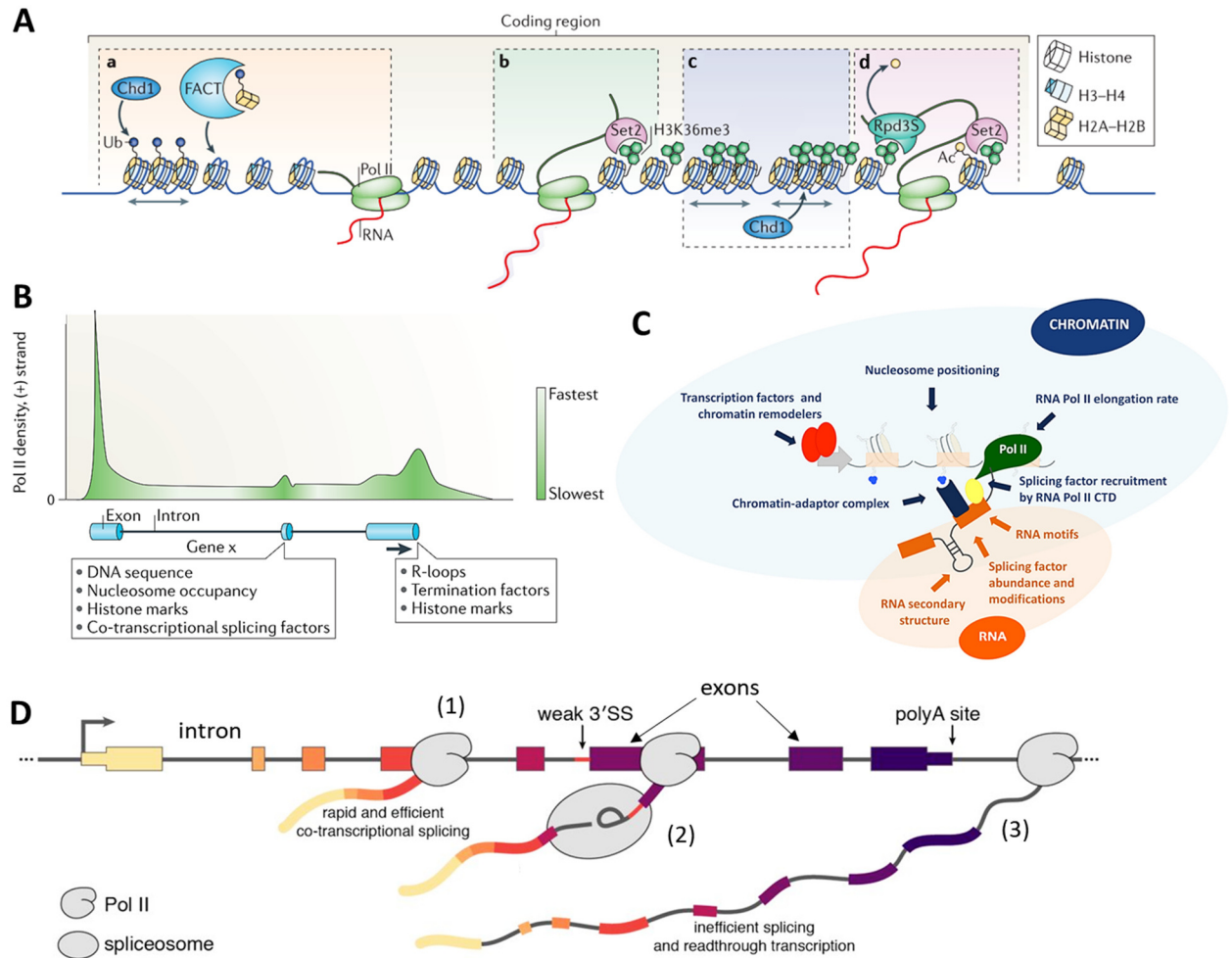


Figure 3 Transcriptional elongation and splicing

A Histone modifications and rearrangements associated with transcriptional elongation. Evicted histones need to be replaced and properly spaced (a&c). H3K36me3 is deposited by Set2/SETD2, marking active transcription and facilitating further transcription by the repulsion of histone acetylation (b&d). Adapted from Venkatesh et al. 2015

B Schematic RNAP2 detection intensity (e.g. by ChIP-seq) is affected by exemplified factors. From Jonkers and Lis 2015

C Schematic of factors that affect splicing. From Luco 2011

D (1) RNAP2 proceeds through the gene body and co-transcriptional splicing is faithful. (2) Weak splice sites displaying no additional modifications can barely slow down RNAP2 and require splicing after the polymerase has already passed the region. (3) If the splicing machinery cannot be properly recruited, transcription reads through the gene body and leads to non-functional transcripts that are most likely degraded. Adapted from Reimer et al. 2020

When transcription is ongoing, only major obstacles or signals can bring the elongating RNAP2 to a halt, dissociate it and terminate transcription. In almost all eukaryotes, transcription termination of nearly all genes is mediated by the pathway that revolves around the Cleavage and Polyadenylation Factor (CPF). An

initial step is a pausing of the RNAP2 that is evoked by either DNA-binding proteins or special termination signals like secondary structures of the nascent RNA. The most prominent and essential termination signal, however, is the poly-adenylation signal or PAS (core sequence: AAUAAA) that follows the coding sequence of a gene in its 3' untranslated region. When RNAP2 passes through and transcribes a PAS, the CPF complex is then recruited to the poly-A RNA and affects the RNAP2 to pause it 18-30 bp downstream of the PAS. This pausing coincides with the cleavage of the nascent RNA from RNAP2, followed by polyadenylation of the mRNA to prepare it for translation. The remaining snippet of RNA at RNAP2 is degraded by the 5'-3' Exoribonuclease 2 (XRN2), a process thought to also dissociate the RNAP2 from the DNA. Details on the precise timing and modes of regulation of transcriptional termination are yet to be discovered (Calvo, 2018; Porrua et al., 2016).

Gene transcription is a very basic process of life and has evolved in its complexity along with the demands of increasingly complex organisms. Proper functioning of metazoan cells require a highly orchestrated control of gene usage to ensure that not just the process of transcription itself is nominally executed, but also that levels and variants of transcriptional output are appropriate to the tissue and the environmental circumstances.

2.2 DNA Base Excision Repair and the Thymine DNA Glycosylase

The DNA base excision repair (BER) pathway provides protection of DNA against the potential mutagenic effect of endogenous and exogenous sources of DNA base damage, the most frequent insult on DNA of all. DNA base damage includes spontaneous hydrolytic decay of the backbone but also base loss, spontaneous deamination, base alkylation or oxidation through metabolites including reactive oxygen species (ROS). The resulting small chemical damage to DNA bases can be cytotoxic or mutagenic and needs to be repaired immediately if proper genome function is to be guaranteed (Kim and M. Wilson III, 2012; Lindahl, 1993). At least 20'000 damaged bases can be measured at any given time in a human cell and this includes only the known and measurable forms of damage. However, measurements of apurinic/apyrimidinic sites (AP-sites), representing intermediates of BER revealed that, depending on cell type, at least 50'000 and up to 200'000 such repair events can be recorded in human cells at any time point, a difference that cannot be explained by spontaneous depurination (Lindahl and Nyberg, 1972; Nakamura and Swenberg, 1999). These numbers therefore indicated early on that there is high turnover of certain types of lesions and at the same time underline the necessity of a functional and highly efficient and versatile BER system to avoid mutations and cell death.

2.2.1 General Mechanisms of Base Excision Repair

BER is has been described as a pathway with two sub-pathways. One of them is associated with the synthesis of a patch of multiple nucleotides after the excision of one damaged base and is therefore termed long-patch BER (Figure 4, right). This work, however, focusses on the short-patch BER pathway that can be roughly divided into five mechanistic steps. Short-patch BER starts with the damage recognition and the name-giving excision of a single corrupted base by DNA glycosylases. The detection of a damaged base is still an incompletely defined process and differs between DNA glycosylases. One currently accepted model postulates an unspecific interaction of the DNA glycosylase with the DNA, followed by sliding along the double helix and non-intrusive interrogation of nucleotides to search for a damaged site. This process can be supported by spontaneous unwrapping of the DNA from nucleosomes through damage-induced steric hindrance, granting the DNA glycosylase increased accessibility to the damaged site (Friedman and Stivers, 2010; Odell et al., 2013; Tian et al., 2021). While many DNA glycosylases display a spectrum of up to 10 and more possible substrates *in vitro*, the processing in the biological context is focussed on the substrates with the highest affinity. Once the DNA glycosylase has engaged with its substrate, it flips the base out of the DNA-double helix and cuts the N-glycosidic bond by a nucleophilic attack, using a water molecule or a deprotonated Lysine residue, depending on the glycosylase (Dodd et al., 2018; Jacobs and Schär, 2012), generating a base-less AP-site (Figure 4, step A, Figure 5A). BER can notably employ mono- or bifunctional glycosylases of which the latter are able to perform a DNA backbone incision in addition to the recognition and excision of the substrate base.

Following the action of a monofunctional glycosylase like the Uracil DNA Glycosylase (UDG/UNG) or Methyl-CpG Binding Domain 4 (MBD4), Abasic-site Endonuclease/Exonuclease 1 (APE1/APEX1) associates with the AP-site and cleaves the phospho-diester bond 5' of the excised base, leaving behind a single-stranded break (SSB) with a 3'-OH and a 5'-deoxyribose phosphate residue (5'-dRP) (Figure 4, step B). In the case of a bifunctional glycosylase (e.g. Oxoguanine Glycosylase (OGG) or Nei Endonuclease-like (NEIL)), base excision is directly accompanied by DNA strand incision performed by the DNA glycosylase. Depending on the bifunctional glycosylase engaged in the back-bone incision, APE1 or the Polynucleotide Kinase/Phosphatase (PNKP) are required to "clean-up" the DNA termini, which is the third step of BER (Figure 4, step B/C). This step prepares the SSB for gap filling and DNA ligation, which ultimately requires a 3'-OH and a 5'-phosphate. To create these ligatable ends, the DNA polymerase β (POL β) harbours a DNA lyase activity to eliminate the 5'-dRP or PNKP can convert a 5'-OH to a 5'-phosphate, depending on the structure of the BER intermediates. After clean-up of the ends, the DNA intermediate is ready for gap-filling by POL β (Figure 4, step D) and subsequent restoration of DNA strand continuity by DNA ligase 3 (LIG3, Figure 4, step E).

Despite not having an own catalytic activity, the X-Ray Repair Cross Complementing protein 1 XRCC1, is an equally important protein to BER. Interacting with multiple DNA glycosylases as well as POL β and LIG3,

XRCC1 is considered as a scaffold protein, essential for the proper coordination of BER and general SSB repair factors (London, 2015). In addition to facilitating the concentration of multiple BER factors at the site of repair, XRCC1 was also shown to increase repair efficiency of the Thymine DNA Glycosylase (TDG), by

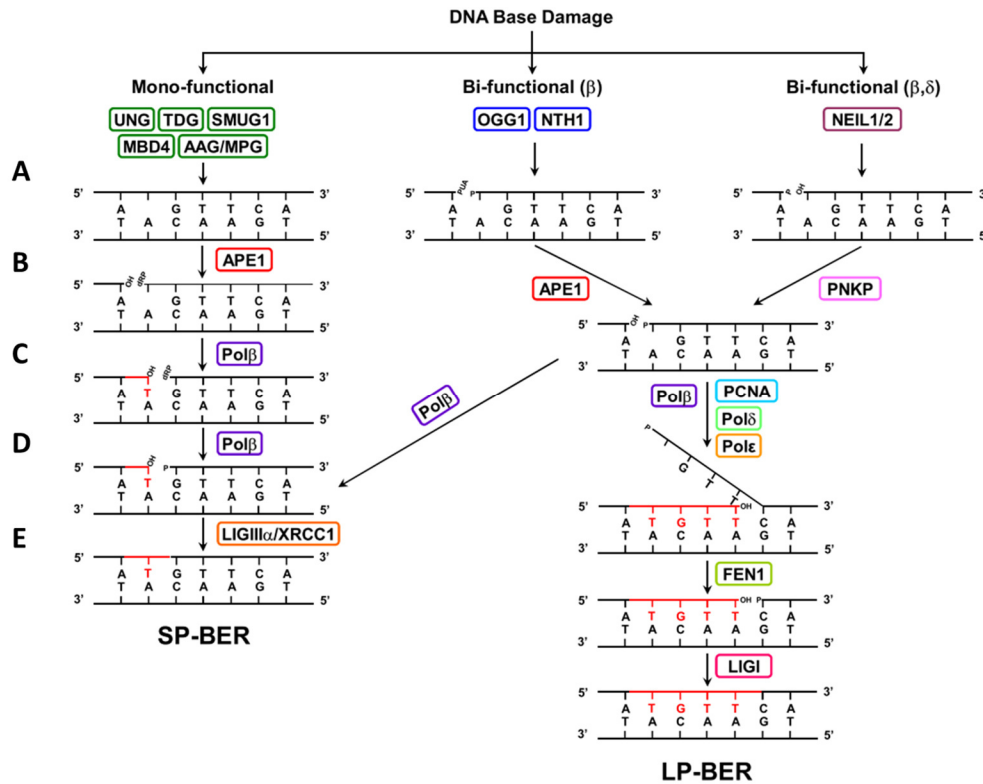


Figure 4 Overview of base excision repair pathways

Short-patch BER (SP-BER) is acting in non-dividing cells, while the long-patch BER (LP-BER) can associate with replication forks through its interaction with proliferating-cell-nuclear-antigen PCNA. Adapted from Kim and M. Wilson III, 2012

transferring a small ubiquitin-like modifier (SUMO) modification to TDG after base excision. This SUMOylation facilitates the dissociation of TDG from the AP-site that it is guarding and allows strand incision by APE1 (Hardeland et al., 2002; Steinacher et al., 2019).

Another XRCC1- interacting factor, the Poly-[ADP-ribose] Polymerase PARP1, has been described to take part in the process of BER (and SSB repair in general), based on the fact that it rapidly recognises and binds SSBs. Binding to SSBs activates the ADP-ribosylation function of PARP1, which then polymerises chains of ADP-ribose moieties on itself (auto-PARYlation) and nearby proteins, including histones. These PAR-chains at the site of DNA base damage or repair are considered as a signalling beacon and recruitment platform for the BER complex (BERosome), comprising XRCC1, as well as the downstream repair factors POL β and LIG3. Indeed, PARP1-proficient cells were shown to increase the speed of SSB repair when compared to PARP1-deficient cells (Fisher et al., 2007). Synthesis of PAR-chains following base excision and/or SSB formation, however, seems to be obsolete, considering that the BERosome may already be assembled at the damaged site. Current investigations in the Schär laboratory try to resolve this mechanistic conundrum, and recent data suggests that PARP1 is not just PARYlating itself and histones, but also members of the BER

machinery, including TDG, APE1, XRCC1, POL β and LIG3 in order to facilitate the dissolution of the BERosome after completion of the repair (see 4.2/appendix II).

In summary, BER is a multi-step DNA repair mechanism, engaging a multitude of proteins that is highly coordinated and dynamic, essential for the survival of a cell.

2.2.2 The Thymine DNA Glycosylase TDG

TDG is a “monofunctional” glycosylase by structure and enzymatic activity but a protein with various biological functions beyond the repair of DNA base damage. Discovered in the late 1980s, it was first described as enzyme that excises thymines (T) when mispaired in DNA with a guanine (G), as it can occur by DNA polymerase errors during replication or as consequence of hydrolytic deamination of a methylated cytosine (5mC) (Wiebauer and Jiricny, 1989, 1990). Such a mismatch would lead to a C to T transition in a subsequent round of DNA replication, and therefore has to be repaired to maintain genomic information. Aside from G•T mismatches, TDG displays high activity towards G•U, G•hydroxy-methyl Uracil (5hmU) and base adducts from exogenous agents like fluorouracil (5FU) (Cortázar et al., 2007; Jacobs and Schär, 2012). The activity on these substrates is well conserved throughout the kingdoms of life and the amino acids crucial for catalysis are located in its well defined catalytic pocket residing in a central core domain of the protein (Figure 5A/B). An apparent second biological function of TDG was discovered when it was found to physically and functionally interact with several TFs like the tumour protein p53, Retinoic Acid Receptor (RAR), c-Jun, or the histone acetyltransferases CBP (CREB-binding Protein) and p300 (Chevray and Nathans, 1992; Kim and Um, 2008; Léger et al., 2014; Tini et al., 2002; Um et al., 1998). While the molecular action of TDG in this context has not been described further than the interaction with said TFs and the consequence of cooperative gene activation, it is likely that TDG is recruited to the targeted genes to exert its third biological function, namely active DNA demethylation.

TDG and Active DNA Demethylation

The most intriguing substrates for TDG were discovered much later in the form of two oxidized derivatives of 5mC, namely 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) (Maiti and Drohat, 2011). These two derivatives are produced iteratively by the α -ketoglutarate-dependent oxidation activity of the ten-eleven-translocation (TET) proteins on 5-hydroxymethylcytosine (5hmC). The possibility of TDG to recognise and excise these two modified bases, allows TDG to exert its third major biological function as a modulator of the epigenetic landscape (and therefore transcriptional control, see more in chapter 2.3). The reinstatement of unmodified cytosines after methylation by DNA Methyl Transferases (DNMTs, Leonhardt et al., 1992; Okano et al., 1999) was initially thought to be impossible by an enzymatic process. Consistently, a direct DNA demethylase activity could never be identified in animal cells, despite strong efforts (Chen and Riggs, 2011; Jin et al., 2008).

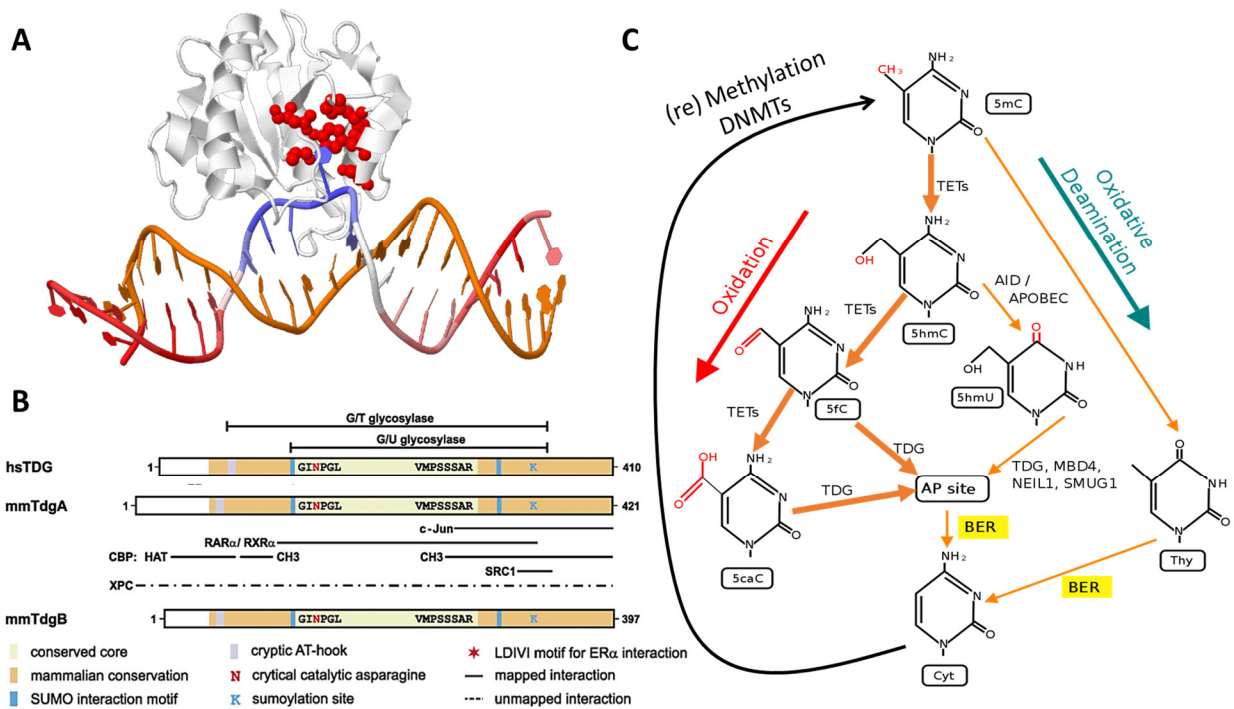


Figure 5 TDG structure and active DNA demethylation

A 3D-Structure of human TDG amino acid residues 123-303, associated with a 5caC flipped out of the DNA double helix. Catalytically important amino acids (N140, A145, H151, Y152, N157, N191) are coloured in red. (PDB 6U16) **B** Schematic primary protein structures of human (hsTDG) and the two murine splice variants of TDG (mmTdgA/B). Conserved catalytic amino acids are written including the crucial N140 (red). Adapted from Cortázar et al. 2007. **C** Cycle of active DNA methylation and demethylation. AID, activation-induced deaminase; APOBEC, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like; MBD4, methyl-CpG binding domain 4. Adapted from Bayraktar et al. 2018

The combined action of TET proteins and TDG, however, provided a plausible mechanism for active DNA demethylation and may explain why TDG is the only mammalian DNA glycosylase essential for embryonic development (lethal after embryonic day E10.5 when disrupted) in mice (Cortázar et al., 2011; Cortellino et al., 2011). Although only having minor effects on 5mC levels in these studies, the consequences of TDG deficiency surfaced as changes of histone methylation (H3K4me2 and H3K27me3) and transcriptional differences when mouse embryonic stem cells (mESCs) underwent differentiation. But soon after these reports, 5fC levels in promoters were then associated with transcription levels of corresponding genes and depletion of TDG in unchallenged ESCs led to a 4 to 5-fold accumulation of 5fC and 5caC, corroborating that TET-TDG mediated active DNA demethylation is taking place without any demand from cellular reorganisation, i.e. differentiation or external stressors (Neri et al., 2015; Raiber et al., 2012; Shen et al., 2013). Along this line, oxidised derivatives of 5mC (oxmCs) were also observed in post-mitotic cells like neurons and could therefore also be uncoupled from cell cycle-mediated passive dilution (also see chapter 2.3.1, Bayraktar and Kreutz, 2018).

TDG was shown to directly interact with the *de-novo* DNA methyltransferases DNMT3A/B (Boland and Christman, 2008; Li et al., 2007) and with TET proteins (Müller et al., 2014; Weber et al., 2016), suggesting that these DNA methylation-modifying proteins cooperate in close proximity. This allows for a highly flexible epigenetic regulation by a constant turnover of DNA methylation, by controlled cyclic methylation,

demethylation and re-methylation (Figure 5C). A mechanistically related concept of DNA demethylation was proposed before it became known that TETs can oxidise 5mC, in which deamination (as opposed to oxidation) of 5mC was shown to create G•T mismatches processed by TDG- (or MBD4)-mediated BER (Métivier et al., 2008). This pathway is plausible but comes with the caveat that deamination is highly mutagenic, and it could not be substantiated as a major route for active DNA demethylation. Nevertheless, deamination of 5mC does occur in many cell types, spontaneously and enzymatically, including mESCs and epiblast-like cells. The proportion of this pathway compared to the oxidation of 5mC, however, seems to be low and rather takes place on free nucleobases than on 5mCs incorporated into the DNA (Spada et al., 2020). 5mC turnover via oxidation has been addressed more intensely and is well established and there is ample evidence for an essential role of TET-TDG-BER-mediated 5mC turnover in cell plasticity, particularly in pluripotent stem cells or in cells undergoing reprogramming to pluripotency (Figure 6A, Ginno et al., 2020; Parry et al., 2020; Rulands et al., 2018). Notwithstanding its obvious biological importance, the regulation and kinetics of this molecular mechanism of cyclic DNA methylation and demethylation as well as its precise function in genome programming remains to be clarified. One possible outcome of TET/TDG-mediated active DNA demethylation was recently uncovered in the Schär group, who investigated 5mC turnover and connected it to chromatin accessibility. In this study, the loss of TDG and therefore most likely loss of 5mC turnover and concomitant BER, can result in reduced histone H1 eviction. These altered histone dynamics in turn reduce chromatin accessibility and finally lead to major re-distribution of histone methylation marks during early mESC differentiation. This implies that TDG and active DNA demethylation are also contributing to higher-order changes in the chromatin landscape (Figure 6B, Zeinab Barekati personal communication, manuscript in preparation, also see chapter 2.3)

The involvement of TDG mediated BER in transcriptional regulation via active DNA demethylation widens the scope of BER from a “simple” but efficient repair system that responds to randomly arising DNA damage, to an actively targeted repair process that helps shaping the epigenetic landscape.

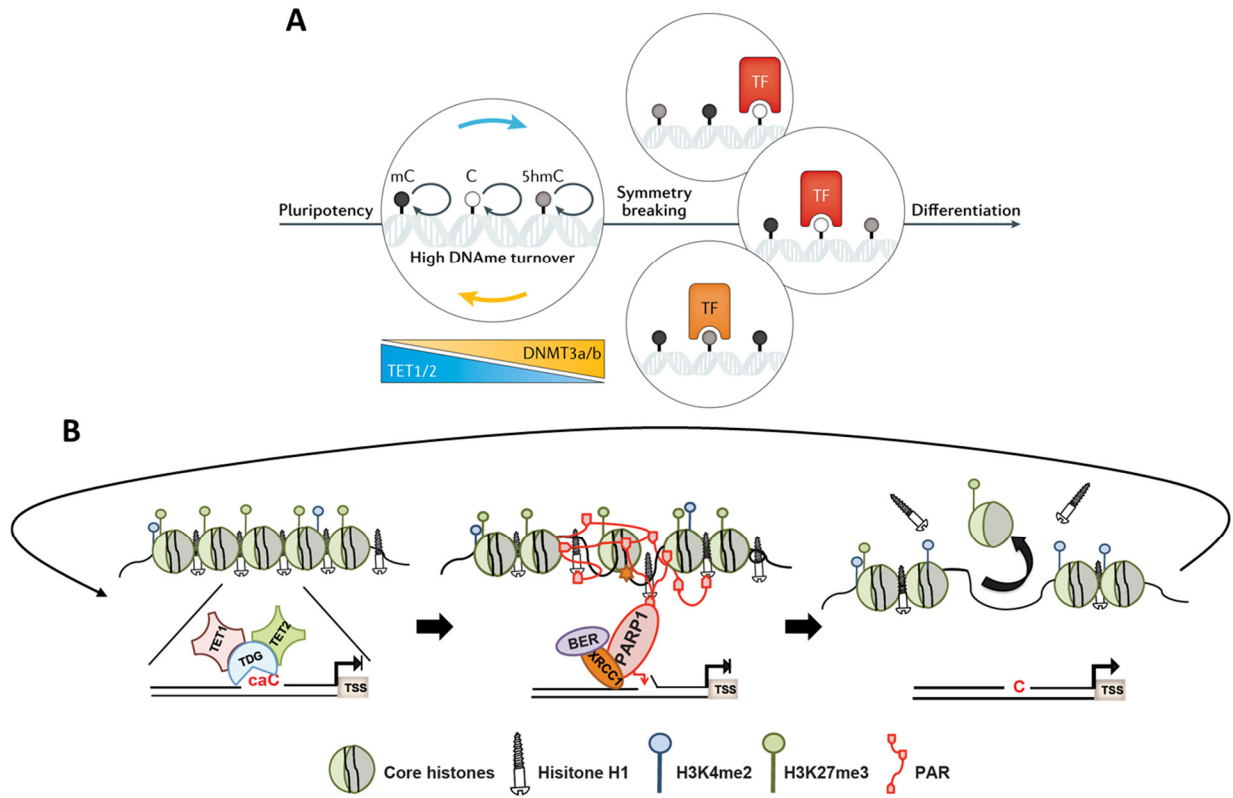


Figure 6 Turnover of DNA methylation and implication on chromatin rearrangements

A Schematic of how differential active DNA demethylation turnover can affect TF binding and tip the balance of expression patterns between cells during differentiation. From Parry et al. 2020 **B** Working model of how active DNA demethylation can lead to histone eviction. Adapted from Barekati et al. (Manuscript in preparation).

2.3 Epigenetic Modifications and their Regulation

The field of epigenetic research started in the 1940ies by Conrad Waddington, who postulated, reminiscent to Jean-Baptiste de Lamarck one and a half a centuries earlier, that development of a species can directly be influenced by environmental cues (Waddington, 1942). He proved it several years later by “canalising” a wild-type population of the fly *Drosophila melanogaster* to display a certain phenotype by the exposure of the *pupae* to a mild “heat shock” (25°C) or ether, which are both non-mutagenic. The trait was maintained without further selection already eight generations after - an unlikely short time frame to generate and stabilize random mutations (Waddington, 1956). He explained this observation with a “readiness” of the population to adapt, based on a pre-existing genetic heterogeneity inherent to a wild-type population, in contrast to a clean-bred laboratory strain, with which this experiment is doomed to fail (Noble, 2015; Waddington, 1957). Waddington applied this principle also to the development of an organism, where the cells of an embryo canalise into their specific function in tissues that are distinct from their surroundings despite the common ancestry, and finally broke it down to the differential regulation and (inter-)action of genes during development. At that time, Waddington was not in a position to grasp the vast interaction network of existing epigenetic actors, but primed modern epigenetic research, in which the highly dynamic modulation of gene expression, without relying on (directly) altering the DNA sequence, is investigated. Thereby the field started to approach an insight where the evolutionary theories of Lamarck and Darwin will finally be combined rather than used to underline contrasts.

The scientific effort conducted until today revealed that the insinuated actors comprise chemical modifications on DNA and DNA-interacting proteins, like histones, as well as the proteins responsible for their positioning. In contrast to the DNA sequence itself, these modifications are reversible and highly interchangeable during the lifetime of one cell and thereby provide the flexibility to alter genetic information without the need of mutations. The following chapters describe how epigenetic regulation accompanies the long-known processes of transcriptional control through transcription-factor binding, from an over-arching genome-wide level, down to fine-tuning of the transcription rate of a single genes.

2.3.1 DNA Methylation

One of the two most studied epigenetic modifications is the methylation of DNA bases, predominantly modifying the carbon-5 position of the cytosine nucleobase (5mC) followed by a G (CpG). Discovered in 1925 (Johnson and Coghill, 1925) as “just” an additional base without apparent function, the understanding of the importance of cytosine methylation started to emerge with its association to the transcriptional state of a gene roughly 60 years later (Bird, 1984a). DNA methylation is an enzymatic process carried out by DNMTs 1-3, that use S-adenosyl methionine (SAM) as methyl donor. The properties of DNMT1 suggest a function in the maintenance of existing methylation patterns during DNA replication, by adding a methyl

group to the complementary CpG on the newly synthesised strand. When the maintenance activity of DNMT1 is inhibited, either by a lower expression of the protein or repulsion by other DNA-binding proteins, CpG methylation patterns levels will not be copied after replication, leading to passive DNA demethylation. Despite being named a “DNA” methyltransferase, DNMT2 is known to have a much higher methylation capability on tRNA than on DNA, if any. DNMT3s are *de novo* methyltransferases as they do not need any hemi-methylated template to copy from. This ability enables them to induce change in DNA methylation patterns and thereby regulate transcription, as discussed later. While this simplified view of role allocation holds true in general, different ways of regulation have also been found, blurring the borders of targeted *de novo* methylation and untargeted maintenance of methylation or its inhibition (reviewed in Lyko, 2018; Mulholland et al., 2020a).

Only about 5% of all Cs are methylated in mammalian genomes, however, affecting the majority of CpG (70-80%) sites in most cell types, except embryonic stem cells (Jabbari and Bernardi, 2004). This non-stochastic distribution hints toward a molecular process with a beneficial function that was selected upon and is underlined by the existence of CpG-islands (CGIs). CGIs are stretches of DNA longer than 500 bp with an increased observed versus expected ratio of CpGs and are frequently found in gene regulatory regions including promoters (Takai and Jones, 2002). In contrast to the genomic average, promoter CGIs are generally hypomethylated, presumably to allow binding of TFs, which would otherwise be hindered by methylated bases. Increased CpG methylation of promoter-CGIs is correlated to the transcriptional repression of associated genes, supporting a regulatory role in transcription via alteration of TF binding. However, not all genes harbour a CGI in their promoter and even in terminally differentiated cells, less than 25% of CGIs have been found to be methylated, suggesting that relationship between CGI methylation and transcription to be more complex than a simple of “on and off” concept (reviewed in Illingworth and Bird, 2009). The simplified view, that methylation generally leads to repression, is nonetheless widely accepted, despite the fact that intragenic DNA methylation has been associated with active transcription (reviewed in Kulis et al., 2013) and activating transcription factors have been found to bind methylated promoters (reviewed in Zhu et al., 2016).

Two factors probably contributed to the establishment of this general conception and both are based on limitations in the experimental approach. One is the long-standing assumption that methylation marks on DNA are not reversible, together with the observation that a genome becomes more and more methylated while undergoing differentiation. A genome-wide increase in DNA methylation can indeed account for the silencing of genes that are not needed in the terminally differentiated cells of tissues and will thereby contribute to stable lineage commitment. Given this global genomic observation, the resolution of initially available/executed techniques of DNA methylation assessment did not allow detection of more nuanced spatio-temporal variation. The second factor lies in the comparably late discovery of oxidised 5mC derivatives

(oxmCs, also see chapter 2.2.2) which focussed the field on the analysis of only 5mC with methods like bisulfite-conversion or antibody-based immunoprecipitation of 5mC. These techniques are, however, not able to detect the presence of oxmCs, meaning that these modified bases were detected as non-methylated Cs despite being a consequence of cytosine methylation - missing thereby an essential part of the of DNA methylation spectrum. Later, these TET-mediated oxidised derivatives of 5mC, namely 5hmC, 5fC and 5caC, were only considered as mere transient by-products of DNA demethylation but it then became clear that these variants also carry their own regulative information. On one hand, they were shown to be bound by distinct readers including transcription factors (Iurlaro et al., 2013; Spruijt et al., 2013). On the other hand, the amount of these oxidised marks were shown to be independent of the levels of other DNA methylation marks, suggesting an individual longevity/stability and, hence, regulatory impact (Bachman et al., 2015; Mulholland et al., 2020b).

DNA methylation has to be viewed as an epigenetic mark that is flexibly positioned and/or further modified throughout the genome in response to internal or external stimuli for the regulation of transcriptional activity. It does not directly impact the coding potential and secondary DNA structure, but harbours information affecting the binding (rate) of proteins to DNA, which indirectly influences higher order chromatin structure. As an example, introducing a 5mC to a binding site can reduce the affinity of a TF that preferentially binds the unmethylated sequence and/or allows the recruitment of a second TF that competes with the first TF and/or in turn changes the rate of transcription initiation itself. This complex balance between binding and dissociation allows the fine-tuning of transcriptional outputs and defects in the control of DNA methylation have frequently been associated with disease (Jin and Liu, 2018; Zhu et al., 2016; Zuo et al., 2017).

DNA Methylation in Gene Bodies

DNA methylation is often considered as the best-studied epigenetic mark, but this is only true for methylation at gene regulatory regions like promoters and enhancers. CpG methylation in gene bodies, however, has been studied to a much lesser degree and most of the evidence available is correlative and often not conclusive. A commonly accepted principle is that higher rates of transcription are accompanied by higher levels of gene body methylation (GBM). Why the modification can be found in GBs and what biological consequence results from it are still an open debate (Anastasiadi et al., 2018; Zilberman, 2017). Studies on GBM have revealed that in the wake of RNAP2-mediated transcription in mESCs, the histone methyltransferase SETD2 accompanies RNA synthesis by RNA polymerase II and deposits H3K36me3 marks, which are also highly correlated with transcriptional activity. H3K36me3 can in turn be bound by the PWWP-domain of DNMT3s that then deposit 5mC in the GB (Baubec et al., 2015). Interestingly, this association was negated a few years earlier, by showing that 5mC deposition is not dependent on SETD2 when examined in human epithelial and colon cancer cells (Hahn et al., 2011). These seemingly contradicting

observations could indicate that a tissue-specific mechanism is responsible for GBM and/or other mechanisms take place that are not yet discovered.

Addressing the putative function of GBM, two main models are currently discussed. The first postulates GBM to regulate splicing as 5mC marks are also found at intron-exon boundaries i.e. splice sites. Indeed, it was shown that a change of methylation correlates with the production of alternatively spliced mRNAs, but the reports so far do not establish a conclusive general concept. Either research was done only at a single or few loci (Marina et al., 2015) or when extended genome-wide, only about 20%-30% of observed alternative splicing events (even less so in plants (Wang et al., 2016b)) could be attributed to different levels of 5mC (Lev Maor et al., 2015; Yearim et al., 2015). Also, the range of effects was often small and/or barely reached statistical significance (Maunakea et al., 2013). Nonetheless, the evidence available suggests that increased methylation of a single CpG at a splice site can change the inclusion frequency of an intron of an artificially introduced single gene (Shayevitch et al., 2018).

The second, widely discussed hypothesis is that intragenic 5mC restricts uncoordinated “cryptic” transcription in the gene body. It was shown that reduced GBM results in an increase of spurious transcripts (Neri et al., 2017). The effect of missing GBM on these transcripts was, however, considerably smaller than the absence of the DNMT3 and SETD2 or its function mediating H3K36 trimethylation. Also do higher levels of GBM generally correlate with higher transcription rates, which contradicts this notion. Even more so, transcription-associated GBM is a ubiquitous phenomenon that is not restricted to genes displaying spurious transcripts and is also not specifically enriched at intragenic promoters (Jeziorska et al., 2017; Jjingo et al., 2012). This leads to the conclusion that this model, again, can only explain a partial function of observed GBM.

Another, less commonly addressed view on intragenic DNA methylation envisages a role of GBM to bookmark genes that have been activated before but are currently silenced by DNA or histone methylation at the promoter (Tompkins et al., 2016). This would keep the promoter as the main switch to turn a gene off and on, but uses GBM as a “memory” to facilitate fast transcriptional elongation upon the reactivation of the gene, which would otherwise quickly be heterochromatinised (also see chapter 2.3.2). This hypothesis relies on a mechanism that repulses silencing factors based on 5mC, which remains to be established. One study indeed showed that transcriptional memory is dependent upon elongating – and not initiating - RNAP2, therefore drawing attention towards events in the gene body, but unfortunately did not address GBM (Iberg-Badeaux et al., 2017)(also see chapter 4.4).

As mentioned before, DNA methylation has to be considered in all its variety, including the oxidised variants; but despite the fact that oxmCs have been described to be generated in the gene body long ago, (Shen et al., 2013), the function of their intragenic presence remains even more elusive than that of 5mC.

2.3.2 Modification on Chromatin Proteins

The DNA double-helix complexes with heteromeric histone protein clusters to form chromatin. Each of these clusters consists of eight subunits comprising two copies of the histone proteins H2A, H2B, H3 and H4 each. A stretch of ~147 bp of DNA winds twice around a histone-octamer to form a nucleosome that is formed directly after DNA replication. A fifth histone protein, H1, can be attached at the nucleosome dyad, where the DNA “enters” and “leaves” the nucleosome, and is thought to tether the DNA, reduce histone sliding and increase chromatin compaction (Figure 7A). The histones H3 and H4 have been extensively investigated, also because these reversible modifications take part in transcriptional control and epigenetic regulation. The vast majority of histones are modified at various amino acid residues on their N-terminal tails, mostly at lysines (K), whereby methylation and acetylation are the most frequent modifications in a large catalogue of additional modifications including ubiquitylation, SUMOylation, PARylation and more. While histone acetylation only occurs as a single molecular modification, up to three methyl groups can be added to a lysine, each representing a signal with specific information (Figure 7B). For example, tri-methylation of K27 on histone 3 (H3K27me3) is associated with transcriptional inhibition and increased compaction of chromatin, while methylation at K4 is generally associated with a permissive transcriptional state. Also, H3K36 methylation is associated with transcriptional activity but is predominantly found in the gene body when tri-methylated. As multiple lysines of one histone can be modified at the same time, an enormous combinatorial variety of modifications could be found at a single or in a stretch of nucleosomes, generating the so-called “histone code”, where each combination carries individual regulatory information (reviewed in Carlberg and Molnár, 2018; Hyun et al., 2017; Simonet et al., 2016). However, the detection of multiple modifications on one histone is experimentally very challenging and only a few of them have been described with confidence (Shema et al., 2016; Villaseñor et al., 2020). The most prominent example is the bivalency of H3, where K27 is tri-methylated and K4 is di- or tri-methylated. These “contradicting” signals for transcriptional activity result in a promoter (or enhancer) state that is neither completely permissive nor inhibitory to transcription. This so-called “poised” state is thought to provide accessibility to TFs while transcription cannot yet be initiated, but is then able to initiate transcription very rapidly when needed (Blanco et al., 2020).

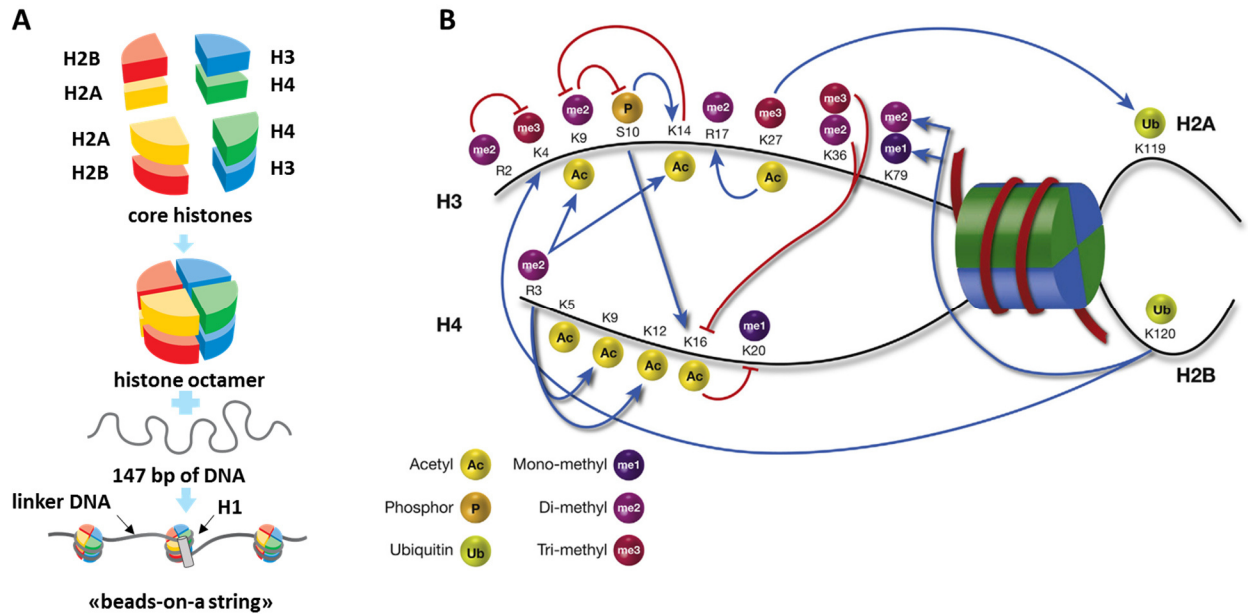


Figure 7 Histone structure and post-translational modifications

A Schematic structure of a histone octamer and winding of DNA. Adapted from Morgan, 2007 **B** Non-exhaustive depiction of modifications at histone tails and their mutual influence. Blue arrows indicate facilitation of modifications, red stunted arrows indicate inhibition. K: lysine, R: arginine, S: serine; Adapted from Simonet et al., 2016

Establishment and Alteration of the Histone Code

Histone modifications are dynamically deposited by “writers” like histone acetyl transferases (HATs) such as p300/CBP (e.g. H3K27ac, H2A.Zac), or histone methyl transferases (HMTs) like Enhancer of Zest Homologue 2 (EZH2 for H3K27me3). They can be enzymatically erased by corresponding histone deacetylases (HDACs) or histone demethylases like Jumonji Domain-containing Protein D3 (JMJD3). Not all histone-binding proteins exert an enzymatic function but recognize a specific stretch of chromatin, and therefore called or “readers”. They, however, frequently interact with erasers or writers to alter the histone modifications to finally change or reinforce the chromatin conformation in a sequence-specific context (Zhang et al., 2015). Well-known examples for such complexes are the Polycomb Group Repressive Complexes PRC1 and PRC2. PRC2, for instance, includes a DNA-binding domain with affinity for CGIs, thereby being a DNA methylation reader in addition. The PRC complexes are essential for the control of H3K27me3 to promote transcriptional silencing and the formation of facultative heterochromatin in a scenario like the following. A gene that was previously transcribed is to be shut down, and the corresponding TF is no longer available to bind the promoter. The un-occupied promoter harbours an unmethylated CGI, as it was needed for accessibility of TFs and formation of the PIC. This unmethylated CGI allows binding of the PRC2 subunit Adipocyte Enhancer-Binding Protein 2 (AEBP2) that recruits the PRC complex 2.2 to the DNA (Figure 8A, step 1). The methyltransferase EZH2 deposits tri-methylation marks at H3K27 in the vicinity, which repulse further TF binding (Figure 8A, step 2). The H3K27me3 mark in turn leads to the recruitment of more PRC2 complex via the subunit Embryonic Ectoderm Development (EED), generating more H3K27me3 marks (Figure 8A, step 3). The H3K27me3 mark can also be bound by

Chromobox (CBX) subunit of PRC1 (Figure 8A, step 4), which ubiquitinylates lysine 119 of histone H2A (H2AK119u), that in turn can again recruit PRC2 via JARID2 or PRC1 via RYBP (Figure 8A, step 5). These interactions therefore pose a very strong positive feedback loop for repression that can quickly spread over a promoter region and eventually into the gene body to decrease chromatin accessibility and hinder transcription (Aranda et al., 2015; Hosogane et al., 2016).

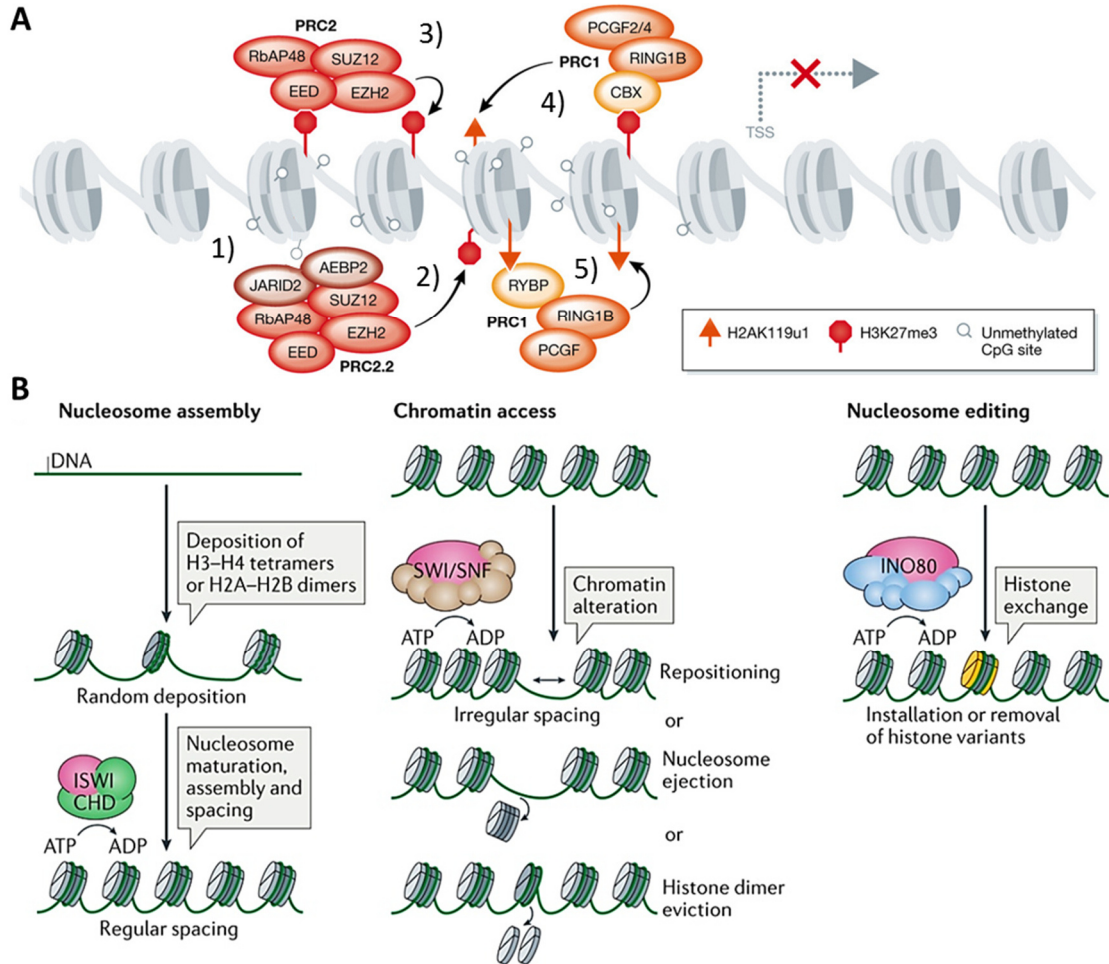


Figure 8 Interaction of histone modifiers and chromatin compaction

A Schematic on the interaction of Polycomb Repressive Complexes 1 and 2 at a promoter of a gene to halt transcription. 1) AEBP-mediated binding of unmethylated CpG, 2) H3K27me3 deposition by EZH2, 3) EED-mediated recognition of H3K27me3 and deposition of more H3K27me3, 4) H2AK119u1 deposition after H3K27me3 recognition by CBX, 5) Ring1B and YY1-binding protein (RYBP)-mediated recognition and deposition of H2AK119u1. Adapted from Zhang et al. 2015 **B** Summary of chromatin remodelling mechanisms. Left: Initial histone deposition and coordinated spacing by Imitation Switch (ISWI) and Chromodomain Helicase DNA-binding (CHD) class proteins. Middle: Histone rearrangements via sliding, whole eviction or partial eviction by members of Switch/Sucrose Non-fermentable (SWI/SNF) group. Right: Complete (or partial) histone exchange by members of the INO80 group. From Clapier et al. 2017

Like DNA methylation, histone methylation does not directly alter the chromatin structure, but charged modifications like acetylation or phosphorylation can impact the association of the positively charged histones with DNA. Acetylation at lysines negates the positive charge and the phosphorylation at serines adds a negative charge to the amino acid at the histone tail, building up a repulsion to other acetylated/phosphorylated histones and nucleosomes as well as the negatively charged DNA leading to a

consequential loosening of chromatin (Fenley et al., 2010; Hong et al., 1993). Conversely, chromatin can be compacted by histone modifications such as H3K9me3 that attracts histone-binding proteins, like Histone Protein 1 (HP1). HP1 is able to bridge two nucleosomes with these modifications and pulls the nucleosomes together. Chromatin remodellers are then engaged to increase the DNA winding and thereby shortening the linker DNA to maximize compaction and form the inaccessible, constitutive heterochromatin (Machida et al., 2018; Watanabe et al., 2018).

Not just the modifications on histones are dynamic, but the histones themselves are as well. As mentioned earlier, histones are not fixed to the stretch of DNA as there exist no covalent bonds between the histone and DNA molecules, but are held together by electrostatic and van-der-Waals forces (Ettig et al., 2011). This allows the nucleosomes to dynamically slide along the DNA, to be rearranged in their composition of subunits, or even completely evicted. While the latter two can happen passively (e.g. during passage of RNAP2 (Kulaeva et al., 2013)), all of the processes can also be actively targeted, which is then summarized under the term chromatin- or histone remodelling. Each of these targeted processes are executed by specialized protein complexes that require the energy source adenosine triphosphate (ATP). They are essential to enable transcription factors to bind DNA and to allow passage to RNA polymerases during transcriptional elongation and DNA polymerases during replication (Figure 8B left/middle, reviewed in Clapier et al., 2017; Venkatesh and Workman, 2015).

Increasing the complexity of chromatin, histones can come in variants. These variants differ from their canonical versions only in a few amino acids and are usually deposited independently from replication. After histone subunits (or the whole octamer) is evicted, so-called histone chaperones take over to reinstate a functional nucleosome (Figure 8B, middle/right, reviewed in Burgess and Zhang, 2013). A well-examined example is the histone variant H3.3 and its chaperone Histone Cell Cycle Regulation Defective Homolog A, short HIRA. H3.3 varies from the canonical H3 in only 5 amino acids and was initially associated with active transcription and chromatin accessibility. (Armache et al., 2020; Schlesinger et al., 2017). But H3.3 is also associated with transcriptional silencing; H3.3 was found to attract the DNA binding protein and SUMO-ligase, Tripartite-motif 28 (TRIM28, or KAP1) to repetitive elements, which in turn recruits the H3K9 methyltransferase SETDB1 to mediate chromatin compaction and heterochromatinisation (also see chapter 2.4.2, Elsässer et al., 2015). Furthermore, H3.3 was shown to be directly involved in transcriptional silencing of genes. More precisely, it is not just the mere presence of H3.3 but also its constant incorporation/turnover that seems necessary to maintain transcriptional states, including also *de novo* silencing of genes (Nashun et al., 2015). The exact molecular mechanisms of this context-dependent regulation remains mostly elusive, and it is therefore difficult to attribute general roles to a histone variant and even more so to the multiple other variants for each histone subunit. Adding the previously mentioned

histone modifications to the picture, which can also be applied to their non-canonical variants, underline the extreme flexibility and variety of epigenetic information on the level of histones.

2.4 Interplay between DNA Methylation, Histone Modification and Transcription

To exemplify how epigenetic control is intertwined and how intricately epigenetic modifications are regulated, two biologically relevant scenarios are described in more detail.

2.4.1 Epigenetic Rearrangements in Stem Cells

Embryonic stem cells (ESC) are an extensively studied experimental model in the field of epigenetics. Their ability to differentiate into cells of any somatic tissue requires a very plastic apparatus for transcriptional regulation, and the major epigenetic rearrangements that take place within the nucleus of a differentiating cell are welcome study objects in epigenetic research.

Blastocyst-derived ESCs are generally DNA hypomethylated as their entire genome needed to be stripped of cell type-specific epigenetic patterns, to create a “clean slate” for the establishment of a new epigenetic landscape in the totipotent zygote (Amouroux et al., 2016). This low level of DNA methylation and restricting histone modifications extend over the bulk genome, exempt from so-called imprinted loci that preserve methylation patterns given by the parental origin of the chromatid. To guarantee pluripotency of the blastocyst, where ESCs are derived from, the hypomethylated state is largely retained but repetitive elements start to become methylated to avoid genomic instability, and, in case of female animals, inactivation of the second X-chromosome starts to take place in mice to ensure dosage compensation in later development; or has already happened in humans at that point.

In vitro culturing of pluripotent ESCs is achieved in two different ways. One is the use of a chemically defined, serum-free medium called “2i” that contains two small molecule inhibitors to slow down proliferation (GSK3-inhibition by CHIR99021) and inhibit expression of differentiation-associated genes via the MAPK/ERK pathway (inhibition by PD0325901). Further adding the cytokine Leukaemia Inhibitory Factor (LIF) supports self-renewal and expression of pluripotency genes (Ying et al., 2008). The other method uses a serum-containing medium also supplemented with LIF. This method relies on the signalling of Bone Morphogenic Protein 4 (BMP4) present in the serum to maintain high levels of pluripotency gene expression. ESCs kept in these two media show similar properties but represent distinct stages of early embryonic development in mice (Figure 9A). The 2i condition favours a very low level of DNA methylation keeping the ESCs in a more naïve state while serum-containing medium leads to an increase of 5mC levels

and “primes” the ESCs for further differentiation and lineage commitment that is happening later in development.

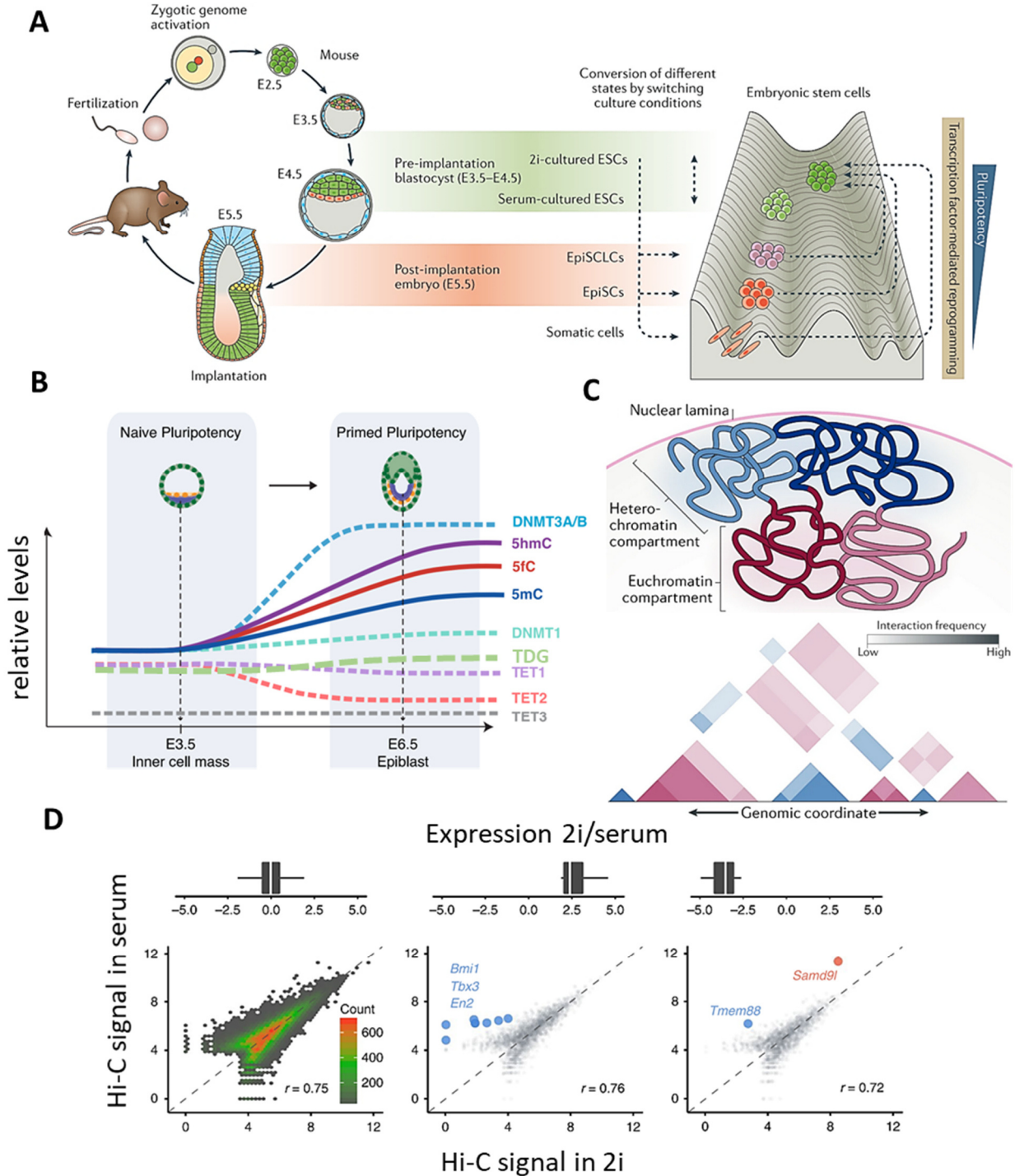


Figure 9 Mouse embryonic development and chromatin interactions in early differentiation

A Overview of mouse embryonic development (left) and a schematic of the differentiation states down the topology of Waddington’s epigenetic landscape. Adapted from Atlasi and Stunnenberg 2017. **B** Overview of relative protein levels as well as levels of 5mC and oxmC derivatives during early differentiation. (Adjusted from Mulholland et al., 2020b). **C** Top: Schematic of chromatin interactions within the nucleus, where heterochromatic regions are often associated with the nuclear lamina. Bottom: Exemplary view of topologically associated domains (TADs) detected by Hi-C, where the triangles indicate the interacting region and its frequency. Darker complexion indicates higher degree of interaction. **D** Hi-C data from Atlasi et al., 2019, stratified by expression change (boxplots on top), showing small changes of chromatin interactions despite high differential expression during 2i to serum transition.

A third and well-investigated state of pluripotent mESCs are epiblast-like cells (EpiLCs), which display a massive increase in DNA methylation and histone modifications and represent the embryo proper at the implantation state (Atlasi and Stunnenberg, 2017; Marks and Stunnenberg, 2014).

The analysis of the transition from naïve to primed pluripotency revealed characteristics of the epigenetic landscape that underline the importance of oxidized 5mC derivatives in development. During the genomic remethylation, the levels of 5hmC, 5fC and, most likely also 5caC, increase over-proportionally to the increase of 5mC. This means that the commitment of ESCs towards certain germ layer cells is not simply relying on the silencing of unnecessary genes by the methylation of CpGs, but highly depends on the control of this process by oxidation of 5mC and active DNA demethylation by TET and TDG (Figure 9B, Kohli and Zhang, 2013; Mulholland et al., 2020b).

Very recent work examined chromatin interactions during this differentiation process, using an experimental technique called chromosome conformation capture (3C or a modification of it, Hi-C). These interactions are considered as key determinants of cell identity as they represent networks of enhancer-promoter contacts for cell type-specific transcriptional programs. In the transition from 2i-cultured, naïve ESCs to pre-implantation cells (serum-cultivated), however, these interactions barely change while the differences in transcription between these two differentiation states are immense (Figure 9D). Only upon entering a later stage of differentiation resembling epiblast cells, minor changes in chromatin interactions become apparent. The report also showed that regulation of genic transcription at this state of differentiation strongly relies on the combined action of DNA methylation and modifications at H3K27. These observations underline the necessity of transcriptional control by epigenetic DNA and chromatin modifications, on top of classical regulation by transcription factors and enhancer-promoter interaction (Atlasi et al., 2019).

2.4.2 Epigenetic Control of Repetitive Elements

DNA and histone methylation are not just known for their regulative role in gene expression but also for the repression of non-coding DNA such as repetitive elements (RE). REs are broadly defined as stretches of DNA that contain patterns of nucleotide sequences that are repeated variable amounts of times. They make up at least 55-60% of the human genome and display a high variation in complexity and length, ranging from long (several kb) multipart arrangements to short and simple repeats (e.g. 10xATAT)(Haubold and Wiehe, 2006). Many of these REs originate from infectious viruses that had integrated into the genome by (retro-)transposition. Accumulated foreign DNA sequences have been considered junk DNA for long (reviewed in Jurka et al., 2007), but as genomic research continues to reveal, such sequences have been adapted for transcriptional control by mutation and selection, and are therefore often found at regulatory regions like enhancers and promoters. Some of these foreign DNA elements have even become fully

functional genes, like the Centromere Protein B CENP-B (Lu et al., 2020; Shapiro and von Sternberg, 2005). This adaptation by utilisation is true for rather ancient sequences where time and evolution resulted in complex rearrangements, but “recent” integrations of such viruses still harbour the potential to be expressed and translocate within the genome. It also seems that DNA transposons are more easily disrupted or rendered inactive by cellular defence mechanisms, as they are barely active in mammalian cells. Retrotransposons (Figure 10A), who transpose via an RNA intermediate, are on the other hand more abundant in the mouse and human genome (Figure 10B) and more often retain the capacity to move (Deniz et al., 2019). Uncontrolled expression of such transposable elements (TEs) has to be countered by transcriptional silencing as it would otherwise pose a threat to genomic stability and genic transcription (Bourque et al., 2018; Makiłowski et al., 2019). The current model on how this silencing is achieved by H3K9 trimethylation and DNA methylation is similar to the situation for genes. The role of these modifications at REs, however, seems to be more clear-cut than at genes; in differentiated cells, the abundance of 5mC and repressive H3K9me3 generally seems to correlate with long-term transcriptional silencing, while demethylation correlates with their re-activation (Goodier, 2016; Minoguchi and Iba, 2008; Walter et al., 2016). But not surprisingly, exceptions to this rule are being discovered as well, particularly in mESCs, where the chromatin is generally hypomethylated and more intricate mechanisms are required to maintain the delicate balance between maximal flexibility and uncontrolled expression (See discussion in Mulholland et al., 2020a).

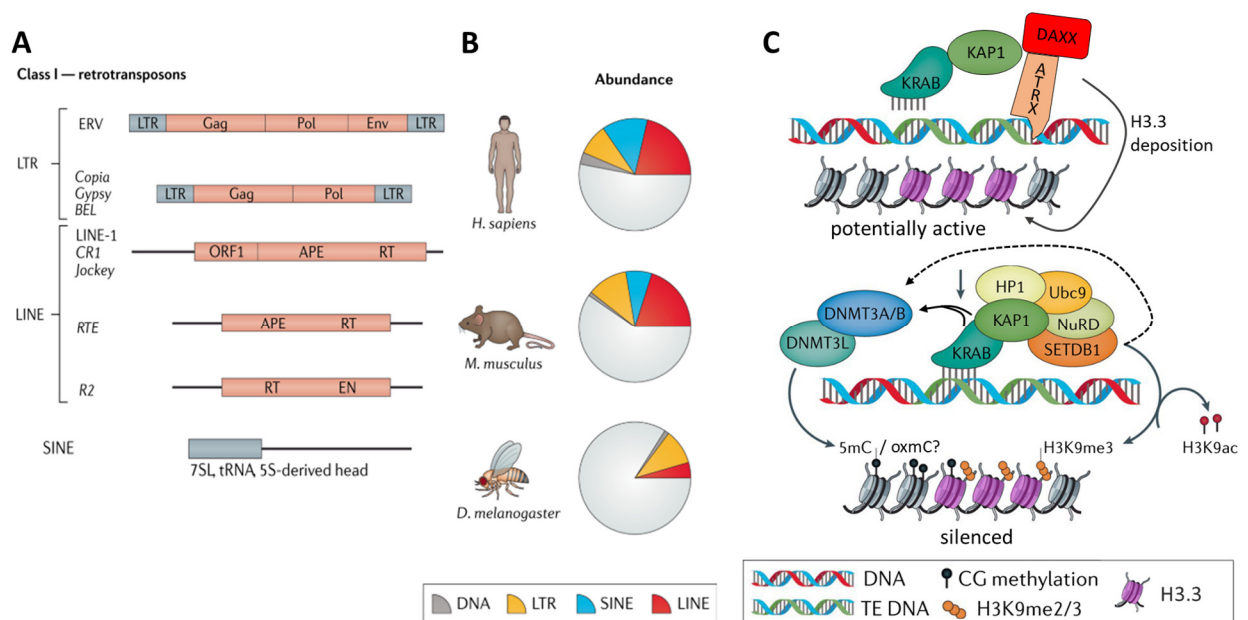


Figure 10 Transposable elements (TE) and epigenetic crosstalk

A An overview of RNA-Transposons and their structure. LTR, Long Terminal Repeat; ERV, Endogenous Retrovirus; Gag, Group-specific antigen; Pol, Polymerase; Env, Envelope; LINE, Long interspersed nuclear element; ORF, Open reading frame; APE, Apurimidinic endonuclease; RT, Reverse transcriptase; RTE, Retrotransposable element; SINE, Short interspersed nuclear element. **B** Abundance of retrotransposable elements in model organisms. **C** Schematic of combined epigenetic silencing of TEs with DNA methylation and H3K9 trimethylation. Adapted from Deniz et al. 2019

Nevertheless, TE silencing appears more complex as recent evidence suggests that oxmCs are continuously generated at such loci in serum-cultivated ESCs, suggesting that DNA methylation is dynamic at TEs even when culture conditions remain stable. Also, at L1 type of long interspersed nuclear elements (LINEs) and endogenous retrovirus-K (ERV-K) elements in human preimplantation embryos, the occurrence of 5fC correlated with a lower expression when compared to element without 5fC, even indicating that active turnover of DNA methylation could contribute to silencing (Gao et al., 2020; Shen et al., 2013).

Another interesting observation with serum-cultivated mESCs was a functional interplay of histone modifiers at a subset of LTRs, called intracisternal A particles (IAPs) (Elsässer et al., 2015). These elements were previously shown to be silenced via H3K9me3 and DNA methylation, indicating a very compact heterochromatic region without transcriptional activity. Despite the high compaction at the loci, the authors were able to demonstrate that the histone chaperone and remodeller Death-domain Associated/Alpha Thalassemia Retardation Syndrome X-linked (DAXX/ATRX) constantly turns over the non-canonical histone variant H3.3, to reinforce the H3K9me3 marks and silence the IAPs and other ERVs. This process, including the interaction between the H3K9 methyl transferase SETDB1 and DNMT3A facilitates the deposition of 5mC that finally prevents transcription (Li et al., 2006; Zuo et al., 2012). Hence, the data suggest that despite low DNA accessibility and silencing of underlying regions, constant rearrangements in the chromatin can be ongoing in order to maintain a silenced state (Elsässer et al., 2015; Navarro et al., 2020; Yang et al., 2015).

The experimental transition of primed mESCs towards naïve mESCs in culture, recapitulating a de-differentiation with rapid loss of DNA methylation, revealed another possibility of silencing of REs. Upon the loss of DNA methylation, H3K9 dimethylation was also lost while H3K9me3 marks remained stable (Walter et al., 2016). This loss of DNA and H3K9 dimethylation led to a de-repression of LINEs and LTRs, which was in turn countered by the novel deposition of H3K27me3 marks to silence the TEs again. This includes H3K27me3 as a flexible silencing mechanism in naïve mESCs and underlines a striking redundancy and variability of epigenetic regulation that is needed to precisely control TEs independent of genic transcription in early development. Interestingly, a well-controlled release of TE-expression in early development is being suggested as a possible driver of advanced mammalian evolution (Mulholland et al., 2020a).

3 Aim of the Thesis

As a major contributor to transcriptional regulation, epigenetic modifications have been investigated heavily for the last thirty and more years. This mode of regulation is involved in every step of the development of an organism as well as in the maintenance of homeostasis in post-replicative cells. While this has become well-accepted in developmental and genome biology, and increasingly also in cancer biology, epigenetic mechanisms still lack general recognition as an integral component of gene regulation. Within the field of epigenetics, a comparable lack of understanding is apparent with respect to the relevance of active DNA demethylation. The biochemical and genetic basis of the process, i.e. the oxidation of 5mC by TET proteins and the subsequent processing of the oxidized base by base excision repair was described ten years ago (Cortázar et al., 2011; Cortellino et al., 2011; He et al., 2011; Ito et al., 2011; Maiti and Drohat, 2011) and has since been extensively investigated. Nonetheless, the molecular mechanisms underlying the contribution of this essential pathway to genome and cell plasticity have not been elucidated.

In my PhD thesis, I therefore aimed at a better understanding of the contribution of BER-mediated active DNA demethylation to general transcriptional regulation, both functionally and mechanistically. More specifically, I aimed to investigate the presence and coordination of TDG-mediated active DNA demethylation on the mechanistical level, especially the generation of DNA single-strand breaks (SSBs). The obtained knowledge would then support assessing the impact of TDG-mediated active DNA methylation and SSBs on transcription as well as chromatin accessibility. Finally, I aimed to analyse the potential role of active DNA demethylation in transcriptional elongation and memory.

To achieve these aims, I used a genetically well-controlled mESC line with a conditionally excisable TDG minigene. Combining this with different transcriptional triggers, I applied a variety of cell-biological, genetic and genomic experimental procedures to investigate the dynamics of active DNA demethylation and transcriptional responses. With others I show that active DNA demethylation is necessary to mediate a genome-wide stress-response and we describe a novel de-repression of repetitive elements upon the PARP inhibitor Talazoparib, of which all is dependent on TET/TDG and the occurrence of SSBs. We could also describe a novel function of PARP1 activity in dissociating BER-factors from their substrates *in vitro* allowing us to differentiate SSB generation at genes regulated by active DNA demethylation from spontaneous SSBs, caused by failed repair, *in vivo*. Generation of a genome-wide dataset for newly transcribed RNA then revealed TDG-dependent alterations in RNAP2 pause-release mediated by changed histone dynamics. Taken together, the obtained data supports the notion of active DNA demethylation as an omnipresent “workhorse” in the cell, particularly the generation of TDG-dependent SSBs in addition to TET-mediated base-oxidation, where most groups stop their investigation of active DNA demethylation.

4 Results & Conclusions

Active DNA demethylation plays a central part in the “toolbox” of mechanisms to regulate the epigenetic landscape. Its involvement in the genome-wide regulation of transcription has been acknowledged but the understanding of its cause, regulation and consequences is still far from conclusive.

Intrigued by the lack of understanding of this complex and importance process, I investigated molecular and morphological phenotypes of cells lacking the DNA glycosylase TDG, the key component in BER-mediated active DNA demethylation. My work in the context of my PhD-thesis, is presented in the two following manuscripts, to which I contributed substantial and critical data. Additional observations and experimental data that are not yet ready for publishing, can be found in the chapter 4.3/appendix III and 4.4. Furthermore, the development of an engineered model system in murine ES cells, that was prerequisite to and extensively applied during my studies, has been published and is provided in the appendix IV. It comprises work mostly obtained during my master studies, and some experimental insights from the PhD and data provided by others.

4.1 PARP Inhibition Induces Cytotoxicity in mESCs by Activating Endogenous Retroviruses

The following manuscript, of which I will be co-first author, is in correction with co-authors and will be submitted to within coming weeks.

4.1.1 Summary

Cancer is the reason for a substantial amount of death in today’s older population and its treatment is a difficult endeavour (Wang et al., 2016a). Today, standard treatments of most cancers still include the rather “imprecise” use of chemicals and/or irradiation that causes random damage to the DNA in a non-cancer cell specific manner. The procedure benefits from the fact, that the fast-dividing fraction of cancer cells is hypersensitive to DNA damage; they fail to activate and execute the necessary DNA repair response and eventually die because proper cell cycle progression is impossible (Dickens and Ahmed, 2018). To reduce the treatment burden and improve treatment success in the long run, therapeutic agents that target specific characteristics of cancer types and cells need to be identified and developed. An approach that seems very promising in treatment of specific subtypes of breast or ovarian cancers, is the administration of a small-molecule inhibitor against the Poly-[ADP-ribose] polymerase 1 (PARP1). Most of the breast cancer types are defective in proteins essential for the repair of double-strand breaks (DSBs) by homologous recombination (HR), a defect which causes synthetic lethality with the effects of PARP-inhibition. The currently accepted cause of action involves the overload of the cellular DNA repair system with increased levels of single-strand breaks (SSBs) caused by PARP-inhibition, which leads to hindrance of SSB repair. This

inability to efficiently repair these SSBs will result in collisions of the lesions with replication forks or elongating RNA polymerases during transcription, ultimately channelling the initial SSB damage into DSBs (Murai et al., 2012). In the case of HR-defective cells, the consequential double-strand breaks (DSBs) cannot be efficiently and correctly repaired and lead to cell death. Unfortunately, however, treatment of HR-deficient cancers with PARP-inhibitors does not guarantee a cure and many patients have to face recurrent tumours (Grinda and Delalogue, 2020; Vinayak and Ford, 2010). To understand this failure in long-term treatment success, more basic research is necessary to understand the consequences of PARP-inhibition at a molecular level.

In this work, we therefore aimed to investigate how cells achieve resistance to PARPi-mediated cytotoxicity by identifying the responsible signalling events on a transcriptional scale as well as how they are executed on a more mechanical level. To this end, we applied the PARP-inhibitor Talazoparib (Tal) to a range of HR-proficient wild-type (WT) cell lines to assess their tolerance to DNA damage. Most of the cell types tested, including bone-marrow derived macrophages or post-mitotic neurons, were not sensitive to the treatment with Tal. An exception to this, however, were pluripotent mouse embryonic stem cells (mESCs) when kept in the chemically defined “2i”-medium to maintain the state of naive pluripotency (see appendix I, Figure 1A). These effects depended on the presence of PARP1 as mESCs depleted of PARP1 were resistant to Tal treatment (Figure 1B/C). One day of Tal treatment resulted in a cell death of about 50% and after a second day of treatment, only ~30% of cells were viable when compared to untreated mESCs (Figure 1E). Cellular responses included an accumulation of cells in S-Phase after 14 h of treatment with 5 nM Tal and a concomitant increase of mESCs with detectable Annexin-V on their surface as a marker for apoptosis (Figure 1F).

To describe the signalling-cascade that finally leads to the observed apoptosis, we performed an mRNA-sequencing experiment in mESCs treated with 50 nM Tal for 6 h and examined the differentially regulated genes. Tal treatment led to a PARP1-dependent upregulation of 1’859 genes but without an obvious signature (Figure 2A). An analysis of the transcription factor binding motif in the promoter of upregulated genes then revealed an enrichment of the tumour-suppressor protein p53 (Figure S2A) and the bulk of upregulated genes was enriched for p53-regulated genes (Figure 2B/C). In line with global transcriptomic data, we confirmed the activation of p53 signalling upon Tal treatment, by demonstration of p53 phosphorylation at Serine 15 and 392 (Figure 2D), as well as an increased recruitment of phosphorylated p53 to the canonical target genes *Fas* and *Tap1* (Figure 2E). The knockdown (KD) of p53 by a siRNA then revealed that the Tal-mediated activation of these targets is indeed dependent on p53. To identify the factors responsible for p53 activation, we analysed the two upstream kinases Ataxia Telangiectasia - Mutated (ATM) and -Rad3 Related (ATR). Treatment with the ATM-inhibitor KU-55933 but not the ATR

inhibitor AZ20, reduced p53 and H2A.X phosphorylation (γ H2A.X, Figure S2D/E) as well as gene expression upon Tal (Figure 2I). ATR activity was further excluded by showing that the ATR targets Checkpoint Kinase 1 (CHK1) and the Replication Protein 2 (RPA) do not get phosphorylated upon Tal (Figure S2F). Based on these results, we concluded that Tal treatment of mESCs leads to an ATM-dependent activation of p53 that is potentially caused by DSBs, despite mESCs, being proficient in DSB repair.

To unravel the p53-mediated response more comprehensively, we performed a total RNA-sequencing of Tal-treated mESCs, including p53-proficient and depleted cells. This showed that 81.5% of significantly deregulated genes in response to Tal are relying on the presence of p53, establishing p53 as a central transcription factor in this scenario (Figure 3A/B). Notably, the overall transcriptional response of Tal treated mESCs, depleted of p53, revealed a strong inverse correlation with p53-proficient cells, expanding the effect to a genome-wide level (Figure 3C). The effect was verified by qPCR (Figure 3D) and we could show that not only the transcriptional response is p53-dependent but also the cytotoxicity, as mESCs transfected with siRNA against p53 become resistant to Tal treatment (Figure 3E/F). This indicated for the first time that p53 activity is responsible for the observed cell death, rather than the initial DNA damage leading to the p53 signalling.

As p53-signalling can affect several hundred genes (Fischer, 2017), we performed a CRISPR/Cas9 screening to identify genes contributing to the Tal-dependent cytotoxicity in mESCs (Figure 4A/B). A synthetic viability approach resulted in PARP1 as the top hit to mediate Tal sensitivity, confirming previous observations of PARP1 dependency and validating the functionality of the screening setup (Figure 4C). The second most-detected hit was the Ten-eleven Translocation protein 1 (TET1) that is known for its function in 5mC oxidation in the context of active DNA demethylation (Figure 4C).

This result was intriguing as it provided a rationale for the peculiar sensitivity of mESCs to Tal, with TET1 (and TET2) being more highly expressed and active in pluripotent stem cells than in most other cell types. High activity of TET proteins, as well as of the Thymine DNA Glycosylase (TDG) are central to pluripotent stem cells, as they are key to the DNA methylation turnover mediating the transcriptional plasticity that is needed for later differentiation (Cortázar et al., 2011; Dawlaty et al., 2014; Mulholland et al., 2020b). We validated the dependency of the Tal-mediated cytotoxicity on TET proteins and saw that the combined depletion of TET1 and TET2 by siRNA provided resistance to the Tal treatment (Figure 4E and S3A/B), and so did the depletion of TDG or its catalytic activity (Figure 4G and S3D-G). This dependence on TET and TDG puts active DNA demethylation as an essential and central mediator in the scenario of Tal-mediated cytotoxicity.

We wondered whether the high activity of active DNA demethylation-induced BER is cause for the damage signalling upon PARPi-mediated inhibition of SSB repair. Consistent with the ATM-dependent phosphorylation of H2A.X, we detected an increase of DSBs in Tal-treated mESCs by pulse field gel

electrophoresis. The formation of DSBs was dependent on PARP1 (Figure S3H) but not on TDG (Figure 4H). We could further show that neither TET1/2 nor TDG depletion leads to the repression of p53 activation upon Tal (Figure 4I and S3I/J), which indicates that active DNA demethylation is unlikely the main cause of Tal-mediated DSB formation upstream of p53-activation.

We therefore addressed the contribution of TET/TDG-induced active DNA demethylation to the signalling after p53 activation. Using a publicly available TDG ChIP-seq dataset (Neri et al. 2015), we identified that 1'594 (86%) of the 1'859 significantly upregulated genes upon Tal, show TDG-binding in their promoter and/or in the gene body (Figure 5A). This is a significant overrepresentation when compared to a random set of regions coinciding with genes, and indicates TDG as a co-factor of p53 in gene activation. We then measured the transcriptional response of the p53 target genes *Alox5*, *Fas* and *Cpz* to Tal in TDG-depleted mESCs, and saw that their upregulation is drastically reduced (Figure 5B and S4A). ChIP-qPCR analysis showed that TDG is enriched at the promoter of these genes, but is not altered upon the induction of Tal (Figure S4B). In contrast to TDG, the presence of p53 at these promoters is strongly increased upon the treatment with Tal in a TDG-dependent manner (Figure 5C), which suggests that TDG facilitates the recruitment of p53 to its target genes.

Interestingly, we also observed that, besides canonical p53 target genes, Tal treatment also upregulated (retro-)transposable elements (TE), including LINEs, SINEs and DNA repeats, but most-notably LTR-containing ERVs (Figure S4D). Consistent with previous reports (Wang et al., 2007), we saw that LTRs display a higher proportion of stress-induced p53-binding when compared to other TEs (Figure S4E), which facilitated their expression upon Tal-treatment when compared to LTRs without p53-binding (Figure 5D). Consequently, de-repression of LTRs upon Tal was inhibited in p53-depleted mESCs not only of the top 50 upregulated LTRs (Figure 5E) but visible amongst all ~57 k significantly detected LTRs (Figure 5F).

We then examined whether the upregulation of these elements is also dependent on TET and/or TDG and showed that depletion of both drastically reduce the Tal-dependent de-repression of TEs, in an epistatic manner (Figure 5G and S4I/J).

We then set out to investigate the role of TET/TDG-mediated active DNA demethylation in the molecular mechanism underlying the p53-dependent activation of canonical targets as well as the TEs. As we had experimental evidence suggesting that TET/TDG-mediated active DNA demethylation facilitates the recruitment of phospho-p53 to and the activation of its targets (Figure 4G), we measured the levels of mC at the promoter of *Fas* and the LTRs "RLTR45" and "RMER19B". We reasoned that the mC levels would drop upon their Tal-mediated transcriptional activation. The steady-state levels of DNA methylation at these targets did, however, not significantly change (Figure S5A). We hypothesised therefore that not the methylation level might be the determining factor but the turnover rate of methylation (Ginno et al., 2020;

Parry et al., 2020) and we carried out a genome-wide mapping of single strand breaks (SSBs) as an approximation for TDG-activity as part of the BER-mediated active DNA demethylation.

We could observe that the method enriches for GpG-containing fragments (Figure S5E) and that SSBs generally enrich at regulatory and genic regions (i.e. enhancers, promoters, 5'UTRs (Figure S5G)). SSB-enriched regions are also coinciding with regions displaying increased TDG occupancy (Figure 6A) as well as p53 occupancy (Figure 6B). TDG-depletion caused a proportional reduction of SSB-enriched sites at p53 as well as TDG sites and displayed also a larger area of SSB enrichment, which suggests that the processing of oxidised mC derivatives is no longer properly coordinated (Figure S5H/I). More importantly, we could show that upon Tal, the SSBs are increased in regions with p53 (Figure 6D) and TDG occupancy (Figure 6E), which correlated with a facilitated transcription of LTRs that display separated and combined p53 or TDG-binding (Figure 6F). In TDG-depleted mESCs however, less SSBs are detected, which correlates with the reduced TE activation (Figure 6G) and hence suggested that TDG-mediated SSBs are necessary for the transcriptional activation of these elements in the p53-mediated response to Talazoparib.

Knowing that DNA repair-associated rearrangement can alter DNA accessibility (Madders and Parsons, 2020; Odell et al., 2013), which in turn strongly affects transcription (Figure 6K&L), we measured DNA accessibility by ATAC-seq. Regions with enriched accessibility coincided well with sites of p53 and TDG enrichment (Figure 6H&I). These levels were hardly affected by Tal treatment or TDG depletion on the genome-wide level, but a significant TDG-dependent increase of accessibility at p53-binding sites was observed upon Tal, when regions of significant ATAC enrichment were used that are unique to the examined conditions (Figure 6J).

We then reasoned that the de-repression of TEs poses a threat to genomic stability and mESCs vitality as they harbour the potential to rearrange genomic loci in different ways and that TE de-repression is therefore a cause for cytotoxicity. Introducing *in vitro*-transcribed dsRNA fragments of ERVs, namely their long terminal repeats, into mESC, we found indeed, a reduction of viability of mESCs that was less severe than the transfection of the dsRNA-mimicking poly I:C (Figure 7A). This transfection, as well as the treatment with Tal was accompanied by the upregulation of genes that are involved in the type I interferon response (Figure S6A-C and 7D) or the necroptosis pathway (Figure 7B and S6A-C) implicating the involvement of these responses in the cytotoxicity of Tal treatment.

In summary, in this collaborative work we could establish sequence of events that causes cytotoxicity in mESCs following treatment with the PARP-inhibitor Talazoparib. We propose an initial damage signal caused by PARPi-mediated impediment of DNA repair that leads to DSB formation and the phosphorylation of p53. This in turn activates a plethora of genes accompanied by a de-repression of transposable elements, which likely evoke a secondary necroptotic and/or interferon-like response. For the first time we show at a

genome-wide scale that the p53-dependent DNA stress response in mESCs entails the activation of TE and is relying on active DNA demethylation. Our data suggest that besides 5mC oxidation by TET proteins, SSB-generation by TDG and BER factors are required for transcriptional activation. Furthermore, we describe a potentially genome-destabilizing and cytotoxic effect of increased TE expression in stem cells that requires consideration in the treatment of patients with Talazoparib.

4.1.2 Contribution

My contribution to this collaborative work started with the observation that TET proteins are an essential contributor to the PARPi-dependent signalling cascade, which eventually leads to cytotoxicity in mESCs. To further elaborate the contribution of active DNA demethylation by TDG and BER, I investigated the rescue of cell viability upon the administration of the PARP1 inhibitor Talazoparib in two TDG knockout ESC model systems. After the conformation of the TDG dependency of this phenotype, I performed the experiments that were done in relation to TDG, and more. Namely:

- To verify the dependence of upregulation of canonical p53-responsive genes and repetitive elements on *Tdg* as well as other factors, I designed and executed the expression analyses by RT-qPCR upon Talazoparib or control conditions, such as Merbarone, Zeocin or Camptothecin. I also validated the transcriptional responses in ESCs with knockdown for *p53*, *Tet1* and *Tet2* by siRNA transfection.
- I investigated whether TDG-depleted cells mimic the phenotype of TET-depleted ESCs regarding cell-cycle profiles and apoptosis, utilizing flow-cytometry.
- To elucidate an involvement of the necroptosis pathway in addition to apoptosis, I designed and performed the expression analysis of necroptosis-related factors like *Ripk3* and *Zbp1* and attempted the knockdown of *Mkl*.
- I performed immunoblot analyses of Tal-treated and control ESCs for p53 phosphorylation to analyse the activation of p53 signalling cascade in dependence of TDG.
- To see whether TDG recruitment to promoters is affected by the presence of p53, I performed TDG-ChIP in p53 depleted cells.
- To link the signalling cascade to the enzymatic activity of TDG and chromatin accessibility, I established, performed and analysed the single-strand break formation on genome-wide level by next-generation sequencing (SSB-seq) and participated in the preparation of the ATAC-seq library.
- To connect and interpret the obtained genome-wide data, I performed the bioinformatic analyses of regions with SSBs in combination with the chromatin accessibility by ATAC-seq, the ChIP-seq data for TDG (Neri et al., 2015) and p53 (Li et al., 2012) with regard to the expression data upon Talazoparib treatment obtained in this study.

All in all, my conceptual inputs, experimental work and data analyses critically contributed to the conclusion of the manuscript that TET/TDG-dependent active DNA-demethylation is essential for the establishment of cytotoxicity via the p53-mediated stress response in mESCs. These data are represented in figures 3-7 as well as in the supplementary figures S3-S7. In addition, I contributed to the writing of the manuscript, to the preparation of figures and will perform the experiments demanded for revisions.

4.1.3 Manuscript, see appendix I

PARP inhibition induces cytotoxicity in mESCs by activating endogenous retroviruses

Jianming Xu^{1,3}, Simon D. Schwarz^{1,3}, Kapila Gunasekera², Roland Steinacher¹, Michael O. Hottiger^{2, 3*} and Primo Schär^{1, 3, 4*}

¹Department of Biomedicine, University of Basel, Mattenstrasse 28, 4058 Basel, Switzerland

²Department of Molecular Mechanisms of Disease, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

³These authors contributed equally.

⁴Lead Contact: Primo Schär, primo.schaer@unibas.ch

*Corresponding authors:

Michael O. Hottiger, michael.hottiger@dmmd.uzh.ch

Primo Schär, primo.schaer@unibas.ch

4.2 Covalent PARylation is an Integral Part of TDG-BER-mediated Active DNA Demethylation in Embryonic Stem Cells

The following manuscript, of which I will be co-first author as well, was submitted on 05.08.2021.

4.2.1 Summary

The poly-[ADP-ribose] polymerase 1 (PARP1) has been implicated in DNA damage repair for a long time. It has been shown to increase single-strand break (SSB) repair efficiency by recognising SSBs and facilitating recruitment of repair factors like the scaffold protein X-Ray Repair Cross Complementing Protein 1 (XRCC1), DNA Polymerase β (POL β) and DNA Ligase 3 (LIG3) (Fisher et al., 2007; Hanzlikova et al., 2016). SSBs arise frequently by spontaneous breakage of the DNA backbone or in the context of DNA base repair via active DNA demethylation. TET-mediated oxidation of methyl cytosine (5mC) creates formylcytosine (5fC) and carboxylcytosine (5caC) that are excised by the thymine DNA glycosylase (TDG) and then replaced with C by DNA base excision repair (BER) (Weber et al., 2016; Zhang and M. Kohli, 2013). The connection of PARP1 to active DNA demethylation through BER is still debated but it implies a potential role of PARP1 in the control of epigenetic DNA modifications and thereby transcriptional control (Ray Chaudhuri and Nussenzweig, 2017). It is widely accepted that the synthesis of poly-[ADP-ribose] (PAR) chains at histones in the vicinity of SSB damage and the auto-PARylation of PARP1 serves as a “signalling beacon” to recruit the BER machinery (Caldecott, 2008, 2014). But a role of PARP and PARylation in the repair process itself, i.e. by modification of proteins engaging in BER has not been described as of yet.

Mouse embryonic stem cells (mESCs) display high levels of TET as well as TDG activity (Mulholland et al., 2020b; Shen et al., 2013) and are therefore a frequently applied experimental model to study DNA (de)methylation dynamics. Along this line, mESCs also display higher levels of PARP1 expression as well as a substantial amount of PAR, when compared to mouse embryonic fibroblasts that are further differentiated (see appendix II, Figure S1A). To identify substrates of PARylation, we used a PAR-specific mass-spectrometry assay (Jungmichel et al., 2013) on protein extracts of mESCs that were either unchallenged, treated with the oxidative stressor H_2O_2 and/or the specific PARP1-inhibitor Talazoparib (Figure S1B). The preliminary data could confirm PARylation of the BER components XRCC1 and LIG3 via (data not shown), which was observed in previous studies with murine induced pluripotent stem cells and human HELA cells (Chiou et al., 2013; Hendriks et al., 2019), but which did not address the molecular impact of PAR-modification to these proteins.

To validate these observations and to investigate the impact of PARylation on the functional interaction of these DNA repair proteins, we took a biochemical approach. Far-western blotting demonstrated an interaction of XRCC1 and LIG3 with auto-PARylated PARP1 but not with unmodified PARP1 (Figure 1A).

Presence of AP-site containing DNA, generated by DNA glycosylases as a BER intermediate, activates the PARylation activity of PARP1 and we observed that in presence of NAD⁺, the combined incubation of PARP1 and unmodified APE1, XRCC1, LIG3 or POL β , resulted in a migratory shift of all the proteins from their calculated and known molecular weight. This showed that PARP1 is able to modify not only XRCC1 and LIG3 but also APE1 and POL β , and that the process was dependent on PARP1 activity, since a catalytic dead mutant or PARP-inhibition by PJ34 prevents PARylation of these substrates (Figure 1B).

To test the mechanistic implications of these modifications in BER, we reconstituted *in vitro* the entire pathway with the highly purified proteins (Figure S2A). In the presence or absence of PARP1 or the co-factor NAD⁺, the uracil DNA glycosylase UDG, XRCC1, APE1, POL β and LIG3 (Figure S2A) were incubated with an artificial DNA homo- or heteroduplex, the latter containing a G•U mismatch. UDG releases the mismatched U to generate an AP-site which is then incised by APE1, generating a SSB. PARylation of APE1, POL β , LIG3 and XRCC1 was observed upon uracil release by UDG (Figure 1C), indicating that PARylation can occur in BER before AP-site incision. To measure the association of PARylated BER factors to DNA, a biotin-labelled 60-mer DNA substrate containing a SSB was incubated with PARP1 and the BER factors in presence of NAD⁺ and precipitated with the help of streptavidin beads. This revealed that upon PARP1 action, PARylated BER factors dissociated from the DNA (Figure 1D/E & S2B/C) while unmodified proteins remained bound under physiological salt concentrations.

We also tested whether PARylation of BER factors affects their enzymatic activity, and observed that both PARylated APE1 and POL β display the same DNA endonuclease and DNA lyase activities as their non-modified counterparts (Figures S2D (left), S2E). Also, polymerisation efficiency of PARylated POL β was the same as of unmodified POL β and only final ligation by LIG3 was slightly impaired by PARylation (Figure S2D right). Consistently, PARylation of all BER factors reduced the overall repair capacity by around 30% (Figure 1F&S2F).

To understand the structural basis of altered DNA binding by PARylation, we used “Rosetta Comparative Modelling” to examine the impact of MAR or PAR coupling to APE1, POL β , XRCC1 and LIG3 *in silico*. Simulated addition of mono-ADP-ribose moieties at sites of previously observed PARylation, revealed that this modification is at the direct interface where APE1 contacts DNA, hence likely interfering with the binding affinity of APE1 to DNA (Figure 2A, S3A&B). The same is true for POL β (Figure 2B, S3C) and modelled PARylation of LIG3 (Figure S3E) revealed a potential steric hindrance and repulsion from the DNA by the increased negative charge of PAR-chains (Abd Elmageed et al., 2012).

These simulations indicated that PARylation of BER factors reduces their affinity to the DNA, providing an explanation for the increased dissociation from the DNA upon modification by PAR.

To investigate the potential impact of PARP1 activity on active DNA demethylation, we analysed the properties of TDG interacting with PARylated PARP1. Far-western immunoblotting revealed that TDG interaction with PARP1 was massively enhanced by the PARylation of the latter, which is dependent on the cooperation of N-terminal and the core domains of hTDG (residues 1-308, Figure 3A). We could confirm that TDG is also PARylated at its N-terminus, the core domain as well as the C-terminus (Figure 3B, S4A) and that the PARylation of TDG results in a 2-fold increase of 5caC base excision activity, exceeding the product/substrate rate of 1 (Figure 3C). This indicates a multiple turn over scenario where product inhibition of TDG is overcome. Analysis of the binding properties of PARylated TDG to a mismatch-containing DNA oligomer also revealed a reduced affinity of PARylated TDG to the DNA, indicating that PARylation of TDG stimulates BER by the increase of its dissociation from the processed substrate (Figure 3D). In a “complete” scenario including all BER factors, the presence of active PARP1 and subsequent PARylation of the examined proteins increased the BER efficiency towards 5caC *in vitro* by nearly 20% (Figure 3E, S4B). It is worth noting that under these circumstances, only max. 50% of each BER protein was PARylated, suggesting the potential of an even higher repair efficiency if the total of proteins were PARylated.

To corroborate these observations *in vivo*, we investigated the occurrence and characteristics of SSBs in mESCs upon the treatment with Tal, making use of our dataset on genome-wide SSB mapping (also see manuscript in 4.1.3/appendix I). mESCs display a particularly high level of active DNA demethylation and based on the so far accepted concept that PARP-inhibition would reduce SSB repair efficiency, we expected a higher amount of SSB occurrence in the genome upon treatment with Tal. This, however, was not the case, as we observed about 40% less regions of significant SSB enrichment in mESCs after Tal treatment (15'528 vs. 28'543). While this effect could be explained by an increased number of random SSBs that do not reach the detection threshold, the signal in regions of SSB enrichment upon Tal treatment is actually stronger than in the untreated ESCs, thereby displaying an even higher signal-to-noise ratio. It therefore seems not to be a dominant effect and rather allows to explain that not just SSB repair but also the initial generation of SSBs can be reduced in the presence of a PARP1-inhibitor.

To focus on SSBs that originate from active DNA demethylation, we used a published dataset that measured accumulation of 5caC in the absence of TDG to mark regions of active DNA oxidation by TETs and excision by TDG (Shen et al., 2013). This revealed that in DMSO-treated ESCs, 24% of all SSB peaks overlap with regions of 5caC excision (6'877) while in Tal-treated cells the coincidence was reduced to 18% (2'792) (Figure 4A). This proportional decrease in the overlap of regions with SSB enrichment and 5caC-excision was also accompanied by a (slight) increase of SSB enrichment further distal to the centre of 5caC excision. Notably, this effect was also apparent when regions of SSB enrichment were analysed that coincide with CpGs with above-average activity of TET-mediated oxidation (Figure 4B, Ginno et al., 2020). Read-based analysis within areas of SSB enrichment and 5caC-excision revealed that Tal reduces the SSB signal to similar

levels as observed in TDG-depleted ESCs either treated with Tal or without. These data indicated that PARP1-mediated PARylation promotes TDG-dependent SSB-formation and hence, TDG/BER-mediated active DNA demethylation.

Addressing the question whether SSBs associated with active DNA demethylation contribute to gene activation, we examined whether the treatment with Tal affects the expression of genes with ongoing active DNA demethylation in their promoter. Treatment of ESCs with Tal significantly reduced SSBs in the promoters of all those genes that also exhibited 5caC excision (Figure 4D). We then analysed expression levels (also see 4.1/appendix I) of genes that show 5caC generation in their promoter and those that additionally show significantly enriched SSBs in their promoter. We saw that genes with SSB enrichment in their promoter exhibit a significantly higher steady-state level of expression when compared to genes without any mark or those with 5caC-excision but no SSB enrichment in their promoter (Figure 4E). This indicates that SSB generation indeed supports more efficient transcription. The treatment with Tal also reduced the upregulation of Tal-responsive genes with significant SSB enrichment in their promoter (data not shown), when compared to those genes without SSB enrichment, suggesting that unhindered SSB generation in the promoter is beneficial for an efficient transcriptional response. Altogether, these data indicate that PARP1-inhibition reduces generation of SSBs at sites of targeted active DNA demethylation, which in turn decreases transcriptional activity at these genes.

Current models of DNA damage formation and repair predict that the vast majority of SSBs occur randomly throughout the genome by the repair of spontaneous DNA lesions, leading to engagement of PARP1 in their repair in an equally random pattern (Caldecott, 2008). This concept does not take into account the high number of targeted SSBs ($5 \times 10^5 - 9 \times 10^5$) continuously generated in ESCs by active DNA demethylation (Steinacher et al., 2019). To contrast the occurrence of instructed, non-random SSBs at sites of active DNA demethylation with random SSBs associated with spontaneous DNA damage, we compared SSB formation at sites of known active DNA demethylation with sites outside such regions. Currently, there are no genomic maps for 8-oxoG generation in ESCs available, which could be used as a proxy for damage through random oxidation. We therefore used simple repeats (including micro-satellites) as well as regions with a skew in the C/G nucleotide content, assuming that, rather than through active DNA demethylation, SSBs at such sites would occur through random processes such as stochastic base damage, base mis-incorporation and/or repair during DNA replication as well as R-loop and G-quadruplex formation (Crossley et al., 2019; Lin and Wilson, 2012; Su and Freudenreich, 2017). These analyses revealed that the proportion of SSBs at sites with predominantly random DNA damage significantly increased (+25% at GC-skews, +95% at simple repeats) when mESCs were treated with Tal. This increase was also visible at the level of SSBs within inter- or intragenic regions harbouring GC skews (Figure 4F, S5C), painting the picture of a redistribution from targeted, active DNA demethylation-associated SSBs in unchallenged mESCs to more randomly occurring

damage caused by reduced BER through PARP1-inhibition.

Interestingly, regions with a GC-skew located at promoters, suggestive of a regulatory role, again showed a reduction of SSBs after the treatment with Tal (data not shown) and thereby corroborate the previous observations from a different point of view.

In summary, we demonstrate in this report a new integral role of PARP1 in BER, where covalent PARylation of BER factors occurs as early as AP-site generation and throughout BER, leading to the dissociation of the BER factors from the DNA (and each other) after repair. Consequently, this increases the turnover efficiency of BER and active DNA demethylation. This important regulatory role of PARP1 adds to/questions the so far dominant view postulating that PARylation of histones and PARP1 itself leads to the non-covalent recruitment of DNA repair factors to DNA damage after SSB-sensing by PARP1. Applying this knowledge to the recently generated SSB-seq data allowed to differentiate instructed SSBs originating from active DNA demethylation, from those that are randomly occurring by DNA damage. Using RNA-seq data, we could show that genes with ongoing active DNA demethylation in their promoter display a generally higher transcriptional activity than genes without this signature and the application of Tal reduces the transcriptional inducibility of these genes. All in all, we tighten the connection between active DNA demethylation and SSB generation and propose a general role of PARylation in BER for transcriptional regulation.

4.2.2 Contribution

My main contribution to this manuscript consists of the bioinformatic analyses of SSBs and RNA-expression data in the context of PARP1 inhibition. More specifically I could show that:

- PARP1 inhibition leads to proportionally less SSB-enriched regions coinciding with regions of active DNA demethylation - defined by 5caC accumulation upon TDG KO (Shen et al., 2013).
- Along this line, SSB signals within regions of 5caC excision were also reduced upon PARP-inhibition.
- This observation was also true for the whole of promoter regions of genes that exhibited 5caC accumulation.
- Through the analysis of mRNA-seq data I showed that genes with ongoing active DNA demethylation in their promoters and showing an enrichment of SSBs displayed increased basal transcription. The reduced inducibility of expression upon Talazoparib indicates that SSBs facilitate transcriptional activity.
- In contrast to regions with SSBs targeted by active DNA demethylation (promoters etc.), regions that are prone to genomic instability and thereby random damage (simple repeats, GC-skewed regions), show increased amount of SSBs, due to the lack of PARP1-activity.

These observations critically contributed to the understanding that within the genome of ESCs, SSBs can be random as well as targeted and that PARP inhibition leads to a re-distribution of SSBs from active DNA demethylation to regions of spontaneous damage. Additionally, the data indicates that loci undergoing targeted SSB formation in the context of active DNA demethylation, especially gene promoters, do so in a repeated and cyclic manner, as they harbour a basal level of SSBs higher than that in regions with random damage. These analyses are reflected in the main figure 4 as well as the supplementary figure S5. In addition, I took part in writing of the manuscript.

4.2.3 Manuscript, see appendix II

Title: Covalent PARylation is an Integral part of TDG-BER-mediated Active DNA Demethylation in Embryonic Stem Cells

Authors:

J. Xu ^{1,2†}, S.D. Schwarz^{1†}, K. Gunasekera^{2,3}, E. Ferrari², M.O. Hottiger², P. Schär^{1*}, R. Steinacher^{1*}

Affiliations:

¹Department of Biomedicine, University of Basel; Basel, Switzerland.

²Department of Molecular Mechanisms of Disease; University of Zurich, Switzerland.

³Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Switzerland.

*Corresponding authors. Email: *Primo.Schaer@unibas.ch, *Roland.Steinacher@unibas.ch

† These authors contributed equally to this work

4.3 Active DNA demethylation Evicts H2A.Z to Facilitate Pause Release of RNA Polymerase 2

The most recent data of my PhD thesis is summarized in a third manuscript. The data collection is not entirely complete but should serve as the framework for a well-supported hypothesis. Missing experiments are discussed and will be done in the near future.

4.3.1 Contribution

Generation of the H2A.Z ChIP-seq dataset was performed by Zeinab Barekati and basic bioinformatic processing (adapter trimming, alignment) was done by Robert Ivanek. Kapila Gunasekera and Florian Geier gave helpful inputs in further analysis of the SSB, ATAC and PRO-seq datasets. Apart from that, I performed all underlying experiments and the bioinformatic analyses. I conceived the hypothesis, designed the figures and wrote the manuscript.

4.3.2 Manuscript, see appendix III

Active DNA demethylation Evicts H2A.Z to Facilitate Pause Release of RNA Polymerase 2

Simon D Schwarz¹, Jianming Xu¹, Kapila Gunasekera², Zeinab Barekati¹, Florian Geier^{1,3}, David Schuermann,¹ Primo Schär^{1,4}

¹ Department of Biomedicine, University of Basel, Basel, Switzerland

² Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Switzerland

³ Swiss Institute of Bioinformatics, Basel, Switzerland

⁴ Corresponding author

Keywords: Active DNA demethylation, TDG, Single-Strand Breaks, H2A.Z, Pause Release

4.4 Additional preliminary data: Intragenic DNA methylation, TDG and SETD2

I introduced in chapter 2.3.1 that intragenic DNA methylation and the oxidised 5mC derivatives are biological marks associated with transcriptional regulation. Their deposition is, at least partly, based on the co-transcriptional trimethylation of H3K36 by the histone methyltransferase SETD2 (Baubec et al., 2015; Rajagopalan et al., 2021; Weinberg et al., 2019). The concomitant recruitment of DNMT3A/B to H3K36me3 and the generation of 5mC marks, may reflect dynamic epigenetic modifications important for transcriptional activity, with the potential to recruit TET and TDG proteins. This notion led to my initial working hypothesis postulating SETD2 as a potential upstream factor in TDG recruitment to the gene body. I reasoned that the connection of TDG with H3K36me3, an epigenetic mark almost exclusively associated with (actively transcribed) gene bodies, could help distinguishing the roles of active DNA (de)methylation at transcriptional regulatory regions from that during elongation. The implicated sequence of events of transcription-mediated deposition of H3K36me3 marks and concomitant deposition of 5mC is attractive as it could, on the one hand, explain how the active DNA demethylation machinery is recruited to the gene body and, on the other hand, provide the basis for understanding its functional consequence in the context of elongating transcription. While the association of 5mC in gene body and active transcription is well-described, the fact that H3K36me3 marks and 5mCs are potent recruiters of polycomb proteins that deposit H3K27me3 marks has been neglected so far. H3K27me3 is well-known for mediating transcriptional silencing and as an early step of chromatin compaction, circumstances that are clearly unfavourable for efficient transcription (Abed and Jones, 2012; Cai et al., 2013; Musselman et al., 2012). I therefore envisaged that cyclic methylation and demethylation of the CpGs in gene bodies could provide a mechanism to counteract premature chromatin condensation through deposition of H3K27me3 marks, during or shortly after transcription. If needed, it would likely also facilitate a quick re-expression of the gene. Altogether, this would extend the molecular roles known of H3K36me3 in gene bodies and, for the first time, describe a role of intragenic DNA (de)methylation as part of an epigenetic, transcriptional bookmark (Jeziorska et al., 2017; Tompkins et al., 2016). Towards the testing of this hypothesis, I present the so far obtained preliminary data and observations in the following chapters. No additional methods were used than those presented in the previous manuscripts.

4.4.1 H3K36me3 Participates in Recruitment of TDG to the Gene Body but not the Processing of oxmCs

To address the potential association between TDG abundance and H3K36me3 marks in the chromatin of murine ESC, I used a published H3K36me3 ChIP-seq dataset (ENCODE: GSM1000109) and superimposed it with TDG ChIP-seq data from our laboratory (Wirz, 2014, Barekati et al. manuscript in preparation). This identified 6'252 (or 18%) of 34'571 significantly enriched intragenic TDG peaks (\log_2 -fold change TDG WT/KO >1.5) in regions with a H3K36me3 signature, an overlap significantly higher than expected by chance (Chi-square test: 0.0002056; Figure 11A/B). The inverse analysis also revealed a slight, but significant enrichment of H3K36me3 marks over intragenic TDG peaks compared to same number of random genomic fragments with same size (Figure 11C). However, SSBs appear to be depleted across the centre of the H3K36me3 mark and increased as much as 1'500 bp away from the modified histone. These data are consistent with a role of SETD2 or H3K36me3 in the recruitment of TDG. BER-mediated processing of oxmC, however, does not seem to take place in the vicinity of the histone H3K36me3 modification.

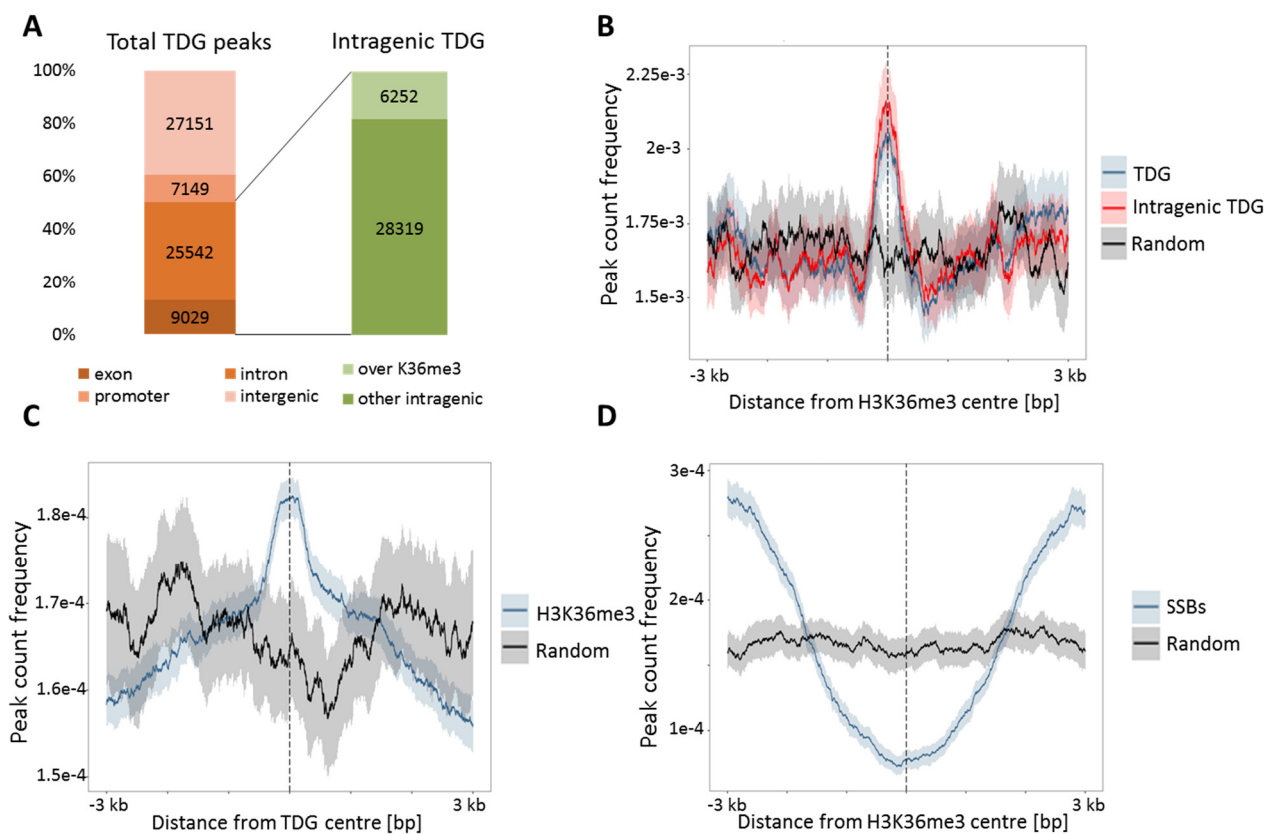


Figure 11 ChIP-seq signal of intragenic TDG coincides with H3K36me3 marks

A left: TDG peak distribution over genic features, right: Fraction coinciding with H3K36me3 or not **B**: Density plot of either all or intragenic TDG peaks around H3K36me3 marks. **B** Distribution of all TDG (blue) and only intragenic (red) regions with TDG enrichment over H3K36me3 enriched regions compared to a randomised peak set (black). **C** Analysis of H3K36me3 marks distributed over enriched intragenic TDG regions. **D** Distribution of regions enriched for SSBs in wild type cells over H3K36me3 marks. Shadings around the curves indicate the 95% confidence interval.

4.4.2 TDG and SETD2 are Required for Transcriptional Efficacy

To investigate whether localisation of TDG is affected by SETD2, I disrupted the reading frame of the *SetD2* gene at the 5' end of exon 4 in our inducible *Tdg* KO mESC line, using CRISPR/Cas9 (gRNA: 5'-ACGGCTACGTCTCTGGTAA-3'). I established three clones (SetD2 KO #1-3) that show a reduction of H3K36me3 protein expression to less than 5% of the level in wild type cells. This low residual level of H3K36me3 is apparent in most publications using *SetD2* KO cells and might be explained by an unspecific activity of NSD (Nuclear Receptor Binding SET containing domain) proteins, usually responsible for H3K36 mono and dimethylation (Huang and Zhu, 2018). The following experiments were done with one or several of these SetD2 KO clones.

I treated ESCs with 2 mM dithiotreitol (DTT) for 30 to 60 minutes to evoke the unfolded protein response (UPR) (Oslowski and Urano, 2011; Yang et al., 2016) resulting in an upregulation of the genes *Ddit3* (DNA Damage Inducible Transcript 3) and *Herpud1* (Homocysteine-responsive Endoplasmic Reticulum-resident Ubiquitin-like Domain Member 1). Analysis of this stress response revealed that mESCs depleted of either protein are not able to upregulate the two genes to the same extent as WT mESCs (Figure 13A). The effect of altered transcriptional regulation was visible with all three SetD2 KO clones and indicates the necessity for TDG and SETD2 for an efficient transcriptional stress response. These experiments were repeated only twice and need further investigation to obtain solid data.

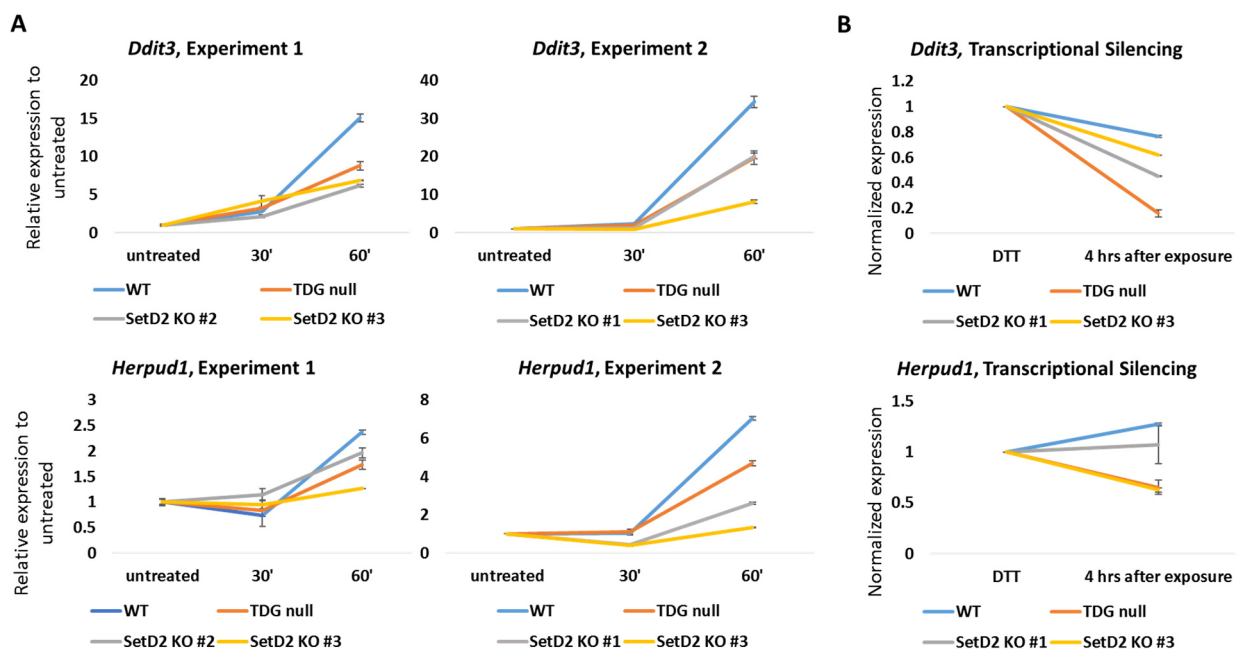


Figure 12 Transcriptional response and silencing in ESCs depleted of TDG or SETD2

A qPCR analysis of mRNA expression levels of indicated genes upon treatment with 2 mM DTT for the indicated time. The two experiments are shown separately as different SetD2 KO clones were used. Error bars show the standard deviation of technical duplicates. **B** Expression levels of indicated genes were measured 4 h after the removal of DTT containing medium and normalized to the expression level after 60 min of exposure. Error bars show standard deviation of technical triplicates

This observed effect on stress response can be either attributed to reduced transcriptional initiation due to lack of DNA demethylation in the promoter when TDG is missing, or to reduced transcriptional elongation by RNAP2 when SETD2 is missing. When DTT as the stressor was removed from the medium and the expression levels measured again 4 hours after, mESCs depleted of either TDG or SETD2 exhibited a stronger reduction of the stress-response genes than the WT ESCs, consistent with a premature transcriptional inactivation of the genes (Figure 12B).

To gain better insight, the memory aspect of transcription should be addressed in a manner not only based on transcriptional levels after the removal of a trigger, but by the potential of cells to reactivate the genes after a certain time. Furthermore should the mechanistical aspect be approached by measurement of H3K27me3 deposition as a sign of gene silencing. In this regard, it should also be considered to perform the essays in ESCs cultured in serum or upon further differentiation, as naïve mESCs in 2i-medium are known to have lower levels of H3K27me3, which might dampen the effect (Marks et al., 2012).

I also investigated transcriptional responses in ESCs depleted of TDG or SETD2 that were undergoing differentiation either by cultivation for four days in serum-containing medium without leukaemia inhibitory factor (LIF), or by the addition of all-trans retinoic acid (RA), which drives differentiation towards neuronal lineages (Bibel et al., 2007; Yu et al., 2012). These approaches showed a weak trend toward a faster silencing of examined genes, however also displayed very high levels of variability between experiments and were therefore discontinued.

4.4.3 Loss of SETD2 Affects Recruitment of TDG to the Chromatin

To investigate whether this reduced efficiency of transcriptional activation coincides with differential recruitment of TDG, I performed TDG ChIP followed by qPCR analysis along the gene body of the two DTT-responsive genes in the SetD2 KO #1 cell line. In two experiments, indeed, TDG enrichment after 30 minutes of DTT treatment was overall affected by the lack of SETD2 and tended to be less abundant in the gene body of both genes (Figure 13A/B) and also reduced in the promoter of *Ddit3*.

This observation would further benefit from an expanded analysis of targets and by methods yielding a complete coverage of genic regions. Both may be achieved by ChIP followed by next-generation sequencing, making use of the novel TDG-antibodies that are currently being established in the Schär lab.

To follow up on this observation with a different approach, chromatin fractionation (protein fractions provided by Federica Richina) was applied to address whether this reduced TDG recruitment is a global genomic phenomenon. There was no overt change in the distribution of TDG in nuclear soluble and chromatin fraction in the absence of SetD2 (SetD2 KO #1 and #2) but we observed a strong increase of SUMOylated TDG present in the soluble fraction of SetD2 KO #1 and #2 cells (Figure 13 C). Densitometric analyses revealed a 2.3 - 4.2-fold increase of TDG-SUMO in the nuclear soluble fraction of SetD2 KO #1 and a less pronounced 1.43 - 1.76-fold increase in SetD2 KO #2, when normalized to histone H3 and the same

fraction in TDG WT ESCs. When the ratio of total soluble TDG (modified and unmodified) to the total of the insoluble fraction was compared, I saw that both, SetD2 KO #1 and #2 cells, exhibited a 2.8 and 3.5-fold higher level of TDG in the soluble fraction, respectively.

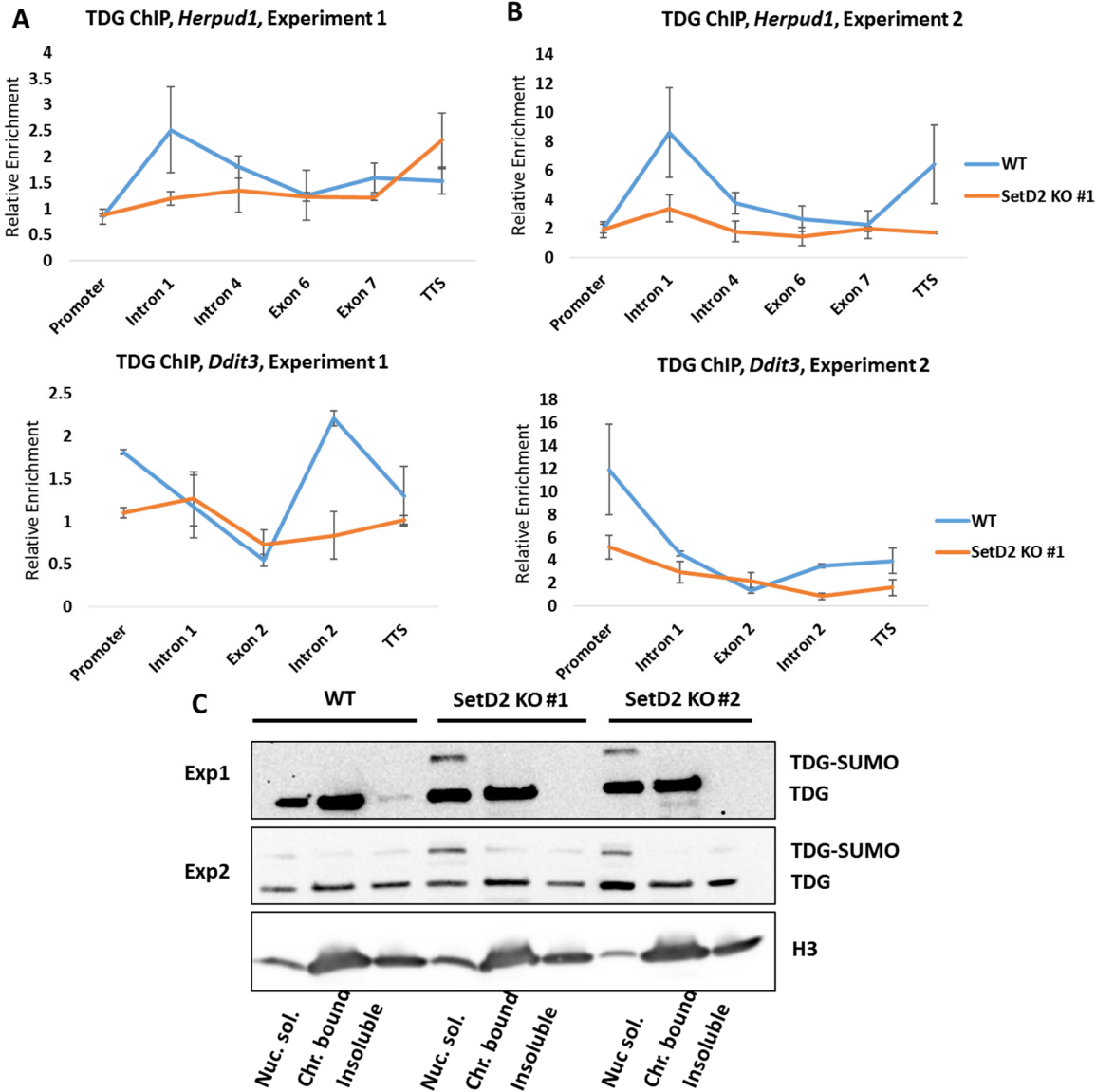


Figure 13 Altered association of TDG to chromatin upon loss of SETD2

A TDG ChIP-qPCR along DTT-responsive genes after 30 min of DTT treatment. Values were first normalized to the input and then to a negative control region on chromosome 9. **B** Same as in **A** but the ChIP-protocol included an additional fixation with the crosslinker disuccinimidyl glutarate (DSG). **C** Immunoblots of chromatin fractionations in indicated SetD2 KO clones and WT mESCs without challenge. Fractions are indicated below the blot (Nuc. sol.: nuclear soluble; Chr. Bound: chromatin bound)

While the presented observation points towards the hypothesis, the reduced enrichment of TDG at these genes upon transcription could also be explained by altered dissociation rates of TDG and needs further examination. Nonetheless, differential recruitment of TDG to the gene body, in dependence of transcriptional levels or SETD2 has not been described until now.

The substantial increase of TDG-SUMO, which is known to significantly reduce DNA binding affinity of TDG

and the increased total soluble TDG in SetD2 KO mESCs, indicates a considerable reduction of chromatin-bound TDG, even though there was no reduction of TDG recruitment visible on the global level. However, a potential underlying mechanism how SETD2 depletion is affecting the levels of SUMOylated TDG remains elusive. Changed expression of SUMO-ligases or proteases are, for example, unlikely to be the cause as these proteins are generally highly expressed in mESCs (Theurillat et al., 2020) and not changed upon SetD2 KO (Zhang et al., 2014). Possible explanations could include an increase of TDG SUMOylation on the chromatin, through increased or mistargeted activity of BER, but also through a mistargeting of the SUMO-ligase that would otherwise modify SETD2 (Hendriks et al., 2017; Martinelli et al., 2018). Another possibility presents hyper-SUMOylation as a general stress response upon the loss of SETD2 (Enserink, 2015). Experiments to measure SSB generation upon depletion of *SetD2* or the competence of the SETD2 SUMOylating ligase towards TDG, could answer these questions. Which SUMO ligase that could be seems unfortunately be unknown at this point.

4.4.4 TDG and SETD2 Double Knockout Affects Cell Viability

Excising *Tdg* in mESCs that are depleted of SETD2 allows for addressing the epistatic relationship between the two gene functions. When preparing such an experiment, I realised that a considerable amount of SETD2 KO mESCs died or showed signs of dying 24 h after the addition of OHT, which triggers the excision of *Tdg*. Measuring metabolic activity of the mESCs as an approximation for cell viability by WST-8 assay, I could confirm that cell viability is reduced up to 50% upon the combined KO of TDG and SETD2 (Figure 14A), suggesting a synthetic rather than epistatic phenotype. To explore the cause of cell death, I examined cell cycle profiles and observed a very small increase of 4-6% of cells in G1-phase upon the depletion of TDG in the *Setd2* KO background (Figure 14B), as well as near doubling of sub-G1 cells in SetD2 KO #1 (Figure 14C/D). While these effects were rather small, analysing the activity/expression levels of proteins relevant for cell cycle progression, like cyclin-dependent kinases, could more clearly reveal potential effects on cell cycle. The synthetic sickness of TDG/SETD2 DKO ESCs was observed repeatedly but varied in its strength and among the three SetD2 KO clones with clone #2 being the one with the lowest sensitivity to TDG depletion. This mESC clone-dependent specificity could be rooted in clonal adaptation from the generation of the cell line, and the high variation could point towards a generally high sensitivity of these cells to environmental changes (e.g., stress or differentiation). To gain more insight into this, surviving ESCs depleted of both proteins could be cultured and examined for basic abilities in transcriptional regulation or differentiation. While the observation of synthetic sickness upon TDG and SetD2 KO is very interesting and links TDG and SetD2-dependent functions, it provides no mechanistical explanation as of yet. Further investigation should address the functional relationships of TDG and SetD2 in transcription associated processes with methods by that provide mechanistic insight as mentioned.

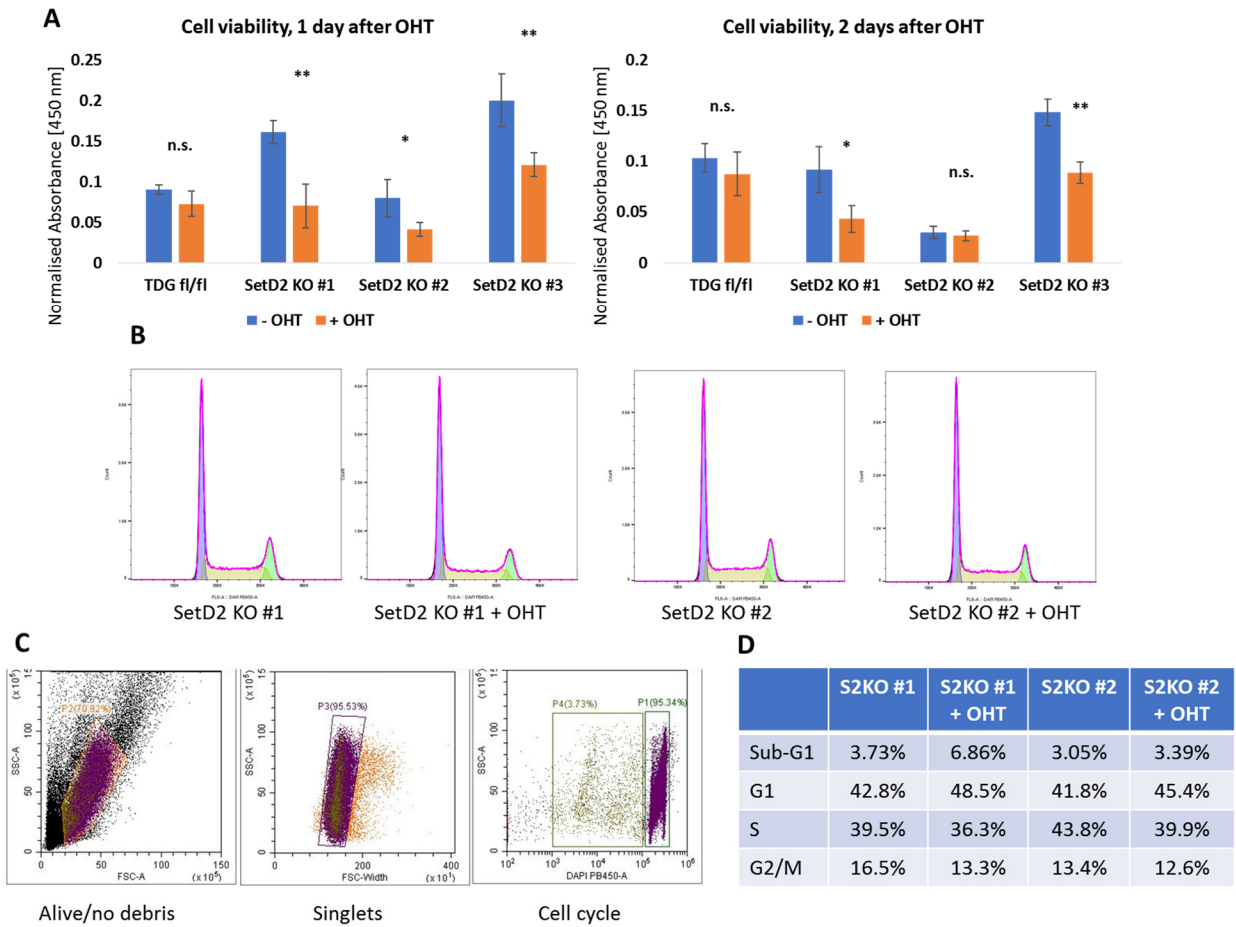


Figure 14 Synthetic lethality after TDG and SETD2 disruption in mESCs

A Cell viability with WST-8 cell counting kit, 24 h or 48 h after depletion of TDG by OHT treatment. mESCs were incubated for 2 h in the presence of WST-8 solution and absorbance at 450 nm was measured. Error bars indicate SD of technical quadruplicates from one experiment. Asterisks indicate statistical p-values of: *: $p \leq 0.05$, **: $p \leq 0.01$. **B** Cell cycle distribution of two SetD2 KO clones with additional TDG KO (+OHT), 24 h after OHT administration. **C** Gating strategy to count cells in sub-G1 phase. **D** Quantification of ESCs in according cell cycle phases.

5 Conclusion and Outlook

The impact of epigenetic modifications, in particular DNA methylation, on transcription has first been observed in the early 1980ies (Bird, 1984a, 1984b) and conceptually been acknowledged in the early 1990ies (Razin and Cedar, 1991). At that stage, few have envisaged how complex and all-encompassing the epigenetic concept of gene control will become and how much it will drive research in the fields of developmental and cancer biology, genetics and transcription. 30-40 years later, research in these fields is still accelerating and while general concepts can be grasped, big efforts have still to be undertaken to explain epigenetic mechanisms and their functional relationships in detail.

The reversibility of epigenetic marks is a key to their biological role in gene expression programming and re-programming but also brings along a high level of complexity. While initially thought that epigenetic histone and DNA modifications are stable in a given type of cell, it has been realised that these modifications are in a constant flux even in the time frame of one cell cycle (Kheir and Lund, 2010). This concept was brought to a new level upon the discovery of an active mechanism to remove methylated bases in the DNA. While in animal cells there is no single enzyme to remove 5mCs, as there is plant cells (Agius et al., 2006; Brooks et al., 2014), oxidation of 5mC by TET proteins to 5fC and 5caC provides a substrate for excision by TDG and repair by the BER pathway to reinstate an unmodified C (He et al., 2011; Ito et al., 2011; Maiti and Drohat, 2011). This revealed the possibility of a cyclical/constant turnover of DNA methylation and DNA demethylation (Kohli and Zhang, 2013; Parry et al., 2020). A comparable mechanism based on deamination of 5mC was already proposed earlier but the biological impact to active DNA demethylation was revealed to be low (Métivier et al., 2008; Spada et al., 2020). Dynamic DNA methylation, which was notably also shown to occur in post-mitotic cells (Bayraktar and Kreutz, 2017) also explains the complex phenotypes observed in mice and mESCs depleted of the involved factors (Schuermann et al., 2016). The TET proteins and TDG, for example, are part of the same pathway but their depletion in mESCs or mice leads to very different phenotypes. While TET1/2/3-depletion (TKO) leads to hypermethylation at regulatory regions and consequential deregulation of ~2500 genes during early differentiation, TKO ESCs can still contribute to proper embryo formation, albeit very poorly (Dawlaty et al., 2014). TDG-deficient mESCs on the other hand, barely show any impediment in homeostasis, revealed much fewer (~300) differentially expressed genes along with minor changes in DNA methylation during early differentiation, but nonetheless fail to develop into a mouse past the embryonic day 11.5 with no conclusive pathological pattern (Cortázar et al., 2011; Cortellino et al., 2011). These convoluted phenotypes now start to be understood when the dynamics and turnover of active DNA demethylation are being examined. Along this line it has to/will be acknowledged that interfering with a process that mediates/modulates a context-specific frequency of DNA methylation and demethylation (Ginno et al., 2020), is bound to result not in a black and white, but a variety of phenotypes, and is distinct to each factor. Like this, DNA hypomethylation phenotypes can be explained

when DNA demethylation factors are inhibited and vice versa (López-Moyado et al., 2019). While this reasoning seems to be very “simple”, it is yet very generalized and only few studies have contributed to this perception so far. To increase the understanding and support this principle, it is therefore imperative to examine active DNA demethylation in the control of transcription by including the interactions of the processes as well as the DNA methylation marks, with regard to the frequency of their enzymatic activity and turnover, respectively.

For my PhD, I therefore decided to investigate the role of TDG in the context of transcriptional regulation with locus-specific and genome-wide approaches. With a recently developed assay to detect SSBs along with classical molecular genetic approaches, I approached the question of TDG-induced BER-activity *in vivo*, and the interference of cellular homeostasis with transcriptional activators as well as DNA repair inhibitors, allowed me to examine active DNA demethylation from different angles.

Active DNA Demethylation in Genome-wide Transcriptional Response and Cytotoxicity, Mediated by PARP1-inhibition

It was previously shown that TDG interacts with the transcription factor p53 (Kim and Um, 2008). This conclusion was based mostly on *in vitro* data and verification of the cooperative effect on transcription by a transient overexpression approach including transcriptional readout for single candidate genes. In our study on the cytotoxicity by PARP1-inhibition in naïve ESCs, we established that TDG is necessary to transmit a stress response following the activation of p53, involving more than one and a half thousand genes at once. Additionally, the first examination of SSB formation in this context allowed to clearly link the catalytic activity of TDG and BER to the measured transcriptional response. This supports the concept of TDG-mediated active DNA demethylation as a basic process in transcriptional stress response.

Early attempts to explain why TET and TDG-depletion leads to the rescue of PARPi-mediated cytotoxicity, included the high DNA demethylation activity in naïve ESCs (Mulholland et al., 2020b; Tovy et al., 2017). Inhibition of PARP1 would lead to the hindrance of the associated BER process and lead to increased levels of DSBs caused by the collision of BER lesions with replication forks. We could, however, show that DSB formation as well as p53 activation is independent of TET and TDG, and therefore rule out a dominant role of active DNA demethylation in the activation of the stress response. This raises the question of the initial increase of DNA damage. Possible answers lie in the multiple engagement of PARP1 in DNA repair that is independent of TET/TDG-initiated BER. For example, PARylation was shown to attract XPC and PARP1 to interact with DDB complex of the NER pathway. Furthermore PARP1 is crucial for the detection of free DNA ends of DSBs and mediates recruitment of the protein Meiotic Recombination 11 Homolog 11 (MRE11) to reliably execute HR-mediated repair (reviewed Ray Chaudhuri and Nussenzweig, 2017). It is therefore conceivable that the PARPi treatment impedes the overall DNA damage repair in cells, which, however, seems manageable for most (somatic) cell types.

Addressing the mESC-specific toxicity mediated by TET/TDG-activity downstream of p53 activation, we also showed that the Tal-induced stress response caused a de-repression of several thousand repetitive elements. Despite mESCs being mostly known for a non-functional interferon (IFN)-response, we saw that this burst of dsRNA abundance elicits the expression of genes associated to the type I IFN response as well as necroptosis, both of which are connected to retroviral infections and sensing of dsRNA (Kalai et al., 2002). It is currently being discussed that the reason for the suppressed IFN response is based on the fact that hypomethylated DNA of mESCs not only facilitates flexible gene expression but also the controlled release of repeat elements as advantageous for pluripotency and potentially evolutionary fitness (Eggenberger et al., 2019; Mulholland et al., 2020a). Hence, the cause of PARPi-mediated cytotoxicity in mESCs might lie in the unlikely hijacking of a type of programmed cell death upon the sheer overload of “foreign” transcripts. Control of repetitive elements is mediated by the epigenetic machinery. Although silencing of these elements by DNA (and histone) methylation is a well-known and accepted concept (Weisenberger et al., 2005), it was observed that DNA methylation at these elements are not static. Namely, oxidized 5mC derivatives were reported at repeats without apparent expression (Deniz et al., 2018; de la Rica et al., 2016; Shen et al., 2013). It is therefore possible that active DNA demethylation also partakes in transcriptional silencing at such loci. This finds support in my PRO-seq data, where a subset of ~13'000 repeats, including SINEs, LINES, DNAREPs and LTRs, were upregulated at least 2-fold, simply by the depletion of TDG (data not shown). How this silencing through DNA (de)methylation turnover is achieved remains elusive. It could be that an unknown interactor of either TETs or TDG leads to recruitment of transcriptional repressors as was reported for the turnover of histone H3.3 to recruit H3K9 methylation marks, which mediate silencing (Elsässer et al., 2015; Schlesinger et al., 2017).

A Role of Active DNA Demethylation in Transcriptional Elongation

While approaching the question on how TDG and active DNA demethylation affect transcription independent of promoter demethylation, the PRO-seq analyses revealed a more general aspect of active DNA demethylation. TDG depleted mESCs showed an increase in promoter-proximal pausing of RNAP2, which was even more pronounced at genes that otherwise become upregulated upon Tal in WT mESCs. Information about the potential role of DNA methylation in RNAP2 pausing is so far very scarce and indirect. The reports available were either describing CTCF-mediated pausing of RNAP2 to facilitate splicing (Shukla et al., 2011) or describing increased RNAP2 pausing at regions with a high GC-skew (including CpG islands) but without any mechanistical approach (Kellner et al., 2015). In this regard, my analyses provide the basis for a novel mechanistic concept that controls RNAP2 pause release via a DNA (de)methylation-mediated modulation of H2A.Z deposition downstream of the TSS. Recent reports showed that H2A.Z eviction is important for RNAP2 pause release (Giaimo et al., 2019; Mylonas et al., 2020) and my data suggests that TDG facilitates the eviction of H2A.Z (variants) that, together with negative elongation factors, keep

initiated RNAP2 in a non-elongating configuration. Although I see a very strong increase of H2A.Z in ESCs upon the deletion of TDG, the mechanism how the histone is actually evicted remains to be investigated. Currently I see two, non-mutually exclusive scenarios how that might happen directly. First, acetylation of the generally more unstable H2A.Z (Placek, 2005) by the TDG-interacting histone acetyltransferase p300 (Henry et al., 2016; Tini et al., 2002) reduces the DNA-binding affinity of the histone variant to render it easily displaceable by RNAP2 (Kulaeva et al., 2013; Ranjan et al., 2020; Tramantano et al., 2016). Second, the generation of active DNA demethylation-mediated SSBs, which are well correlated with H2A.Z(ac) (own data and Reid et al., 2021), cause local DNA structural rearrangement that evicts H2A.Z or potentially the entire nucleosome (Madders and Parsons, 2020; Odell et al., 2013). A third, more indirect mode of action, involves TDG in facilitating transcription at enhancers upon an according stimulus, like estrogen (Kolendowski et al., 2018), which can then affect RNAP2 pause release via NELF dissociation by enhancer RNAs (Core and Adelman, 2019; Schaukowitch et al., 2014). All these hypotheses are based on observations from own data combined with publicly available datasets and are well supported by current literature. While the first two scenarios have clear and directly testable predictions, the examination of TDG-controlled enhancer transcription and its interaction with promoters, involves more factors and therefore requires more time to investigate. One obvious experiment, however is a type of chromatin conformation capture to assess the enhancer-promoter interaction frequency in dependence of TDG.

The concept of histone eviction by active DNA demethylation was addressed earlier in the Schär laboratory. This work showed that TDG mediated DNA demethylation is associated with PARP-dependent histone H1 eviction to render loci accessible for transcription (Barekati et al., manuscript in preparation), potentially in co-operation with pioneering transcription factors (Deckard et al., 2019; Pantier et al., 2019; Sardina et al., 2018). In the case of a genome that is already very accessible, such as in mESCs cultivated in 2i-medium, H1 eviction might not be as important. The principle, however, of BER-mediated histone eviction may be applied to H2A.Z at the TSS and in the gene body where its presence is correlated with transcriptional inhibition (Coleman-Derr and Zilberman, 2012).

These considerations relate to my very initial hypothesis stating that active DNA demethylation protects a transcribed gene body from premature compaction to provide transcriptional fidelity and stability (Aranda et al., 2015; Hosogane et al., 2016), initiated by the co-transcriptional deposition of DNA methylation (Anastasiadi et al., 2018; Baubec et al., 2015). Based on the strong correlation between transcriptional activity and the occurrence of intragenic H3K36me₃, I investigated whether TDG is recruited to the gene body in dependence of the histone methyltransferase SETD2. Preliminary results indicated a weak dependence of intragenic TDG localisation on the presence of SETD2. Transcriptional analysis of DTT-induced stress response, however, showed an accelerated downregulation of DTT-responsive genes in TDG and/or SETD2-depleted mESCs after the withdrawal of DTT. This suggests that ESCs with functional

H3K36me3 deposition and/or active DNA demethylation in gene bodies, are able to maintain a transcriptional memory after the transcriptional cue is gone (Tompkins et al., 2016). It is possible that the oxidation of 5mC in gene bodies is enough to prevent PRC binding and the activity of TDG is a mere by-product. It is however also possible that accompanying SSBs equally contribute to the prevention of chromatin condensation and transcriptional hindrance. Additional phenotypes regarding TDG and transcription upon the depletion of SETD2, i.e. reduced TDG-SUMO on chromatin, demand further step-by-step investigations before precise conclusions can be drawn. An over-arching pattern, however, seems to appear that the function of active DNA demethylation exceeds the oxidation of 5mC and the excision of its derivatives to reinstate an unmodified C. In doing so, this process creates conformational changes in the chromatin and, depending on the genetic context, provides new access to transcription factors (i.e. at promoters/enhancers), polymerases (at gene bodies) or chromatin remodellers that might even mediate long-term silencing of the locus.

Regulation of SSB in active DNA demethylation and TDG-deficient mESCs

When I created the SSB data set for the analysis of the stress-response to PARP inhibition (Tal), I noticed a substantial decrease in genomic regions showing significant SSB enrichment in mESCs treated with Tal (-45%). The expectation, at that point, was that interference with SSB-repair by PARP-inhibition is bound to lead to an increased amount of SSBs, caused by continuously arising stochastic damage and DNA repair intermediates, as discussed before (Caldecott, 2008, 2014). Given that ESCs cultivated in 2i-medium are kept in a hypomethylated state (Marks and Stunnenberg, 2014), we hypothesised that PARP-inhibition could, as well, lead to a reduced (re)initiation of the DNA methylation/demethylation cycle due to the impaired repair efficiency, and/or that PARP1 modulates BER in a so-far unknown way. Addressing a potential effect of PARP inhibition on BER, we explored the possibility of TDG-modification by PARP1. This led to the discovery that TDG itself, and all of the BER proteins downstream of TDG, can be PARylated. We then found by *in vitro* reconstitution of the BER process that the overall efficiency of BER is increased by PARylation because of an accelerated dissociation of the repair proteins from the DNA - and from each other - after completion of the repair function. Hence inhibition of PARylation by Tal leads to a generalised retardation of BER already at the level of TDG-mediated base excision and AP-site incision by APE1. Notably, read densities of SSBs in unchallenged WT mESCs were generally higher over regions of active DNA demethylation than in regions without obvious targeting of TET or TDG. Also, SSBs were reduced at “scheduled” demethylation loci upon the treatment with Tal, while regions without detectable TET or TDG association (e.g. simple repeats) showed an increase in SSBs. These observations suggests that PARP-inhibition shifts the balance of SSBs originating predominantly from active DNA demethylation in unchallenged mESCs, towards more SSBs arising from spontaneous DNA damage across genome in Tal treated mESC.

Two more important observations should be addressed at this point. First, the SSB read density at sites of significant enrichment in PARP-inhibited mESCs was larger than in unchallenged mESCs but restricted to a narrower genomic region. This could suggest that at regions where SSBs are generated, PARP inhibition reduces the rate of their repair, facilitating labelling and detection. Furthermore, the SSB regions observed in PARP-inhibited mESCs are slightly shifted from those observed in unchallenged mESCs, which could reflect a blockage of the targeted DNA demethylation event (e.g. “trapped” PARP1 (Murai et al., 2012)) and potentially new generation of SSBs in the vicinity due to mistargeting of TET and/or TDG or activation of alternative pathways.

Second, mESCs depleted of TDG show an increased amount of SSB-enriched regions in their genome. While this was not the first time we observed more DNA damage in TDG null ESCs (Barekati, et al., manuscript in preparation; Alexandra Hrovat, personal communication) it is still a puzzling phenomenon, as no backup pathway is known that could process oxmCs the same way as TDG-mediated BER does. These SSB-enriched regions do still co-localise to a large extent (~50%) with those detected in WT ESCs but generally have a larger genomic footprint (i.e. peak-size +15-20%). As it is unlikely for oxmC species to generate spontaneous SSBs, this observation could point towards a “backup” pathway with a repair event larger than that associated with TDG-mediated BER. This notion is supported by another observation, namely that the vast majority of APE1-dependent damage in mESCs is dependent on TDG, suggesting that whatever alternative repair is applied in the absence of TDG does not involve APE1 or short-patch BER. Potential candidates for activity are the NEIL glycosylases or the nucleotide-excision repair pathway (NER). NEILs do not process 5fC or 5caC *in vitro* but do bind them with high affinity (Mallick et al., 2019; Schomacher et al., 2016), and their overexpression in mESCs depleted of TDG partially reduced the accumulation of 5fC and 5caC (Müller et al., 2014). NEILs can associate with PCNA, initiating long-patch BER and provide a potential explanation for an increased repair event (Dou et al., 2008).

The XPC protein, a central component of nucleotide excision repair (NER) involved in DNA damage recognition, was also shown to interact with TDG and affect DNA methylation by increasing the turnover of active DNA demethylation in the presence of TDG (Ho et al., 2017; Shimizu et al., 2010). NER engages the DNA damage as a complex with the endonucleases XPF-ERCC1 cutting 5' and XPG 3' to the lesion, evoking a larger repair event of 25-30 bp (Canturk et al., 2016). NER is however mainly processing bulky DNA lesions, to which neither 5fC nor 5caC belong and it is therefore unclear how exactly NER can associate with these substrates in the absence of TDG. Direct evidence for an activity on 5fC or 5caC is, missing but it was observed that livers of ERCC1 defective mice have reduced levels of 5fC and 5caC (Zarakowska et al., 2018), suggesting a functional interaction in some form with active DNA methylation. Also, XPG and XPF were shown to increase local DNA demethylation in cooperation with the CCCTC-binding factor CTCF, in the context of transcription (Le May et al., 2012). Notably, during the writing of this thesis, a study was

published that showed the XPC subunit RAD23B can be recruited to the BERosome by binding XRCC1 that was ubiquitylated by UHRF2 upon binding to hmC (Liu et al., 2021) and potentially also after binding to 5fC and 5caC (Schneider et al., 2020). These observations hint at a possible role of NER or NER factors in active DNA demethylation as an alternative to the TDG-BER dependent process. Given that alternative pathways may act on 5fCs and 5caCs, it can be reasoned that the accumulation of 5fC or 5caC in cells depleted of TDG, does not reflect the total of all 5fC and 5caCs generated by TETs. This could be determined by measurement of these species in a combined depletion of TDG and above-mentioned candidate proteins. Further hints as to which repair factors might take part in the alternative processing of oxmCs could soon be revealed by the CRISPR/CAS9 screen for synthetic lethality in the absence of APE1, currently performed in the Schär laboratory by Alexandra Hrovat.

All in all, the research presented in my PhD thesis extends the insight into the biological role of TDG/BER-mediated active DNA demethylation and the molecular mechanisms involved in transcriptional control. I could show that active DNA demethylation-associated BER, affects transcription at regulatory regions including enhancers and TSSs, but also at repetitive elements and within the gene body. The knowledge gained will serve as a basis for future research into the underlying mechanisms for which a number of novel and testable hypotheses have been proposed and discussed.

At all loci examined, TET protein and BER activity create a dynamic DNA methylation state in a constant cross-talk with histone variants and their modifications. While my work focused on the impact of active DNA demethylation on histone H2A.Z and the trimethylation of H3K36, many others are investigating the interactions of DNA (de)methylation with modulators of H3K4 or H3K27 modification to further explore the intricate way of transcriptional regulation mediated by the epigenetic machinery. An important observation from my work but also the work of others is that the functional interactions between DNA (de)methylation, histone modifications, alteration of chromatin states as well as transcriptional initiation and elongation, cannot be seen as a linear, hierarchical process where one comes before the other (Sandoval and Reich, 2021). But, as generic as it may sound, everything seems to impact everything and every consequence will serve as a new biological cause. And yet, underlying this complex network of molecular biochemical processes are simple chemical reactions. Processes like the oxidation of DNA bases, the repair, sensing and signalling of the resulting DNA interruption, the modification of responses in chromatin dynamics and transcription - some of which are newly implicated in my PhD thesis - all are subject to fundamental physical rules. Epigenetic modifications thus tip the balance in the chemical equilibria of these protein-DNA interactions that are otherwise governed by the laws of thermodynamics and ultimately, entropy.

Finally, studying of the thymine DNA glycosylase beautifully exemplifies how evolution happens. Initially described, TDG serves as DNA repair protein to correct G•T and G•U DNA mispairings (Cortázar et al., 2007).

The emergence of methylated bases and their oxidation presented a novel type of “mismatch” and the properties of TDG were hijacked to establish a system for chromatin and gene regulation through a multi-step pathway involving the generation of DNA strand breaks (Liu et al., 2020; Mulholland et al., 2020a; Parker et al., 2019). The concept of changing chromatin modifications through “DNA damage” creates a challenge for genome stability but at the same time an opportunity to introduce disruptive changes into chromatin. This exemplifies how evolution is not directed to develop the most efficient or direct mechanisms, but “utilises” pre-existing tools that have proven reliable and brings forth novel adaptations based on possibilities and given circumstances (Jeffery, 1999; Koliadenko and Wilanowski, 2020).

The development of epigenetic regulation also contributes to evolution. Conrad Waddington defined the epigenetic information as heritable, and it is without doubt maintained through (most) cell divisions. There is, however, a discussion whether epigenetic marks are trans-generationally inheritable, which is in my view obsolete. That is, as long as the machinery that conveys epigenetic marks, e.g. DNA (de-)methylation proteins, is trans-generationally inherited, it is also able to maintain the epigenetic regulation. This means, even if epigenetic patterns acquired through adaptation, are being erased during the formation of offspring, a consistent external (selective) pressure will lead to the same transcriptional response and the re-establishment of the same epigenetic pattern. Is this selective pressure present for multiple generations of an organism (for example in humans more than three), the consequential prolonged “monotonous” expression pattern will be reflected by a corresponding monotonous CpG methylation pattern. It then stands the chance to become “fixed” in the genome, of which a hint was shown in ageing cells (Hernando-Herraez et al., 2019), and is clearly represented by the selection and creation of CpG-islands (Illingworth and Bird, 2009; Sharif et al., 2010; Weber et al., 2007). We therefore can acknowledge the principle of evolution proposed by Charles Darwin, which applies natural selection, combined with Jean-Baptiste Lamarck’s principle that organisms constantly adapt to their surroundings as good as they can – these principles are not mutually exclusive but complementary.

6 References

- Abd Elmageed, Z.Y., Naura, A.S., Errami, Y., and Zerfaoui, M. (2012). The Poly(ADP-ribose) polymerases (PARPs): New roles in intracellular transport. *Cell. Signal.* *24*, 1–8.
- Abed, J.A., and Jones, R.S. (2012). H3K36me3 key to Polycomb-mediated gene silencing in lineage specification. *Nat. Struct. Mol. Biol.* *19*, 1214–1215.
- Agius, F., Kapoor, A., and Zhu, J.-K. (2006). Role of the Arabidopsis DNA glycosylase/lyase ROS1 in active DNA demethylation. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 11796–11801.
- Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W.S., D’Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., et al. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* *18*.
- Anastasiadi, D., Esteve-Codina, A., and Piferrer, F. (2018). Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* *11*, 37.
- Aranda, S., Mas, G., and Di Croce, L. (2015). Regulation of gene transcription by Polycomb proteins. *Sci. Adv.* *1*, e1500737.
- Armache, A., Yang, S., Martínez de Paz, A., Robbins, L.E., Durmaz, C., Cheong, J.Q., Ravishankar, A., Daman, A.W., Ahimovic, D.J., Klevorn, T., et al. (2020). Histone H3.3 phosphorylation amplifies stimulation-induced transcription. *Nature* 1–6.
- Arnold, P.R., Wells, A.D., and Li, X.C. (2020). Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Front. Cell Dev. Biol.* *7*, 377.
- Atlasi, Y., and Stunnenberg, H.G. (2017). The interplay of epigenetic marks during stem cell differentiation and development. *Nat. Rev. Genet.* *18*, 643–658.
- Atlasi, Y., Megchelenbrink, W., Peng, T., Habibi, E., Joshi, O., Wang, S.-Y., Wang, C., Logie, C., Poser, I., Marks, H., et al. (2019). Epigenetic modulation of a hardwired 3D chromatin landscape in two naive states of pluripotency. *Nat. Cell Biol.* *21*, 568–578.
- Bachman, M., Uribe-Lewis, S., Yang, X., Burgess, H.E., Iurlaro, M., Reik, W., Murrell, A., and Balasubramanian, S. (2015). 5-Formylcytosine can be a stable DNA modification in mammals. *Nat. Chem. Biol.* *11*, 555–557.
- Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schübeler, D.

(2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* 520, 243–247.

Bayraktar, G., and Kreutz, M.R. (2017). Neuronal DNA Methyltransferases: Epigenetic Mediators between Synaptic Activity and Gene Expression? *Neurosci.* 107385841770745.

Bayraktar, G., and Kreutz, M.R. (2018). The Role of Activity-Dependent DNA Demethylation in the Adult Brain and in Neurological Disorders. *Front. Mol. Neurosci.* 11, 169.

Bibel, M., Richter, J., Lacroix, E., and Barde, Y.-A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat. Protoc.* 2, 1034–1043.

Bird, A.P. (1984a). DNA methylation versus gene expression. *J. Embryol. Exp. Morphol.* 83 Suppl, 31–40.

Bird, A.P. (1984b). Gene expression: DNA methylation — how important in gene control? *Nature* 307, 503–504.

Birzele, F., Csaba, G., and Zimmer, R. (2008). Alternative splicing and protein structure evolution. *Nucleic Acids Res.* 36, 550–558.

Blanco, E., González-Ramírez, M., Alcaine-Colet, A., Aranda, S., and Di Croce, L. (2020). The Bivalent Genome: Characterization, Structure, and Regulation. *Trends Genet.* 36, 118–131.

Boland, M.J., and Christman, J.K. (2008). Characterization of Dnmt3b:Thymine-DNA Glycosylase Interaction and Stimulation of Thymine Glycosylase-Mediated Repair by DNA Methyltransferase(s) and RNA. *J. Mol. Biol.* 379, 492–504.

Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., et al. (2018). Ten things you should know about transposable elements. *Genome Biol.* 19, 199.

Brooks, S.C., Fischer, R.L., Huh, J.H., and Eichman, B.F. (2014). 5-methylcytosine recognition by arabidopsis thaliana DNA glycosylases DEMETER and DML3. *Biochemistry* 53, 2525–2532.

Brunelle, M., Nordell Markovits, A., Rodrigue, S., Lupien, M., Jacques, P.É., and Gévry, N. (2015). The histone variant H2A.Z is an important regulator of enhancer activity. *Nucleic Acids Res.* 43, 9742–9756.

Burgess, R.J., and Zhang, Z. (2013). Histone chaperones in nucleosome assembly and human disease. *Nat. Struct. Mol. Biol.* 20, 14–22.

Cai, L., Rothbart, S.B., Lu, R., Xu, B., Chen, W.Y., Tripathy, A., Rockowitz, S., Zheng, D., Patel, D.J., Allis, C.D., et al. (2013). An H3K36 Methylation-Engaging Tudor Motif of Polycomb-like Proteins Mediates PRC2

Complex Targeting. *Mol. Cell* 49, 571–582.

Caldecott, K.W. (2008). Single-strand break repair and genetic disease. *Nat. Rev. Genet.* 9, 619–631.

Caldecott, K.W. (2014). DNA single-strand break repair. *Exp. Cell Res.* 329, 2–8.

Calvo, O. (2018). Sub1 and RNAPII, until termination does them part. *Transcription* 9, 52–60.

Canturk, F., Karaman, M., Selby, C.P., Kemp, M.G., Kulaksiz-Erkmen, G., Hu, J., Li, W., Lindsey-Boltz, L.A., and Sancar, A. (2016). Nucleotide excision repair by dual incisions in plants. *Proc. Natl. Acad. Sci.* 113, 4706–4710.

Carlberg, C., and Molnár, F. (2018). The Histone Code BT - Human Epigenomics. C. Carlberg, and F. Molnár, eds. (Singapore: Springer Singapore), pp. 75–88.

Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution - Trashing old rules to forge new ones. *Cell* 157, 77–94.

Chen, Z., and Riggs, A.D. (2011). DNA Methylation and Demethylation in Mammals. *J. Biol. Chem.* 286, 18347–18353.

Chevray, P.M., and Nathans, D. (1992). Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5789–5793.

Chiou, S.-H., Jiang, B.-H., Yu, Y.-L., Chou, S.-J., Tsai, P.-H., Chang, W.-C., Chen, L.-K., Chen, L.-H., Chien, Y., and Chiou, G.-Y. (2013). Poly(ADP-ribose) polymerase 1 regulates nuclear reprogramming and promotes iPSC generation without c-Myc. *J. Exp. Med.* 210, 85–98.

Clapier, C.R., Iwasa, J., Cairns, B.R., and Peterson, C.L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nat. Rev. Mol. Cell Biol.* 18, 407–422.

Coleman-Derr, D., and Zilberman, D. (2012). Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS Genet.* 8, e1002988–e1002988.

Cooper, S.J. (2005). Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* 16, 1–10.

Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: A nexus of gene regulation. *Genes Dev.* 33, 960–982.

Cortázar, D., Kunz, C., Saito, Y., Steinacher, R., and Schär, P. (2007). The enigmatic thymine DNA glycosylase. *DNA Repair (Amst).* 6, 489–504.

Cortázar, 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.

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.

Dawlaty, M.M., Breiling, A., Le, T., Barrasa, M.I., Raddatz, G., Gao, Q., Powell, B.E., Cheng, A.W., Faull, K.F., Lyko, F., et al. (2014). Loss of tet enzymes compromises proper differentiation of embryonic stem cells. *Dev. Cell* *29*, 102–111.

Deckard, C.E., Banerjee, D.R., and Sczepanski, J.T. (2019). Chromatin Structure and the Pioneering Transcription Factor FOXA1 Regulate TDG-Mediated Removal of 5-Formylcytosine from DNA. *J. Am. Chem. Soc.* *141*, 14110–14114.

Deniz, Ö., de la Rica, L., Cheng, K.C.L., Spensberger, D., and Branco, M.R. (2018). SETDB1 prevents TET2-dependent activation of IAP retroelements in naïve embryonic stem cells. *Genome Biol.* *19*, 6.

Deniz, Ö., Frost, J.M., and Branco, M.R. (2019). Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* *20*, 417–431.

Dickens, E., and Ahmed, S. (2018). Principles of cancer treatment by chemotherapy. *Surg.* *36*, 134–138.

Dodd, T., Yan, C., Kossmann, B.R., Martin, K., and Ivanov, I. (2018). Uncovering universal rules governing the selectivity of the archetypal DNA glycosylase TDG. *Proc. Natl. Acad. Sci. U. S. A.* 201803323.

Dou, H., Theriot, C.A., Das, A., Hegde, M.L., Matsumoto, Y., Boldogh, I., Hazra, T.K., Bhakat, K.K., and Mitra, S. (2008). Interaction of the Human DNA Glycosylase NEIL1 with Proliferating Cell Nuclear Antigen: THE POTENTIAL FOR REPLICATION-ASSOCIATED REPAIR OF OXIDIZED BASES IN MAMMALIAN GENOMES*. *J. Biol. Chem.* *283*, 3130–3140.

Eggenberger, J., Blanco-Melo, D., Panis, M., Brennand, K.J., and TenOever, B.R. (2019). Type I interferon response impairs differentiation potential of pluripotent stem cells. *Proc. Natl. Acad. Sci.* *116*, 1384–1393.

Elsässer, S.J., Noh, K.-M., Diaz, N., Allis, C.D., and Banaszynski, L.A. (2015). Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* *522*, 240–244.

Enserink, J.M. (2015). Sumo and the cellular stress response. *Cell Div.* *10*, 4.

Ettig, R., Kepper, N., Stehr, R., Wedemann, G., and Rippe, K. (2011). Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of DNA unwrapping. *Biophys. J.* *101*, 1999–2008.

Fenley, A.T., Adams, D.A., and Onufriev, A. V (2010). Charge state of the globular histone core controls stability of the nucleosome. *Biophys. J.* *99*, 1577–1585.

Field, A., and Adelman, K. (2020). Evaluating Enhancer Function and Transcription. *Annu. Rev. Biochem.* *89*, 213–234.

Fischer, M. (2017). Census and evaluation of p53 target genes. *Nat. Publ. Gr.* *36*, 3943–3956.

Fisher, A.E.O., Hohegger, H., Takeda, S., and Caldecott, K.W. (2007). Poly(ADP-Ribose) Polymerase 1 Accelerates Single-Strand Break Repair in Concert with Poly(ADP-Ribose) Glycohydrolase. *Mol. Cell. Biol.* *27*, 5597–5605.

Friedman, J.I., and Stivers, J.T. (2010). Detection of Damaged DNA Bases by DNA Glycosylase Enzymes. *Biochemistry* *49*, 4957–4967.

Gao, Y., Li, L., Yuan, P., Zhai, F., Ren, Y., Yan, L., Li, R., Lian, Y., Zhu, X., Wu, X., et al. (2020). 5-Formylcytosine landscapes of human preimplantation embryos at single-cell resolution. *PLOS Biol.* *18*, e3000799.

Giaimo, B.D., Ferrante, F., Herchenröther, A., Hake, S.B., and Borggrefe, T. (2019). The histone variant H2A.Z in gene regulation. *Epigenetics and Chromatin* *12*, 1–22.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. *Mol. Cell* *45*, 814–825.

Ginno, P.A., Gaidatzis, D., Feldmann, A., Hoerner, L., Imanci, D., Burger, L., Zilbermann, F., Peters, A.H.F.M., Edenhofer, F., Smallwood, S.A., et al. (2020). A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nat. Commun.* *11*, 2680.

Goodier, J.L. (2016). Restricting retrotransposons: a review. *Mob. DNA* *7*, 16.

Grinda, T., and Delaloge, S. (2020). Survival benefits of PARP inhibitors in advanced breast cancer: a mirage? *Ann. Oncol.* *31*, 1432–1434.

Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* *19*, 621–637.

Hahn, M.A., Wu, X., Li, A.X., Hahn, T., and Pfeifer, G.P. (2011). Relationship between gene body DNA methylation and intragenic H3K9ME3 and H3K36ME3 chromatin marks. *PLoS One* *6*.

Hanzlikova, H., Gittens, W., Krejcikova, K., Zeng, Z., and Caldecott, K.W. (2016). Overlapping roles for PARP1 and PARP2 in the recruitment of endogenous XRCC1 and PNKP into oxidized chromatin. *Nucleic Acids Res.* *45*, gkw1246.

- Hardeland, U., Steinacher, R., Jiricny, J., and Schär, P. (2002). Modification of the human thymine-DNA glycosylase by ubiquitin-like proteins facilitates enzymatic turnover. *EMBO J.* *21*, 1456–1464.
- Hassan, H.M., Kolendowski, B., Iovic, M., Bose, K., Dranse, H.J., Sampaio, A. V., Underhill, T.M., and Torchia, J. (2017). Regulation of Active DNA Demethylation through RAR-Mediated Recruitment of a TET/TDG Complex. *Cell Rep.* *19*, 1685–1697.
- Hassan, H.M., Iovic, M., Kolendowski, B., Bauer-Maison, N., Onabote, O., Cecchini, M., Haig, A., Maleki Vareki, S., Underhill, T.M., and Torchia, J. (2020). Loss of Thymine DNA Glycosylase Causes Dysregulation of Bile Acid Homeostasis and Hepatocellular Carcinoma. *Cell Rep.* *31*.
- Haubold, B., and Wiehe, T. (2006). How repetitive are genomes? *BMC Bioinformatics* *7*, 541.
- He, Y., Li, B., Li, Z., Liu, P., Wang, Y., and Tang, Q. (2011). Tet-Mediated Formation of 5-Carboxyl cytosine and Its Excision by TDG in Mammalian DNA. *Sci. (New York, N.Y.)* *333*, 1303–1307.
- Hendriks, I.A., Lyon, D., Young, C., Jensen, L.J., Vertegaal, A.C.O., and Nielsen, M.L. (2017). Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. *Nat. Struct. Mol. Biol.* *24*, 325–336.
- Hendriks, I.A., Larsen, S.C., and Nielsen, M.L. (2019). An advanced strategy for comprehensive profiling of ADP-ribosylation sites using mass spectrometry-based proteomics. *Mol. Cell. Proteomics* *18*, 1010–1024.
- Henry, R.A., Mancuso, P., Kuo, Y.-M., Tricarico, R., Tini, M., Cole, P.A., Bellacosa, A., and Andrews, A.J. (2016). Interaction with the DNA Repair Protein Thymine DNA Glycosylase Regulates Histone Acetylation by p300.
- Hernando-Herraez, I., Evano, B., Stubbs, T., Commere, P.H., Jan Bonder, M., Clark, S., Andrews, S., Tajbakhsh, S., and Reik, W. (2019). Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. *Nat. Commun.* *10*, 1–11.
- Ho, J.J., Cattoglio, C., McSwiggen, D.T., Tjian, R., and Fong, Y.W. (2017). Regulation of DNA demethylation by the XPC DNA repair complex in somatic and pluripotent stem cells. *Genes Dev.* *31*, 830–844.
- Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* *325*, 626–628.
- Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., and Bradbury, E.M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 “tail” to DNA. *J. Biol. Chem.* *268*, 305–314.
- Hosogane, M., Funayama, R., Shirota, M., and Nakayama, K. (2016). Lack of Transcription Triggers

H3K27me3 Accumulation in the Gene Body. *Cell Rep.* 16, 696–706.

Hu, X., Zhang, L., Mao, S.-Q., Li, Z., Chen, J., Zhang, R.-R., Wu, H.-P., Gao, J., Guo, F., Liu, W., et al. (2014). Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* 14, 512–522.

Huang, C., and Zhu, B. (2018). Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. *Biophys. Reports* 4, 170–177.

Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* 49, 324.

Iberg-Badeaux, A., Collombet, S., Laurent, B., van Oevelen, C., Chin, K.-K., Thieffry, D., Graf, T., and Shi, Y. (2017). A Transcription Factor Pulse Can Prime Chromatin for Heritable Transcriptional Memory. *Mol. Cell Biol.* 37.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. *FEBS Lett.* 583, 1713–1720.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303.

Iurlaro, M., Ficiz, G., Oxley, D., Raiber, E.-A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S., and Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* 14, R119.

Jabbari, K., and Bernardi, G. (2004). Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. *Gene* 333, 143–149.

Jacobs, A.L., and Schär, P. (2012). DNA glycosylases: In DNA repair and beyond. *Chromosoma* 121, 1–20.

Jeffery, C.J. (1999). Moonlighting proteins. *Trends Biochem. Sci.* 24, 8–11.

Jeziorska, D.M., Murray, R.J.S., Gobbi, M. De, Gaentzsch, R., Garrick, D., Ayyub, H., Chen, T., Li, E., Telenius, J., Lynch, M., et al. (2017). DNA methylation of intragenic CpG islands depends on their transcriptional activity during differentiation and disease. *Proc. Natl. Acad. Sci.* 201703087.

Jin, Z., and Liu, Y. (2018). DNA methylation in human diseases. *Genes Dis.* 5, 1–8.

Jin, S.-G., Guo, C., and Pfeifer, G.P. (2008). GADD45A Does Not Promote DNA Demethylation. *PLoS Genet.* 4, e1000013.

Jjingo, D., Conley, A.B., Yi, S. V., Lunyak, V. V., and King Jordan, I. (2012). On the presence and role of human gene-body DNA methylation. *Oncotarget* 3, 462–474.

- Johnson, T.B., and Coghill, R.D. (1925). The Discovery of 5-Methyl-cytosine in Tuberculinic Acid, the Nucleic Acid of the Tubercle Bacillus. *J. Am. Chem. Soc.* *47*, 2838–2844.
- Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* *16*, 167–177.
- Jungmichel, S., Rosenthal, F., Altmeyer, M., Lukas, J., Hottiger, M.O., and Nielsen, M.L. (2013). Proteome-wide Identification of Poly(ADP-Ribosyl)ation Targets in Different Genotoxic Stress Responses. *Mol. Cell* *52*, 272–285.
- Jurka, J., Kapitonov, V. V, Kohany, O., and Jurka, M. V (2007). Repetitive Sequences in Complex Genomes: Structure and Evolution. *Annu. Rev. Genomics Hum. Genet.* *8*, 241–259.
- Kalai, M., Van Loo, G., Vanden Berghe, T., Meeus, A., Burm, W., Saelens, X., and Vandenabeele, P. (2002). Tipping the balance between necrosis and apoptosis in human and murine cells treated with interferon and dsRNA. *Cell Death Differ.* *9*, 981–994.
- Kellner, W.A., Bell, J.S.K., and Vertino, P.M. (2015). GC skew defines distinct RNA polymerase pause sites in CpG island promoters. *Genome Res.* *25*, 1600–1609.
- Kheir, T.B., and Lund, A.H. (2010). Epigenetic dynamics across the cell cycle. *Essays Biochem.* *48*, 107–120.
- Kim, E.J., and Um, S.J. (2008). Thymine-DNA glycosylase interacts with and functions as a coactivator of p53 family proteins. *Biochem. Biophys. Res. Commun.* *377*, 838–842.
- Kim, Y.-J., and M. Wilson III, D. (2012). Overview of Base Excision Repair Biochemistry. *Curr. Mol. Pharmacol.* *5*, 3–13.
- Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* *502*, 472–479.
- Kolendowski, B., Hassan, H., Krstic, M., Isovich, M., Thillainadesan, G., Chambers, A.F., Tuck, A.B., and Torchia, J. (2018). Genome-wide analysis reveals a role for TDG in estrogen receptor-mediated enhancer RNA transcription and 3-dimensional reorganization. *Epigenetics and Chromatin* *11*, 5.
- Koliadenko, V., and Wilanowski, T. (2020). Additional functions of selected proteins involved in DNA repair. *Free Radic. Biol. Med.* *146*, 1–15.
- Kulaeva, O.I., Hsieh, F.K., Chang, H.W., Luse, D.S., and Studitsky, V.M. (2013). Mechanism of transcription through a nucleosome by RNA polymerase II. *Biochim. Biophys. Acta - Gene Regul. Mech.* *1829*, 76–83.
- Kulis, M., Queirós, A.C., Beekman, R., and Martín-Subero, J.I. (2013). Intragenic DNA methylation in

transcriptional regulation, normal differentiation and cancer. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829**, 1161–1174.

Kunz, C., Focke, F., Saito, Y., Schuermann, D., Lettieri, T., Selfridge, J., and Schär, P. (2009). Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil. *PLoS Biol.* **7**, e91.

de la Rica, L., Deniz, Ö., Cheng, K.C.L., Todd, C.D., Cruz, C., Houseley, J., and Branco, M.R. (2016). TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biol.* **17**.

Larsson, A.J.M., Johnsson, P., Hagemann-jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, Å., Rivera, C.M., Ren, B., and Sandberg, R. (2018). Genomic encoding of transcriptional burst kinetics. *Nature*.

Léger, H., Smet-Nocca, C., Attmane-Elakeb, A., Morley-Fletcher, S., Benecke, A.G., and Eilebrecht, S. (2014). A TDG/CBP/RAR α ternary complex mediates the retinoic acid-dependent expression of DNA methylation-sensitive genes. *Genomics, Proteomics Bioinforma.* **12**, 8–18.

Lev Maor, G., Yearim, A., and Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. *Trends Genet.* **31**, 274–280.

Li, H., Rauch, T., Chen, Z.-X., Szabó, P.E., Riggs, A.D., and Pfeifer, G.P. (2006). The Histone Methyltransferase SETDB1 and the DNA Methyltransferase DNMT3A Interact Directly and Localize to Promoters Silenced in Cancer Cells. *J. Biol. Chem.* **281**, 19489–19500.

Li, M., He, Y., Dubois, W., Wu, X., Shi, J., and Huang, J. (2012). Distinct Regulatory Mechanisms and Functions for p53-Activated and p53-Repressed DNA Damage Response Genes in Embryonic Stem Cells. *Mol. Cell* **46**, 30–42.

Li, Y.Q., Zhou, P.Z., Zheng, X.D., Walsh, C.P., and Xu, G.L. (2007). Association of Dnmt3a and thymine DNA glycosylase links DNA methylation with base-excision repair. *Nucleic Acids Res.* **35**, 390–400.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715.

Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610–3618.

Liu, J., Hu, H., Panserat, S., and Marandel, L. (2020). Evolutionary history of DNA methylation related genes in chordates: new insights from multiple whole genome duplications. *Sci. Rep.* **10**, 1–14.

Liu, X., Xu, B., Yang, J., He, L., Zhang, Z., Cheng, X., Yu, H., Liu, X., Jin, T., Peng, Y., et al. (2021). UHRF2 commissions the completion of DNA demethylation through allosteric activation by 5hmC and K33-linked ubiquitination of XRCC1. *Mol. Cell* 1–15.

- London, R.E. (2015). The structural basis of XRCC1-mediated DNA repair. *DNA Repair (Amst)*. 30, 90–103.
- López-Moyado, I.F., Tsagaratou, A., Yuita, H., Seo, H., Delatte, B., Heinz, S., Benner, C., and Rao, A. (2019). Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. *Proc. Natl. Acad. Sci.* 201903059.
- Lu, J.Y., Shao, W., Chang, L., Yin, Y., Li, T., Zhang, H., Hong, Y., Percharde, M., Guo, L., Wu, Z., et al. (2020). Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Rep.* 30, 3296–3311.e5.
- Lyko, F. (2018). The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* 19, 81–92.
- Machida, S., Takizawa, Y., Ishimaru, M., Sugita, Y., Sekine, S., Nakayama, J., Wolf, M., and Kurumizaka, H. (2018). Structural Basis of Heterochromatin Formation by Human HP1. *Mol. Cell* 69, 385–397.e8.
- Madders, E.C.E.T., and Parsons, J.L. (2020). Base Excision Repair in Chromatin and the Requirement for Chromatin Remodelling BT - Mechanisms of Genome Protection and Repair. D.O. Zharkov, ed. (Cham: Springer International Publishing), pp. 59–75.
- 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.
- Makałowski, W., Gotea, V., Pande, A., and Makałowska, I. (2019). Transposable Elements: Classification, Identification, and Their Use As a Tool For Comparative Genomics. pp. 177–207.
- Mallick, M., Han, D., Schomacher, L., Schu, K.M., Musheev, M.U., Karaulanov, E., Krebs, L., and Seggern, A. Von (2019). NEIL1 and NEIL2 DNA glycosylases protect neural crest development against mitochondrial oxidative stress. 1–38.
- Marina, R.J., Sturgill, D., Bailly, M.A., Thenoz, M., Varma, G., Prigge, M.F., Nanan, K.K., Shukla, S., Haque –, N., and Oberdoerffer, S. (2015). TET-catalyzed oxidation of intragenic 5-methylcytosine regulates CTCF-dependent alternative splicing. *EMBO J.* 35, 1–21.
- Marks, H., and Stunnenberg, H.G. (2014). Transcription regulation and chromatin structure in the pluripotent ground state. *Biochim. Biophys. Acta* 1839, 129–137.
- Marks, H., Kalkan, T., Menafrá, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Francis Stewart, A., Smith, A., et al. (2012). The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell* 149, 590–604.

Martinelli, G., Mancini, M., De Benedittis, C., Rondoni, M., Papayannidis, C., Manfrini, M., Meggendorfer, M., Calogero, R., Guadagnuolo, V., Fontana, M.C., et al. (2018). SETD2 and histone H3 lysine 36 methylation deficiency in advanced systemic mastocytosis. *Leukemia* 32, 139–148.

Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* 23, 1256–1269.

Le May, N., Fradin, D., Iltis, I., Bougnères, P., and Egly, J.-M. (2012). XPG and XPF Endonucleases Trigger Chromatin Looping and DNA Demethylation for Accurate Expression of Activated Genes. *Mol. Cell* 47, 622–632.

McDaniel, S.L., and Strahl, B.D. (2017). Shaping the cellular landscape with Set2/SETD2 methylation. *Cell. Mol. Life Sci.* 74, 1–18.

Métivier, R., Gallais, R., Tiffoche, C., Le Péron, 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.

Minoguchi, S., and Iba, H. (2008). Instability of Retroviral DNA Methylation in Embryonic Stem Cells. *Stem Cells* 26, 1166–1173.

Morgan, D.O. (2007). *The cell cycle: principles of control.* (London: New Science Press), p.

Mulholland, C.B., Nishiyama, A., Ryan, J., Nakamura, R., Yiğit, M., Glück, I.M., Trummer, C., Qin, W., Bartoschek, M.D., Traube, F.R., et al. (2020a). Recent evolution of a TET-controlled and DPPA3/STELLA-driven pathway of passive DNA demethylation in mammals. *Nat. Commun.* 11.

Mulholland, C.B., Traube, F.R., Ugur, E., Parsa, E., Eckl, E.-M., Schönung, M., Modic, M., Bartoschek, M.D., Stolz, P., Ryan, J., et al. (2020b). Distinct and stage-specific contributions of TET1 and TET2 to stepwise cytosine oxidation in the transition from naive to primed pluripotency. *Sci. Rep.* 10, 12066.

Müller, U., Bauer, C., Siegl, M., Rottach, A., and Leonhardt, H. (2014). TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.* 42, 8592–8604.

Murai, J., Huang, S.N., Das, B.B., Renaud, A., Zhang, Y., Doroshow, J.H., Ji, J., Takeda, S., and Pommier, Y. (2012). Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. *Cancer Res.* 72, 5588 LP – 5599.

Musselman, C.A., Avvakumov, N., Watanabe, R., Abraham, C.G., Lalonde, M.-E., Hong, Z., Allen, C., Roy, S., Nuñez, J.K., Nickoloff, J., et al. (2012). Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. *Nat. Struct. Mol. Biol.* 19, 1266–1272.

- Mylonas, C., Auld, A.L., Lee, C., Cisse, I.I., and Boyer, L.A. (2020). A dual role for H2A.Z.1 in modulating the dynamics of RNA Polymerase II initiation and elongation. *BioRxiv* 2020.06.22.165373.
- Nakamura, J., and Swenberg, J.A. (1999). Endogenous apurinic/aprimidinic sites in genomic DNA of mammalian tissues. *Cancer Res.* *59*, 2522–2526.
- Nashun, B., Hill, P.W.S., Smallwood, S.A., Festenstein, R.J., Kelsey, G., Nashun, B., Hill, P.W.S., Smallwood, S.A., Dharmalingam, G., and Amouroux, R. (2015). Continuous Histone Replacement by Hira Is Essential for Normal Transcriptional Regulation and De Novo DNA Methylation during Mouse Oogenesis Article Continuous Histone Replacement by Hira Is Essential for Normal Transcriptional Regulation and De Novo DNA . *MOLCEL* *60*, 611–625.
- Navarro, C., Lyu, J., Katsori, A.M., Caridha, R., and Elsässer, S.J. (2020). An embryonic stem cell-specific heterochromatin state promotes core histone exchange in the absence of DNA accessibility. *Nat. Commun.* *11*, 1–14.
- Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Anselmi, F., Parlato, C., Medana, C., Dal Bello, F., and Oliviero, S. (2015). Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Rep.* *10*, 674–683.
- Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., Maldotti, M., Anselmi, F., and Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*.
- Noble, D. (2015). Conrad Waddington and the origin of epigenetics. *J. Exp. Biol.* *218*, 816–818.
- Odell, I.D., Wallace, S.S., and Pederson, D.S. (2013). Rules of engagement for base excision repair in chromatin. *J. Cell. Physiol.* *228*, 258–266.
- Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev.* *10*, 2657–2683.
- Osowski, C.M., and Urano, F. (2011). Measuring ER Stress and the Unfolded Protein Response Using Mammalian Tissue Culture System. *Methods Enzymol.* *490*, 71–92.
- Pal, S., Gupta, R., Kim, H., Wickramasinghe, P., Baubet, V., Showe, L.C., Dahmane, N., and Davuluri, R. V (2011). Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. *Genome Res.* *21*, 1260–1272.
- Pantier, R., Tatar, T., Colby, D., and Chambers, I. (2019). Endogenous epitope-tagging of Tet1, Tet2 and Tet3 identifies TET2 as a naïve pluripotency marker. *Life Sci. Alliance* *2*, e201900516.
- Parker, M.J., Weigele, P.R., and Saleh, L. (2019). Insights into the Biochemistry, Evolution, and

Biotechnological Applications of the Ten-Eleven Translocation (TET) Enzymes. *Biochemistry* 58, 450–467.

Parry, A., Rulands, S., and Reik, W. (2020). Active turnover of DNA methylation during cell fate decisions. *Nat. Rev. Genet.*

Placek, B.J. (2005). The H2A.Z/H2B dimer is unstable compared to the dimer containing the major H2A isoform. *Protein Sci.* 14, 514–522.

Porrau, O., Boudvillain, M., and Libri, D. (2016). Transcription Termination: Variations on Common Themes.

Raiber, E.-A., Beraldi, D., Ficz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.* 13, R69.

Rajagopalan, K.N., Chen, X., Weinberg, D.N., Chen, H., Majewski, J., Allis, C.D., and Lu, C. (2021). Depletion of H3K36me2 recapitulates epigenomic and phenotypic changes induced by the H3.3K36M oncohistone mutation. *Proc. Natl. Acad. Sci.* 118, e2021795118.

Ranjan, A., Nguyen, V.Q., Liu, S., Wisniewski, J., Kim, J.M., Tang, X., Mizuguchi, G., Elalaoui, E., Nickels, T.J., Jou, V., et al. (2020). Live-cell single particle imaging reveals the role of RNA polymerase II in histone H2A.Z eviction. *Elife* 9.

Ray Chaudhuri, A., and Nussenzweig, A. (2017). The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nat. Rev. Mol. Cell Biol.* 18, 610–621.

Razin, A., and Cedar, H. (1991). DNA methylation and gene expression. *Microbiol.Rev.* 55, 451–458.

Reid, D.A., Reed, P.J., Schlachetzki, J.C.M., Nitulescu, I.I., Chou, G., Tsui, E.C., Jones, J.R., Chandran, S., Lu, A.T., McClain, C.A., et al. (2021). Incorporation of a nucleoside analog maps genome repair sites in postmitotic human neurons. *Science* (80-). 372, 91–94.

Rulands, S., Lee, H.J., Clark, S.J., Angermueller, C., Smallwood, S.A., Krueger, F., Mohammed, H., Dean, W., Nichols, J., Rugg-Gunn, P., et al. (2018). Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency. *Cell Syst.* 7, 63-76.e12.

Sandoval, J.E., and Reich, N.O. (2021). p53 and TDG are dominant in regulating the activity of the human de novo DNA methyltransferase DNMT3A on nucleosomes. *J. Biol. Chem.* 296, 100058.

Sardina, J.L., Collombet, S., Tian, T. V, Gómez, A., Di Stefano, B., Berenguer, C., Brumbaugh, J., Stadhouders, R., Segura-Morales, C., Gut, M., et al. (2018). Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. *Cell Stem Cell* 23, 727-741.e9.

- Schaukowitch, K., Joo, J.-Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.-K. (2014). Enhancer RNA Facilitates NELF Release from Immediate Early Genes. *Mol. Cell* 56, 29–42.
- Schlesinger, S., Kaffe, B., Melcer, S., Aguilera, J.D., Sivaraman, D.M., Kaplan, T., and Meshorer, E. (2017). A hyperdynamic H3.3 nucleosome marks promoter regions in pluripotent embryonic stem cells. *Nucleic Acids Res.* 45, 12181–12194.
- Schneider, M., Trummer, C., Stengl, A., Zhang, P., Szwagierczak, A., Cardoso, M.C., Leonhardt, H., Bauer, C., and Antes, I. (2020). Systematic analysis of the binding behaviour of UHRF1 towards different methyl- and carboxylcytosine modification patterns at CpG dyads. *PLoS One* 15, e0229144.
- Schomacher, L., Han, D., Musheev, M.U., Arab, K., Kienhöfer, S., von Seggern, A., and Niehrs, C. (2016). Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nat. Struct. Mol. Biol.* 23, 116–124.
- Schuermann, D., Weber, A.R., and Schär, P. (2016). Active DNA demethylation by DNA repair: Facts and uncertainties. *DNA Repair (Amst)*. 44, 92–102.
- Shapiro, J.A., and von Sternberg, R. (2005). Why repetitive DNA is essential to genome function. *Biol. Rev.* 80, 227–250.
- Sharif, J., Endo, T.A., Toyoda, T., and Koseki, H. (2010). Divergence of CpG Island promoters: A consequence or cause of evolution? *Dev. Growth Differ.* 52, 545–554.
- Shayevitch, R., Askayo, D., Keydar, I., and Ast, G. (2018). The importance of DNA methylation of exons on alternative splicing. *RNA* 24, 1351–1362.
- Shema, E., Jones, D., Shores, N., Donohue, L., Ram, O., and Bernstein, B.E. (2016). Single-molecule decoding of combinatorially modified nucleosomes. *Science (80-)*. 352, 717 LP – 721.
- 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.
- Shimizu, Y., Uchimura, Y., Dohmae, N., Saitoh, H., Hanaoka, F., and Sugawara, K. (2010). Stimulation of DNA Glycosylase Activities by XPC Protein Complex: Roles of Protein-Protein Interactions. *J. Nucleic Acids* 2010, 1–12.
- Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* 479, 74–79.

- Simonet, N.G., Rasti, G., and Vaquero, A. (2016). Chapter 21 - The Histone Code and Disease: Posttranslational Modifications as Potential Prognostic Factors for Clinical Diagnosis. J.L.B.T.-E.B. and D. García-Giménez, ed. (Boston: Academic Press), pp. 417–445.
- Spada, F., Schiffers, S., Kirchner, A., Zhang, Y., Arista, G., Kosmatchev, O., Korytiakova, E., Rahimoff, R., Ebert, C., and Carell, T. (2020). Active turnover of genomic methylcytosine in pluripotent cells. *Nat. Chem. Biol.* 1–9.
- Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146–1159.
- Steinacher, R., Barekati, Z., Botev, P., Kuśnierczyk, A., Slupphaug, G., and Schär, P. (2019). SUMOylation coordinates BERosome assembly in active DNA demethylation during cell differentiation. *EMBO J.* 38, e99242.
- Takai, D., and Jones, P. a (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U. S. A.* 99, 3740–3745.
- Tan-Wong, S.M., Dhir, S., and Proudfoot, N.J. (2019). R-Loops Promote Antisense Transcription across the Mammalian Genome. *Mol. Cell* 76, 600-616.e6.
- Tellier, M., Maudlin, I., and Murphy, S. (2020). Transcription and splicing: A two-way street. *WIREs RNA* 11, e1593.
- The C. elegans Sequencing Consortium (1998). Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science* (80-). 282, 2012–2018.
- Theurillat, I., Hendriks, I.A., Cossec, J.-C., Andrieux, A., Nielsen, M.L., and Dejean, A. (2020). Extensive SUMO Modification of Repressive Chromatin Factors Distinguishes Pluripotent from Somatic Cells. *Cell Rep.* 32, 108146.
- Tian, J., Wang, L., and Da, L.-T. (2021). Atomic resolution of short-range sliding dynamics of thymine DNA glycosylase along DNA minor-groove for lesion recognition. *Nucleic Acids Res.*
- Tini, M., Benecke, A., Um, S.-J., Torchia, J., Evans, R.M., and Chambon, P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. *Mol. Cell* 9, 265–277.
- Tirard, S., Morange, M., and Lazcano, A. (2010). The Definition of Life: A Brief History of an Elusive Scientific Endeavor.
- Tompkins, J.D., Jung, M., Chen, C., Lin, Z., Ye, J., Godatha, S., Lizhar, E., Wu, X., Hsu, D., Couture, L.A., et al.

(2016). Mapping Human Pluripotent-to-Cardiomyocyte Differentiation: Methylomes, Transcriptomes, and Exon DNA Methylation “Memories.” *EBioMedicine* 4, 74–85.

Tovy, A., Spiro, A., McCarthy, R., Shipony, Z., Aylon, Y., Alton, K., Ainbinder, E., Furth, N., Tanay, A., Barton, M., et al. (2017). p53 is essential for DNA methylation homeostasis in naïve embryonic stem cells, and its loss promotes clonal heterogeneity. 1–14.

Tramantano, M., Sun, L., Au, C., Labuz, D., Liu, Z., Chou, M., Shen, C., and Luk, E. (2016). Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex. *Elife* 5.

Um, S., Harbers, M., Benecke, A., Pierrat, B., Losson, R., and Chambon, P. (1998). Retinoic acid receptors interact physically and functionally with the T:G mismatch-specific thymine-DNA glycosylase. *J. Biol. Chem.* 273, 20728–20736.

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., and Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.* 24, 896–905.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* 16, 178–189.

Viktorovskaya, O. V, and Schneider, D.A. (2015). Functional divergence of eukaryotic RNA polymerases: unique properties of RNA polymerase I suit its cellular role. *Gene* 556, 19–26.

Villaseñor, R., Pfaendler, R., Ambrosi, C., Butz, S., Giuliani, S., Bryan, E., Sheahan, T.W., Gable, A.L., Schmolka, N., Manzo, M., et al. (2020). ChromID identifies the protein interactome at chromatin marks. *Nat. Biotechnol.* 38.

Vinayak, S., and Ford, J.M. (2010). PARP Inhibitors for the Treatment and Prevention of Breast Cancer. *Curr. Breast Cancer Rep.* 2, 190–197.

Vitas, M., and Dobovišek, A. (2019). Towards a General Definition of Life.

Waddington, C.H. (1942). Canalization of Development and the Inheritance of Acquired Characters. *Nature* 150, 563–565.

Waddington, C.H. (1956). Genetic Assimilation of the Bithorax Phenotype. *Evolution (N. Y.)* 10, 1–13.

Waddington, C.H. (1957). *The Strategy of the Genes* (Routledge).

Walter, M., Teissandier, A., Pérez-Palacios, R., and Bourc’his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* 5, e11418.

Wang, H., Naghavi, M., Allen, C., Barber, R.M., Bhutta, Z.A., Carter, A., Casey, D.C., Charlson, F.J., Chen, A.Z., Coates, M.M., et al. (2016a). Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 388, 1459–1544.

Wang, T., Zeng, J., Lowe, C.B., Sellers, R.G., Salama, S.R., Yang, M., Burgess, S.M., Brachmann, R.K., and Haussler, D. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl. Acad. Sci.* 104, 18613–18618.

Wang, X., Hu, L., Wang, X., Li, N., Xu, C., Gong, L., and Liu, B. (2016b). DNA Methylation Affects Gene Alternative Splicing in Plants: An Example from Rice. *Mol. Plant* 9, 305–307.

Watanabe, S., Mishima, Y., Shimizu, M., Suetake, I., and Takada, S. (2018). Interactions of HP1 Bound to H3K9me3 Dinucleosome by Molecular Simulations and Biochemical Assays. *Biophys. J.* 114, 2336–2351.

Weber, A.R., Krawczyk, C., Robertson, A.B., Kuśnierczyk, A., Vågbo, C.B., Schuermann, D., Klungland, A., and Schär, P. (2016). Biochemical reconstitution of TET1–TDG–BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat. Commun.* 7, 10806.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* 39, 457–466.

Weinberg, D.N., Papillon-Cavanagh, S., Chen, H., Yue, Y., Chen, X., Rajagopalan, K.N., Horth, C., McGuire, J.T., Xu, X., Nikbakht, H., et al. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* 573, 281–286.

Weisenberger, D.J., Campan, M., Long, T.I., Kim, M., Woods, C., Fiala, E., Ehrlich, M., and Laird, P.W. (2005). Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res.* 33, 6823–6836.

Wiebauer, K., and Jiricny, J. (1989). In vitro correction of G o T mispairs to G o C pairs in nuclear extracts from human cells. *Nature* 339, 234–236.

Wiebauer, K., and Jiricny, J. (1990). Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G.T mispairs in nuclear extracts from human cells. *Proc. Natl. Acad. Sci. U. S. A.* 87, 5842–5845.

Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., and Adelman, K. (2015). Pausing of RNA Polymerase II Regulates Mammalian Developmental Potential through Control of Signaling Networks. *Mol. Cell* 58, 311–322.

- Wirz, A. (2014). Linking active DNA demethylation by Thymine DNA Glycosylase with epigenetic regulation of gene expression. Dr. Thesis.
- Yang, B.X., EL Farran, C.A., Guo, H.C., Yu, T., Fang, H.T., Wang, H.F., Schlesinger, S., Seah, Y.F.S., Goh, G.Y.L., Neo, S.P., et al. (2015). Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells. *Cell* 163, 230–245.
- Yang, Y., Cheung, H.H., Tu, J., Miu, K.K., and Chan, W.Y. (2016). New insights into the unfolded protein response in stem cells. *Oncotarget* 7, 54010–54027.
- Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J.P., Nissim-Rafinia, M., Cohen, A.H., Rippe, K., Meshorer, E., et al. (2015). HP1 Is Involved in Regulating the Global Impact of DNA Methylation on Alternative Splicing. *Cell Rep.* 10, 1122–1134.
- Ying, Q., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* 453, 519–524.
- Yu, S., Levi, L., Siegel, R., and Noy, N. (2012). Retinoic Acid Induces Neurogenesis by Activating Both Retinoic Acid Receptors (RARs) and Peroxisome Proliferator-activated Receptor β/δ (PPAR β/δ). *J. Biol. Chem.* 287, 42195–42205.
- Zarakowska, E., Czerwinska, J., Tupalska, A., Yousefzadeh, M.J., Gregg, S.Q., Croix, C.M.S., Niedernhofer, L.J., Foksinski, M., Gackowski, D., Szpila, A., et al. (2018). Oxidation Products of 5-Methylcytosine are Decreased in Senescent Cells and Tissues of Progeroid Mice. *Journals Gerontol. Ser. A* 73, 1003–1009.
- Zhang, Y., and M. Kohli, R. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *8*, 1385–1395.
- Zhang, T., Cooper, S., and Brockdorff, N. (2015). The interplay of histone modifications – writers that read. *EMBO Rep.* 16, 1467–1481.
- Zhang, Y., Xie, S., Zhou, Y., Xie, Y., Liu, P., Sun, M., Xiao, H., Jin, Y., Sun, X., Chen, Z., et al. (2014). H3K36 Histone Methyltransferase Setd2 Is Required for Murine Embryonic Stem Cell Differentiation toward Endoderm. *Cell Rep.* 8, 1989–2002.
- Zhu, H., Wang, G., and Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation. *Nat. Rev. Genet.* 17, 551–565.
- Zilberman, D. (2017). An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol.* 18, 87.
- Zuo, X., Sheng, J., Lau, H.-T., McDonald, C.M., Andrade, M., Cullen, D.E., Bell, F.T., Iacovino, M., Kyba, M.,

Xu, G., et al. (2012). Zinc Finger Protein ZFP57 Requires Its Co-factor to Recruit DNA Methyltransferases and Maintains DNA Methylation Imprint in Embryonic Stem Cells via Its Transcriptional Repression Domain. *J. Biol. Chem.* 287, 2107–2118.

Zuo, Z., Roy, B., Chang, Y.K., Granas, D., and Stormo, G.D. (2017). Measuring quantitative effects of methylation on transcription factor-DNA binding affinity.

7 Appendix

I – PARP Inhibition Induces Cytotoxicity in mESCs by Activating Endogenous Retroviruses

II – Covalent PARylation is an Integral Part of TDG-BER-mediated Active DNA Demethylation in Embryonic Stem Cells

III - Active DNA demethylation Evicts H2A.Z to Facilitate Pause Release of RNA Polymerase 2

IV - Inducible TDG Knockout Models to Study Epigenetic Regulation

PARP inhibition induces cytotoxicity in mESCs by activating endogenous retroviruses

Jianming Xu^{1,3}, Simon D. Schwarz^{1,3}, Kapila Gunasekera², Roland Steinacher¹, Michael O. Hottiger^{2, 3*} and Primo Schär^{1, 3, 4*}

¹Department of Biomedicine, University of Basel, Mattenstrasse 28, 4058 Basel, Switzerland

²Department of Molecular Mechanisms of Disease, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

³These authors contributed equally.

⁴Lead Contact: Primo Schär, primo.schaer@unibas.ch

*Corresponding authors:

Michael O. Hottiger, michael.hottiger@dmmd.uzh.ch

Primo Schär, primo.schaer@unibas.ch

Summary

PARP inhibitors (PARPi) have been approved for cancer therapy based on their synthetic lethality in combination with defects in homology-directed DNA double strand-break repair (DSBR), such as apparent in BRCA-deficient cancers. How PARPi affect DSBR proficient cells remains an important unanswered question. Here, we show that DSBR proficient naïve murine embryonic stem cells (mESCs) are highly sensitive to PARP inhibitors such as Talazoparib (Tal). Tal acts on PARP1 and activates p53 via ATM-dependent DNA damage signalling. CRISPR screening identified the 5-methylcytosine hydroxylases TET1/2 as key mediators of Tal-induced cytotoxicity. Depletion of TET1/2 as well as TDG rescued Tal sensitivity of mESCs, implicating active DNA demethylation in mediating p53-dependent Tal toxicity in mESCs. We show that TET and TDG facilitate Tal-induced, p53-dependent DNA stress response including not only the activation of conventional p53 targets but also a strong transcriptional induction of endogenous retroviral elements, contributing to the observed cytotoxicity. Our data establish a role active DNA demethylation in p53-dependent DNA stress response induced by PARPi in DSBR proficient stem cells, providing insight into molecular mechanism and markers underlying potential cytotoxic and epigenetic effects of PARPi-based cancer therapy.

Keywords: PARP inhibitors, synthetic lethality, p53, active DNA demethylation, TET, TDG, endogenous retroviruses.

Introduction

Protein ADP-ribosylation is catalyzed by ADP-ribosyltransferases (ARTs also called Poly-ADP-ribose polymerases; PARPs). ARTD1 or PARP1, the best-characterized intracellular ART family member, is responsible for the majority of nuclear ADP-ribosylation. This post-translational modification serves in the nucleus as a signal to recruit and regulate multi-protein complexes involved in genotoxic stress responses (Ciccia and Elledge, 2010). The recent development of PARP inhibitors (PARPi) as anti-cancer drugs made use of the multifaceted role of ADP-ribosylation in DNA repair. The successful application of PARPi in cancer therapy relies primarily on the concept of “synthetic lethality”, according to which homologous recombination (HR)-deficient cells (e.g. BRCA1 or BRCA2-deficient) fail to repair DNA double strand-breaks (DSBs) that accumulate following PARP inhibition (Bryant et al., 2005; Farmer et al., 2005). One possible source of such DSBs is “PARP trapping”. Inhibition of PARP1 by PARPi causes PARP1 to be trapped on DNA repair intermediates (Murai et al., 2012), particularly during DNA base excision repair (BER) (Brown et al., 2017; Ronson et al., 2018). This may in turn cause obstruction of replication forks, which requires BRCA-dependent HR to be resolved. Yet, a direct demonstration of an allosteric trapping mechanism *in vivo* is so far missing (Hopkins et al., 2015) and biochemical evidence of PARP1 trapping on chromatin seems to require high levels of DNA damage, such as induced by alkylating agents like methyl-methanesulfonate (Murai et al., 2014). Also, whether and exactly how inhibition of PARP affects HR proficient cells is an important unsolved question.

Mouse embryonic stem cells (mESCs) are derived from the inner cell mass of blastocysts; they are pluripotent and can replicate indefinitely. Different stages of *in vitro* pluripotency have been defined. When cultured in the presence of serum and leukemia inhibitory factor (Clifford et al. 2003), mESCs assume a “primed” state, in which the population is heterogeneous in terms of gene expression and developmental competence (Hayashi et al., 2008). When cultured in 2i medium, containing inhibitors of MEK1/2 and GSK3 β , mESCs remain in a “ground state” of pluripotency, characterized by a more homogeneous population and increased developmental competence (Silva et al., 2009). One hallmark of “2i” mESC is a globally hypomethylated, inner cell mass (ICM)-like genome (Leitch et al., 2013). DNA hypomethylation can be achieved by either a passive or active DNA demethylation processes. Passive DNA demethylation occurs via inhibition of the maintenance DNA methyltransferase (DNMT) and subsequent dilution of the 5-methylcytosine (5mC) during DNA replication; active mechanisms involve dioxygenases of the ten-eleven translocation (TET) protein family, which convert 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-

carboxylcytosine (5caC) in sequential oxidation steps (Wu and Zhang, 2014). Thymine DNA glycosylase (TDG)-mediated DNA base excision repair (BER) then replaces 5fC and 5caC with unmethylated cytosines (Jacobs and Schär, 2011; Kohli and Zhang, 2013). Both these processes contribute to the establishment and maintenance of the hypomethylated genome of “2i” mESCs (Sim et al., 2017; von Meyenn et al., 2016). Genomic regions that maintain DNA methylation also in “2i” mESCs coincide with inactive chromatin marked by trimethylation of H3K9 (H3K9me3) and include endogenous retroviral elements (ERVs) as well as imprinted loci (Habibi et al., 2013). ERVs are long terminal repeats (LTRs)-containing retroelements and account for approximately 10% of the mouse genome (Mouse Genome Sequencing et al., 2002). Due to their ability to replicate and re-integrate into the genome, their expression can be a source for genetic mutation and instability (Jern and Coffin, 2008; Maksakova et al., 2006). While ERV expression is stably silenced in most somatic cell types predominantly by DNA methylation (Bulut-Karslioglu et al., 2014), their loci are more permissive for expression in naïve mESCs, where they are mainly regulated by histone methylation (Deniz et al., 2018; Walter et al., 2016). The expression of certain ERVs in mESCs as well as in 2-cell stage embryos (Macfarlan et al., 2012) suggests a function of these elements in the regulation of host gene expression (Kunarso et al., 2010). In addition, certain families of LTR-containing ERVs contain near-perfect p53 binding sites (Wang et al., 2007) and, notably, loss of p53 was shown to increase genome-wide DNA methylation, including also ERV loci (Tovy et al., 2017).

In this study, we investigated the responses of HR-competent cells to inhibition of PARP. Comparing different cell lines, we found that naïve mESCs, although being BRCA-proficient, are highly sensitive to PARP1 inhibition by Talazoparib. Through genome-wide gene expression analyses and CRISPR/Cas9 screening we identified p53 and TET1 as a central component of the PARPi response in mESCs. We found that PARPi treatment induced genotoxic stress signalling via ATM and p53 in a TET and TDG-independent manner but p53-dependent cytotoxicity and gene expression responses downstream of TET and TDG action. Investigating these downstream events, we uncovered a role for TET/TDG mediated active DNA demethylation in facilitating p53-dependent upregulation of certain ERVs. Thus, our work demonstrates a critical role of p53 and TET/TDG-dependent active DNA demethylation in mediating cell death in naïve mESCs in response to PARP1 inhibition.

Results

Naïve mESCs are highly sensitive to PARPi

To investigate the response of HR-competent cells to PARPi, we exposed a variety of mouse cells lines with different properties (primary, immortalized; proliferating, post-mitotic) to Talazoparib (Tal; IC50, 0.57 nM for ARTD1) (Shen et al., 2013). As expected from current models of PARPi action, HR proficient NIH 3T3 cells, differentiated C2C12 cells (myotubes), bone marrow derived macrophages, and post-mitotic neurons were all resistant to Tal treatment (Figure 1A, right). Naïve mESCs cultivated in 2i medium (containing MEK1/2 and GSK3 β inhibitors), however, showed high sensitivity to Tal (Figure 1A, left). When using an unrelated mESC line (159, aka HA36CB1, derived from mixed 129-C57Bl/6 background blastocysts), which spontaneously accumulates differentiated cells when cultured in serum-containing media, we observed that Tal treatment affected predominantly the pluripotent (grown as spheres) but not the more differentiated cells (grown as monolayer).

While Tal treatment induced massive cell death in wild type mESCs, mESCs with either a constitutive (*Parp1*^{-/-}, Figure 1B, S1A) or a conditional PARP1 defect (*Parp1*^{fl/fl} upon Cre-induction, Figure 1C, S1B), were resistant to Tal treatment. The same effect was observed with a second PARPi – Olaparib (Ola; IC50, 5 nM for PARP1) indicating high specificity to PARP1 (Meneer et al., 2008) (Figure 1D and Figure S1C).

To clarify whether PARPi sensitivity of mESCs is related to their naïve pluripotency state, we conditioned the “2i” naïve mESCs in “serum” to primed mESCs. Serum-adaptation increased the tolerance of wild type mESCs to Tal without significantly affecting the proliferation rate (Figure 1E). Notably, addition of “2i” inhibitors did not sensitize NIH 3T3 cells to Tal (Figure S1D), suggesting that the naïve pluripotency state rather than the inhibition of MEK1/2 and/or GSK3 β sensitizes mESCs to Tal. Flow cytometry indicated that Tal treatment of mESCs induces a delay in cell cycle progression and accumulation in S phase, 14 hrs post treatment (Figure 1F, bottom). Upon prolonged treatment (36 hrs), cells arrested in G2/M phase and became apoptotic as seen by Annexin V presentation and propidium iodide penetration (Figure 1F, upper). By contrast, *Parp1*^{-/-} mESCs showed no response to Tal, neither in cell cycle progression nor in apoptosis up to 36 hrs of treatment (Figure 1F).

Taken together, these data document that HR-proficient naïve mESCs are highly sensitive to PARPi, that this sensitivity depends on PARP1, and that it is associated with the ground state pluripotency of the cells.

PARPi activates an ATM-dependent p53 response in naïve mESCs

To identify pathways responding to Tal treatment in naïve mESCs, we examined the effect of Tal exposure on gene expression by mRNA sequencing. PolyA⁺ RNA from WT and *Parp1*^{-/-} mESCs treated with Tal for 6, 12 or 24 hours. Wild type mESC showed pronounced transcriptional changes within the first 6 hours of Tal treatment, a majority of which lasted up to 24 hrs of treatment. By contrast, Tal treatment in *Parp1*^{-/-} cells induced only minor changes in gene expression (Figure 2A), with most of the genes activated in wild type mESCs being unchanged or even downregulated in the absence of PARP1 (Clusters 1-3, and 5, Figure 2A). Gene ontology analysis did not reveal any specific pathway affected. We therefore searched for transcription factors regulating those genes activated by Tal within the first 6 hrs of treatment and discovered the tumor suppressor protein p53 and components of the Polycomb repressive complex 2 (PRC2) as the most enriched factors (Figure S2A). To test the association with p53, we defined authentic p53 target genes responding to doxorubicin treatment and that have p53 bound at their promoters. Based on published p53 ChIP-seq data in mESCs (Lee et al., 2010), we identified 444 p53 target genes that were activated by doxorubicin and 219 genes (49%) were also upregulated within the first 6 hrs of Tal treatment while only 7 were downregulated (Figure 2B). Gene set enrichment analysis confirmed a significant enrichment of these authentic p53 target genes (Normalized Enrichment Score = 1.29, FDR < 0.05) (Figure 2C). To validate p53-dependent gene activation upon Tal treatment, we assessed the response of the well-known p53 target genes *Fas*, *Tap*, and *Wnt9a* by qPCR that were also detected in the mRNA-seq. Tal not only activated the direct p53 target genes, but also the non-p53 but potential secondary target *Hey2* in a p53-dependent manner (Figure S2B). These results establish that Tal treatment of mESCs activates the p53 pathway.

We next assessed p53 phosphorylation following Tal treatment of mESCs by immunoblotting with Ser15 and Ser392 phospho-specific antibodies. Tal induced p53 phosphorylation at both residues to a similar amount and comparable to the effect of topoisomerase 2 inhibition by Mer (Figure 2D). Neither Tal nor Ola was able to activate p53 in U2OS or RPE-1 cells under the tested condition (Figure S2C), indicating that activation of the p53 pathway by Tal is a specific feature of mESCs. Chromatin-immunoprecipitation (ChIP) established that Tal-induced p53 phosphorylation in wild type mESCs correlated with its increased enrichment at target gene promoters of *Fas* and *Tap1* but not at the non-p53 target promoters of *Dll1* and *Hey2* (Figure 2E). Tal treatment of *Parp1*^{-/-} mESCs had no significant effect on p53 enrichment at any of the analysed promoters (Figure 2E). Consistent with p53 dependent gene activation, Tal also increased trimethylation of histone H3 lysine 4 (H3K4me3)

at the *Fas* and *Tap1* promoters, which was not detectable at the *Dll1* and *Hey2* promoters (Figure 2F).

To determine the signalling events leading to p53 activation in Tal-exposed mESCs, we tested the contribution of the two transducer kinases ATM and ATR, making use of specific inhibitors. Treatment of wild type mESCs with the ATM-specific inhibitor KU-55933 (Hickson et al., 2004) partially repressed Tal-induced gene activation (Figure 2G). Besides p53, ATM also phosphorylates histone H2A.X at Ser139 in response to DNA damage, generating γ H2A.X aggregates, (Ismail and Hendzel, 2008; Kinner et al., 2008). Tal induced H2A.X phosphorylation as well as γ H2A.X foci in wild type mESCs (Figure S2D/E), which were both reduced by ATM but not by ATR inhibition. In addition, targets of ATR-dependent phosphorylation, CHK1 and RPA2, were not phosphorylated upon Tal treatment (Figure S2F). siRNA-mediated depletion of ATM and its downstream checkpoint kinase (CHK2) (Matsuoka et al., 2000) either individually or in combination reduced Tal-induced p53 phosphorylation (Figure 2H) as well as gene activation (Figure 2I and S2G), implicating both ATM and CHK2 in p53 activation upon PARP inhibition.

P53 is the central mediator of Tal-induced cytotoxicity

Given the role of p53 in the activation of target genes upon PARP1 inhibition, we used total RNA-seq to investigate the dependency of this response on p53 expression. mESCs were treated with control siRNA or a sip53 before the treatment with 5 nM Tal for 6 hrs. Out of 1'173 significantly upregulated genes upon Tal in WT mESCs, 956 (81.5%) were downregulated or showed no response when p53 was depleted prior to the application of Tal (Figure 3A/B). In fact, the transcriptional response induced by Tal, showed a strong inverse behaviour on a genome-wide level when p53 was depleted (Figure 3C), which indicates p53 to be a master regulator of Tal-induced transcriptional changes. These observations could also be validated by qPCR (Figure 3D). Furthermore did the depletion of p53 lead to a resistance against Tal-mediated cytotoxicity (Figure 3E/F), indicating that activity of p53 is involved in the apoptotic signature observed earlier (Figure 1F).

TET/TDG-dependent processes mediate Tal-induced cytotoxicity downstream of p53 activation

To further investigate the mechanisms underlying PARPi-induced cell death in mESCs, we performed a genome-wide CRISPR screen (Doench et al., 2016) (Figure 4A and B). A total of 441 sgRNAs induced prolonged cell survival upon Tal treatment in a least one of three replicate cultures. Among the 25 top hits (Figure 4C), two genes were identified by all four targeting

sgRNAs represented in the library, *Parp1* and *Tet1*. As *Parp1*^{-/-} mESCs are resistant to PARPi (Figure 1B and S1B), the identification of *Parp1* confirmed the functionality of the screen. Notably, *Tet1* was identified as the second top hit, indicating that 5mC oxidation may cause Tal toxicity in mESCs. Depletion of TET1, but not TET2, by siRNA, partially rescued the survival of wild type mESCs exposed to Tal, consistent with the screening result (Figure 4E). Simultaneous TET1 and TET2 depletion, however, further enhanced cell survival in the presence of Tal while TET3 depletion did not have any additional effect (Figure 4D and E), consistent with its poor expression in mESCs (Figure S3C) (Huang et al., 2014).

Until now, no function other than mC oxidation is known about TET proteins. To address whether TET proteins mediate Tal sensitivity via TDG/BER-dependent active DNA demethylation, we first used *Tdg*^{-/-} mESCs reconstituted with a *Tdg* mini-gene expressed at endogenous levels TDG (TDGwt) (Cortazar et al., 2011), a catalytic dead mutant (TDGcat, N151A) (Cortellino et al., 2011; Hardeland et al., 2000; Steinacher et al., 2019) or the empty vector only (TDGnull). Consistent with a previous observation (Pettitt et al., 2018), TDGnull mESC showed resistance to Tal, and so did TDGcat mESCs (Figure S3D). The inactivation of TDG not only rescued Tal-induced cell death but also the G2/M cell cycle arrest and expression of apoptosis (Figure S3E/F) This effect was also reproduced with an unrelated mESC cell line carrying a 4-hydroxytamoxifen-inducible (4OHT) TDG knockout allele (*Tdg*^{fl/fl}) (Schwarz et al., 2020) (Figure 4F/G and S3G).

Together with the dependence on TET activity, these results suggested that TDG mediates Tal-induced cell death through its DNA glycosylase activity, presumably in the context of active DNA demethylation via base excision repair (BER). PARP1 inhibition is expected to cause an accumulation of DSBs, which leads to cell death in HR-deficient cells and presumably to the ATM-dependent p53 activation (Figure 2G-I). Performing pulse field gel electrophoresis, we could see that DSB accumulation also happens in HR-proficient mESCs that are treated with Tal and that this effect is PARP1-dependent (Figure S3H). We argued that SSBs, caused by the high activity of TET/TDG-initiated BER in naïve mESCs (Mulholland et al., 2020), could be the reason for the increase of DSBs upon Tal treatment. We therefore assessed DSBs in TDG-depleted mESCs and observed the same level of DSBs in mESCs with and without TDG (Figure 4H), indicating that the DSBs are not the dominant cause of Tal-mediated cytotoxicity.

Consistent with the present DSBs, p53 was still phosphorylated in Tal-exposed cells after TDG depletion (Figure 4I) as well as of TET1 and TET2 (Figure S3I), demonstrating that DNA damage signalling is activated independent of TET/TDG-induced DNA demethylation. The same was observed after Zeocin treatment (Zeo) that is inducing DNA damage unrelated to

substrates of TDG (Ehrenfeld et al., 1987; Hecht, 2000 et al.). Both Tal and Zeo induced comparable p53 phosphorylation independent of the presence of TDG (Figure S3J). Considering that both TET and TDG depletion rescued the Tal-induced cytotoxicity despite the presence of DSBs and p53 activation, these results demonstrate that the toxicity in Tal-exposed mESCs is not caused by active DNA demethylation-induced BER upstream of p53 signalling but rather downstream of p53 activation in signal transduction.

TDG facilitates p53-dependent gene expression

The observation that Tal activates p53 without the need of either TET or TDG, but then causes cytotoxicity that is dependent on TET and TDG, prompted us to investigate the role of TET/TDG in Tal-induced gene expression in mESCs. Out of 22,356 genes represented in our mRNA-seq dataset, 1,859 genes showed a more than 2-fold upregulation upon Tal treatment (Figure 5A). Using a publicly available TDG ChIP-seq dataset (Neri et al. 2015), we detected that 1,594 (86%) of these genes showed TDG enrichment either at the promoter and/or in the gene body, pointing out an overrepresentation of TDG in these genes (p-value of chi-square testing observed over expected: p-value=1.87e-17). To investigate TDG's role in p53-dependent Tal-induced gene activation, we tested the three p53 target genes *Alox5*, *Cpz*, *Fas* that showed significant upregulation in both the mRNA and total RNA-seq data. In all three cases, Tal induced strong activation in TDG wild type mESCs while the activation was significantly lower when TDG was either absent or inactive (Figure S4A). These results were reproduced with the 4OHT-inducible mESC TDG KO cell line. (Figure 5B), suggesting that TDG activity facilitates p53 dependent gene activation in response to Tal. Consistently, all examined p53 target genes (*Alox5*, *Cpz*, and *Fas*) showed positive enrichment of TDG and p53 at their promoters (Figures S4B), and p53 recruitment to the promoter of these genes was significantly increased upon Tal treatment in a TDG dependent manner (Figure 5C). Notably, depletion of TET1 and 2 on top of a TDG KO did not further reduce the gene-inducibility when compared to the separate depletions of TETs and TDG (Figure S4C), consistent with an epistatic interaction between TET and TDG in this context. Hence, these data strongly suggest that TET and TDG-mediated active DNA demethylation facilitates p53-dependent gene expression in mESCs.

Tal activates ERV expression in a p53- and TDG-dependent manner

Previous studies reported the de-repression of repetitive elements that are under control of p53 upon DNA damage, as a potential threat to cellular survival (Leonova et al., 2013 and reviewed

in Tiwari et al., 2018). We therefore made use of our total RNAseq and measured the expression of repetitive elements. We observed that the main categories of retrotransposable elements, including long terminal repeats (LTRs), long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs) as well as DNA repeat elements showed significantly increased expression upon Tal treatment (Figure S4D). The upregulation was most poignant in endogenous retroviral elements (ERVs) as part of LTR-containing elements (Figure S4D). Based on a public ChIP-seq dataset of phosphorylated p53 upon doxorubicin (Li et al 2012), we saw that these elements showed a larger proportion of p53 binding than other repeats, (Figure S4E), thereby facilitating the transcriptional response to PARP1 inhibition (Figure 5D), which is consistent with previous reports (Wang, 2007). Not only activation of the top 50 differentially expressed ERVs was negated upon p53 depletion by siRNA (Figure 5E) but also 89% of all significantly differentially detected LTRs (Figure S4F/G). Moreover, the Tal-induced expression of all significantly detected LTRs depended on the presence of p53 (Figure 5F). Validation of three chosen LTRs (*Rltr10a*, *Rltr45*, and *Rmer19b*) detected from the RNA-seq, confirmed these genome-wide effects, clearly establishing p53 as a necessary mediator of Tal-induced expression of repetitive elements (Figure S4H).

In accordance with the observation that TET and TDG was required for efficient activation of canonical p53 target genes following Tal treatment, we found that TDG depletion nearly abolished Tal-induced ERV expression (Figure 5G, Figure S4I), and combined TDG and TET1/2 KO did not further decrease the Tal-mediated de-repression (Figure S4J).

Taken together, these data showed for the first time that PARP inhibition induces expression of repetitive elements, particularly ERVs, in naïve mESCs in a p53- and TET/TDG-dependent manner.

TET/TDG/BER-mediated DNA single strand-breaks (SSBs) facilitate p53-dependent transcription

To understand the mechanism underlying TDG-facilitated transcription at repetitive elements, we first determined DNA methylation levels with bisulfite conversion, followed by pyrosequencing across selected CpGs at two ERVs and the canonical p53 target *Fas*. CpGs at *Rltr45* and *Rmer19b* showed an average of 40% - 60% methylation whereas CpGs at *Fas* showed 5%-10% methylation (Figure S5A). Treatment with Tal and/or depletion of TDG did not detectably alter the 5mC states at the sites assessed. Global levels of oxidized 5mC derivatives, analyzed by mass-spectrometry, showed an increase of 5fC and 5caC following TDG disruption, but no change in 5hmC (Figure S5B-D). Notably, Tal treatment increased global levels of 5hmC in TDG-proficient and deficient cells and 5fC and 5caC specifically in

TDG deficient cells (Figure S5B-D). These results indicated that Tal increases detectability of active DNA demethylation intermediates in mESCs, despite changes at examined ERV becoming upregulated with Tal are minor (Figure S5A).

Since Tal did not change steady-state levels of DNA methylation, yet induced ERV expression in a TDG-dependent manner, we investigated whether DNA demethylation-associated DNA strand incisions occur at ERVs and whether these facilitate p53-dependent transcriptional activation. We captured and mapped SSBs by SSB-seq (Baranello et al., 2014, 2018) in triplicates of WT and TDG-depleted cells, each with or without Tal treatment. We detected an over-representation of CpGs in the reads of captured fragments when compared to the non-enriched input, indicating an enrichment for regions with active DNA demethylation (Figure S5E, Xu and Schwarz et al., in submission). In line with previous observations, SSB-enriched regions are predominantly found in genic and regulatory regions (i.e., enhancers, promoters and 5'UTRs, Figure S5G) (Cao et al., 2019, Sriramachandran et al., 2020, Baranello et al., 2014). Alignment of all detected SSB peaks showed a strong coincidence with published TDG ChIP-seq data from mESCs (Figure 6A, Neri et al., 2015), as well as with p53 (Figure 6B, Li et al., 2012). Although the association of SSBs with TDG and p53 was reduced in TDG-depleted mESCs, the reduction was unintuitively small. The regions of SSB-enrichment, however, appeared less dense at the peak center and more spread out to the immediate flanking regions as well as spanning over a larger genetic area, suggesting a loss of DNA-repair accuracy/coordination in TDG-deficient mESCs (Figure S5H/I).

These data in combination with the observation that TDG-enriched regions associate very well with p53-enriched regions (Figure 6C), strongly suggested that TDG cooperated with p53 as a transcription factor on a genome-wide level involving DNA strand-break formation. In line with this, Tal treatment resulted in an increased density of SSBs coinciding with TDG and p53 (Figure 6D/E).

Given that Tal induced p53- and TDG-dependent activation of LTR transcription (Figure 5D-G) as well as TDG-dependent SSB formation at sites of p53-enriched chromatin, we wondered whether this also applies to LTRs. p53 is strongly enriched over detected LTRs after damage induction by doxorubicin (Li et al. 2012, Figure S5J), TDG however, only shows minor enrichment in unchallenged mESCs (Neri et al. 2015, Figure S5K). Measuring the SSB signal at all detectable LTRs that become upregulated upon Tal treatment (Figure 5G), showed indeed that SSBs increase in response to Tal treatment, particularly at p53-bound LTRs and even more so at those with additionally detected TDG occupation (Figure 6F). The Tal-induced increase of SSBs at upregulated LTRs was TDG-dependent as in the absence of TDG, SSB were reduced

(Figure 6G). This indicates that TDG/BER-mediated SSBs contribute to efficient transcriptional regulation upon stress-induction.

We wondered how expression of p53 targets upon Tal is facilitated by SSBs, and argued that the targeted DNA repair increased accessibility (Madders et al., 2020, Odell et al., 2013). We therefore performed ATAC-seq to investigate whether TDG activity impacts chromatin accessibility in inducible TDG KO mESCs that were treated with Tal for 24 hrs. We detected a similar number of peaks among all samples (Figure S5L), with a minute increase (1%) upon Tal treatment that was visible in both, TDG pro- and deficient cells. TDG-deficient cells, however, showed slightly lower number of steady-state accessible sites (-1%).

Alignment of total detected ATAC peaks showed a high enrichment of accessible sites coinciding with sites of significant p53 enrichment (ChIP-seq) (Figure 6H). Sites of significant accessibility coincided similarly well with TDG presence in WT mESCs, but were reduced upon Tal (Figure 6I), indicating a certain re-distribution of accessibility from TDG-bound sites in unchallenged mESCs to sites targeted by p53. We therefore subsetted the identified ATAC peaks that are unique to the condition, and saw an increase in the association of p53-binding and accessibility that was TDG-dependent (Figure 6J).

Consistent with a role of chromatin accessibility in facilitating transcription, our data show a strong positive correlation between transcriptional activity (RNA-seq), SSB density (SSB-seq) and local chromatin accessibility (ATAC-seq) (Figure 6K), which is still present, although reduced, in Tal-treated mESCs (Figure 6L). When we focused our analysis to p53-targeted LTRs, we did detect increased accessibility at those sites upon Tal treatment. This increase of accessibility, however, seems largely TDG-independent (Figure S5M).

Taken together, these data demonstrate for the first time a role of TDG in facilitating p53-dependent activation of LTR transcription in Tal treated mESC through BER-associated SSB formation. While the accessibility at LTRs is increased upon Tal treatment, the presence or absence of TDG does not affect it.

Tal-induced cytotoxicity in mESC is potentially caused by activation of ERV transcription leading to necroptosis and an interferon-like response

The aforementioned data showed that under 2i culture conditions, LTR-containing ERVs were activated upon Tal treatment in a p53- and TDG-dependent manner. We wondered whether this could contribute to the cytotoxic phenotype associated with Tal treatment. GSEA showed a significant representation of genes involved in the type 1 interferon response upon Tal-treatment (Figure S6A) and a trend towards significance of genes involved in necroptosis

(S6A/B). Both these cellular responses are associated to viral infections and sensing of dsRNA and displayed a significant underrepresentation in Tal-treated mESCs depleted of p53 (S6B/C). To address whether activation of ERV transcription may be the cause of Tal-induced cytotoxicity, we assessed the cellular responses to transfection of *in vitro* transcribed double-stranded ERV RNAs (Figure S6D). Indeed, we saw a cytotoxic effect to WT mESCs upon the transfection of the ERV dsRNAs, which was less pronounced after transfection of the dsRNA-mimicking poly-inosinic:cytidylic acid (poly-I:C) (Figure 7A). This response was accompanied by the upregulation of *Mkl1* and *Zbp1*, both components of the necroptosis pathway that were also detected in the RNA-seqs (Figure 7B), but no upregulation of canonical p53 targets was detected. We included the human colorectal adenocarcinoma cell line HT-29 as a positive control, known to be sensitive to dsRNA transfection and they displayed heavy cell death after the transfection of ERV dsRNAs as well as the poly-I:C (Figure 7C). As expected, transfection of ERV dsRNA was able to induce an interferon response in HT-29 cells (Figure 7D). In contrast, only two out of four interferon-responders (*Irf7* and *Ifi27*) were activated when we repeated the experiment in mESCs (Figure 7E), suggesting an incomplete interferon response in mESCs. Taken together, these data suggest that Tal-mediated activation of ERV transcription induces cell death in mESCs mediated by an interaction of necroptosis and (type 1) Interferon response (Sarhan et al., 2019).

Discussion

Although PARP inhibitors (PARPi) have been approved for the treatment of BRCA-deficient cancer (Brown et al., 2016; Turk and Wisinski, 2018) our understanding of the mechanisms underlying PARPi-induced cell toxicity is incomplete. The data presented here show that naïve pluripotent mESCs (2i condition), which are proficient in homology directed DSBR and therefore expected to be PARPi resistant, are in fact hypersensitive to the treatment. Both FDA-approved PARPi Talazoparib (Tal) and Olaparib induced cell death in mESCs, which was not observed in PARP1 KO mESCs (Figures 1A-D), suggesting that both tested PARPi specifically targeted PARP. Notably, the sensitivity was reduced but still observable when mESCs were cultured with serum, in line with a recent report performing CRISPR screening for the identification of PARP1 mutants resistant to Tal-induced cell death (Pettitt et al., 2018). Inhibition of PARP1 in several somatic cell lines with intact DSBR also showed that they are not sensitive to PARPi (Figure 1A), suggesting that the toxicity of PARPi in DSBR-proficient cells is related to specific features of pluripotent stem cells. The unique sensitivity of mESCs towards PARPi highlights an important role of PARP1 enzymatic activity in the maintenance and/or establishment of this cellular state, as was also observed in iPSCs generation where PARP1 proficiency significantly promoted iPSCs derivation (Chiou et al., 2013; Doege et al., 2012; Roper et al., 2014; Weber et al., 2013). The molecular mechanisms underlying pluripotency and lineage specification have been largely attributed to the unique structure and dynamics of chromatin in naïve mESCs, which is globally DNA hypomethylated (Habibi *et al.*, 2013; von Meyenn *et al.*, 2016) and decondensed (Meshorer and Misteli, 2006; Meshorer et al., 2006; Orkin and Hochedlinger, 2011).

TET and TDG-mediated active DNA demethylation takes an essential part in the maintenance of the transcriptionally highly active state of mESCs and are necessary for faithful differentiation (Mulholland et al., 2020, Cortázar et al., 2011, Dawlaty et al., 2014). By initiating BER, active DNA demethylation will create SSB and, thereby, substrates for PARP1 binding. TET (Kharat et al., 2020) and TDG-deficient mESCs (Cortazar et al., 2011, Steinacher et al., 2019, Pettitt et al., 2018) were independently observed to be less sensitive to PARPi. Although we cannot exclude TET/TDG/BER-mediated active DNA demethylation leading to some initiation of DNA stress signalling, our data suggest a predominant role of active DNA demethylation in the execution and relay of signalling downstream of p53 activation. This is supported by two key observations: 1. DNA double-strand break formation was unaffected in TDG-deficient mESCs showing resistance to Tal-mediated cytotoxicity - and 2. Tal was able to induce p53 phosphorylation in the absence of TET (Figure S3I) or TDG (Figure 4I and S3J).

Notably, these observations also show that Tal cytotoxicity in DSBR-proficient mESCs is not primarily a consequence of the DSBs formed, as expected, but of TET and TDG-dependent processes downstream of ATM and p53-mediated signalling. Consistently, genome-wide analyses revealed that Tal-induced SSBs at p53 binding sites occur in a TDG-dependent manner and correlate with gene expression at examined p53 targets (Figure 6 and S5). Interestingly, Tal induced gene activation had no obvious effect on steady-state levels of DNA methylation, indicating that cytosine modification turnover rather than the absolute level of the modification is relevant for the transcriptional responses examined. In the context of a globally hypomethylated and well-accessible genome, this concept has recently received increased attention (Wan et al., 2015, Ginno et al., 2020). In addition to the activation of canonical p53 target genes, we observed a strong upregulation of endogenous retroviral elements (ERVs) upon the treatment with Tal. This upregulation led to the activation of genes involved in necroptosis and interferon responses, which likely contributed to the observed cytotoxicity (Figures 7 and S6). While we could not fully quantitate the contribution of aberrant ERV expression to the Tal cytotoxicity, the upregulation of these transposable elements should be taken into consideration when PARPi is applied in the therapeutic context. Notably, hyperactive and promiscuous transcription is considered as a hallmark of pluripotent mESCs. Cancer cells with stem cell features have been reported to share a similar transcriptional regulatory network with embryonic stem cells (Kim and Orkin, 2011; Kim et al., 2010). Thus, our findings have implications on the behaviour of tumors under PARPi-based therapy, suggesting that the stem cell-like fraction of cancer cells in particular might respond by reactivation of silenced genomic areas and increased genetic instability. Also, the observation that PARPi-induced cytotoxicity in mESCs depends on p53 suggests that PARPi treatment of BRCA deficient cancers will reduce tumor mass via the mechanisms of synthetic lethality, but cells with mutated p53 or p53 deficiency would escape the treatment and give rise to relapse. Based on our results, we would therefore suggest that the status and activity of PARP1, TET/TDG and p53 should be taken into consideration for selecting patients that might profit from PARPi based therapy.

Further investigation is required to understand the mechanistic details underlying the differential response of different cell types to PARP inhibition and cellular stresses in general. This will not only advance our understanding of the biology of genome-environment but also help predict responses to and long term effects on cell- and tissue homeostasis of cancer therapy, be it targeted or conventional.

Methods

PARP1 inducible KO mESCs

Parp1-floxed mice (*Parp1^{fl/fl}*) were crossed with Rosa26CreER mice (Ventura et al., 2007) to generate PARP1 conditional KO mESCs (Kunze et al., manuscript in preparation).

Cell culture

Parp1^{+/+}, *Parp1^{-/-}*, and *Parp1^{fl/fl}* mouse ESCs were cultured in “2i” medium (DMEM/F-12 and Neurobasal medium supplemented with N2, B27, 0.1 mM 2-mercaptoethanol, 1× penicillin/streptomycin, 3 μM CHIR99021, 1 μM PD0325901, and 1000 U/ml LIF) on gelatinized plates. *Tdg^{+/+}*, *Tdg^{-/-}*, and *Tdg^{fl/fl}* mESCs were cultured in “2i” medium supplemented with 0.5 mM sodium pyruvate without antibiotics. For the transition into serum culture, 2i medium was replaced with DMEM GlutaMAX supplemented with 10% FBS, 10 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol, 1× non-essential amino acids, 1× penicillin/streptomycin, and 1000 U/ml LIF. Cells were cultured for another 5 passages before adding Tal. The same medium was used to culture 159 mESCs. NIH 3T3 cells were cultured in DMEM GlutaMAX supplemented with 10% FBS and 1× penicillin/streptomycin. HT-29 cells were cultured in DMEM GlutaMAX supplemented with 10% FBS. For siRNA/dsRNA transfection, cells were reverse transfected with either Lipofectamine RNAiMAX or Mirus TransIT-X2 reagent.

Flow cytometry

For cell cycle analysis, mESCs were treated as described and harvested by trypsinization and washed once with ice-cold PBS. After washing, 100 μl of cell suspension was diluted with 900 μl of DAPI staining solution (2 μg/ml DAPI, 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM MgCl₂, and 0.1% Triton X-100) and analyzed on a LSRFortessa (BD Biosciences). For apoptosis analysis, cells were harvested by trypsinization and washed once with ice-cold PBS. After washing, cells were resuspended in 1× binding buffer and stained with Annexin V-FITC and propidium iodide. Apoptotic and necrotic cells were quantified on a LSRFortessa or a CytoFLEX S. Results were analyzed with FlowJo using Watson (pragmatic) model.

Immunofluorescence staining for γH2A.X

Cells grown on glass cover slips were treated as described and fixed with 4% formaldehyde. After fixation, cells were permeabilized with 0.2% Triton X-100 in PBS. Subsequently, cells were blocked by 2% BSA and 0.1% Triton X-100 in PBS. After blocking, cells were first

stained with anti- γ H2A.X Ab (Upstate, #05-636) followed by Cy3-conjugated goat anti-mouse secondary Ab (Jackson ImmunoResearch). Finally, cells were stained with Hoechst 33258 stain solution and mounted with Vectashield mounting solution.

PFGE

Cells were treated as described and harvested by trypsinization and washed once with cold PBS. Typically, after washing 2.5×10^5 cells in 100 μ l PBS were mixed with an equal volume of premelted low melting point agarose. The mixture was transferred to the plug mold (BioRad) and subject to Proteinase K digestion in lysis buffer (100 mM EDTA pH 8.0, 0.2% Sodium deoxycholate, 1% sarcosine and 1 mg/ml Proteinase K) at 37°C for 72 hrs. After digestion, the plug was washed three times with washing buffer (20 mM Tris-HCl pH 8.0 and 50 mM EDTA) and subject to electrophoresis for 21 hrs.

RNA extraction, reverse transcription and quantitative real-time PCR

Cells were treated as described and total RNA was extracted either with TRI Reagent (Sigma-Aldrich) or RNeasy Plus Mini kit (QIAGEN). After extraction, 1 μ g of total RNA was digested with DNase I and reverse transcribed using either MultiScribe reverse transcriptase (ThermoFisher) or RevertAid First Strand cDNA synthesis kit (ThermoFisher) for first strand cDNA synthesis. Real-time PCR was performed with either KAPA FAST SYBR kit (Roche) or SensiFAST SYBR kit (Bioline) on Rotor-Gene Q (QIAGEN). Relative gene expression was determined using primers that spanned exon-exon junctions and either *Actb* or *Rps13* as an internal control for normalization.

Library preparation and sequencing for RNA-seq

Total RNA was extracted with TRI Reagent. The quality of the isolated RNA was determined with a Qubit® (1.0) Fluorometer (Life Technologies, California, USA) and a Bioanalyzer 2100 (Agilent, Waldbronn, Germany). Only samples with a 260 nm/280 nm ratio between 1.8–2.1 and a 28S/18S ratio within 1.5–2 were further processed. For mRNA-seq the TruSeq RNA Sample Prep Kit v4 (Illumina, Inc, California, USA) was used in the succeeding steps. Briefly, total RNA samples (100-1000 ng) were poly-A enriched and then reverse-transcribed into double-stranded cDNA. The cDNA samples were fragmented, end-repaired and polyadenylated before ligation of TruSeq adapters containing the index for multiplexing. Fragments containing TruSeq adapters on both ends were selectively enriched with PCR. The quality and quantity of

the enriched libraries were validated using Qubit® (1.0) Fluorometer and the Caliper GX LabChip® GX (Caliper Life Sciences, Inc., USA).

The TruSeq SR Cluster Kit HS3000 (Illumina, Inc, California, USA) was used for cluster generation using 10 pM of pooled normalized libraries on the cBOT. Sequencing was performed on the Illumina HiSeq 4000 with single end 150 bp using the TruSeq SBS Kit HS3000 (Illumina, Inc, California, USA).

RNA-seq analysis

RNA-seq reads were aligned with the STAR-aligner (Dobin et al., 2013). As a reference, we used the Ensembl mouse genome build GRCm38. Gene expression values were computed with the function `featureCounts` from the R package `Rsubread` (Liao et al., 2013). Differential expression was computed using the generalized linear model implemented in the Bioconductor package `DESeq2` (Love et al., 2014).

The Sushi (Hatakeyama et al., 2016) framework was used to run the data analyses. Heatmaps were generated with R scripts using the function `heatmap.2` of the R package `gplots`.

Genome-wide CRISPR screening

Parp1^{+/+} mESCs were first transduced with lentiviruses expressing Cas9 (lentiCas9-Blast) (Sanjana et al., 2014). lentiCas9-Blast was a gift from Feng Zhang (Addgene viral prep #52962-LV; <http://n2t.net/addgene:52962>; RRID: Addgene_52962). After selection with 10 µg/ml Blasticidin, mESCs stably expressing Cas9 were established. For the screening, Cas9-expressing mESCs were transduced with mouse Brie CRISPR knockout pooled lentiviral library (Doench et al., 2016), which was a gift from David Root and John Doench (Addgene #73633). One day after transduction, cells were selected with 0.75 µg/ml puromycin to get rid of untransduced cells. Puromycin selection was kept until the cells were harvested. Four days after transduction, cells were treated with 5 nM Tal for 4 days. Afterwards, cells were grown without Tal for another 7 days to recover. Fifteen days after transduction, 5 nM Tal was added again. Cells were harvested at 14, 17, and 19 days post transduction. Library was prepared by amplifying the sgRNA sequences using the Illumina Amplicon primers and sequenced with MiSeq.

SSB-seq and library preparation

The detection of endogenous single-strand breaks (SSB) was performed as being published before (Baranello et al., 2014; Baranello et al., 2018) with considerable adaptations.

Briefly, genomic DNA was extracted with the Genomic-Tip 100/G Kit (Qiagen) from freshly treated mESCs. Fifty micrograms of gDNA were subjected to the nick labelling using 5 µl of *E. coli* DNA polymerase I (NEB) and a labelling mix containing 2 µM DIG-dUTPs. The labelled DNA was purified and fragmented with Bioruptor (Diagenode) and a restriction enzyme cocktail (EcoRI, HindIII, XbaI). Immunoprecipitation was performed over night with 2 µg of anti-DIG antibody (Roche) and the antibody-bound complexes were recovered with 40 µl of pre-blocked Protein G Dynabeads (Invitrogen).

Libraries of input and IP samples (10 ng DNA each) were prepared using the KAPA HyperPrep Kit (Roche) following the manufacturers protocol. Subsequent paired-end sequencing (75 cycles) was performed on an Illumina MiSeq system at the genomics facility Basel to an average depth of ~50 mio reads per sample before filtering.

SSB-seq analysis

Reads were aligned to the mouse genome (mm10 UCSC version) with bowtie2 (version 2.3.4.2) and extra options "--maxins 2000 --no-mixed --no-discordant --local --mm". Duplicates were marked with picard tools (version 2.9.2) and resulting BAM files were filtered by removing reads falling into ENCODE blacklist regions (version 2014 + manual removal of high coverage artefact regions) using Rsamtools (R version 3.6, Bioconductor version 3.10). Leading to an average coverage of at least 25 mio reliable reads per sample. Peaks were called across all replicates of a sample group using macs2 (version 2.1.2.1) and extra options "--t \$IPs -c \$INPUTs -f BAM -g 2652783500 --keep-dup all --nomodel --shift -100 --extsize 200 -broad". Peaks were called across all replicates of a sample group using HOMER (version 2.1.2.1) with specifying commands "--style histone -o auto". Genome-wide counts per million (CPM) were normalized to the total amount of filtered reads in 500bp windows applying a trimmed mean of M-values normalization using edgeR (Version 3.30). Analysis and visualizations were performed with following packages under R version 4.0.3: GenomicRanges_1.14, ChIPpeakAnno_3.22.2, ChIPseeker_1.24, metagene2_1.4, ggplot2_3.3.

ATAC-seq, library preparation, sequencing and processing

Tdg^{fl/fl} mESCs were treated with 5 nM Tal for 24 hrs. According to the "Omni-ATAC" seq protocol (Corces et al 2017), potential cell-free DNA was digested with 200 U/ml DNase I in 2i medium for 30 min at 37 °C. About 100'000 cells were harvested for each condition and replicate. Chromatin tagmentation and library preparation was done by using the ATAC-seq kit from Active Motif (#53150). ATAC-seq library was sequenced on the Illumina NovaSeq PE150

at Novogene (UK), to a coverage of at least 150 mio reads per sample.

FastQC (v 0.11.9) was used to check the quality of the reads in ATACSeq fastq files and adaptor sequence contamination. Reads were mapped to the mm10 mouse genome assembly using STAR algorithm (v 2.7) with specifying commands `--alignIntronMax 1 --outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --outFilterMatchNmin 0 --outFilterMismatchNmax 2`. High resolution peak calling over replicates was performed using Genrich (v 0.6) with the specifying commands `-p 0.005 -q 0.005 -a 20.0 -j -m 30 -r -v -l 50`. Signals of SSBs or ATAC outside of significant enrichment (peaks) were counted as log₂ fold enrichment of IP over Input by using “bamCount” (bamsignals package V1.22.0) and normalization by trimmed mean of M-values according to the library size using edgeR (V3.13). Signals of accessibility outside of significant enrichment (peaks) were counted as log₂ counts per million by using “bamCount” (bamsignals package V1.22.0) and normalization by trimmed mean of M-values according to the library size using edgeR (V3.13).

Cell fractionation

Cells were treated as described and harvested by trypsinization and resuspended in ice-cold hypotonic buffer A (10 mM Hepes pH 8.0, 10 mM KCl, 5 mM MgCl₂, 0.1% Triton X-100, 1 mM DTT and 1× protease inhibitor cocktail). After centrifugation, the cytoplasmic fraction was precipitated by TCA and the nuclear pellet was resuspended in ice-cold buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT and 1× protease inhibitor cocktail). Chromatin was separated from the nucleoplasm by centrifugation. Cellular proteins were detected by WB using the following antibodies: anti-TOP2A Ab (abcam, ab52934), anti-ARTD1 Ab (Santa Cruz, sc-53643), anti- α -Tubulin Ab (Sigma-Aldrich, T6199).

Protein extraction and analysis

Protein levels were analyzed by western blotting of 50 μ g of NP-40 whole-cell extracts, separated by SDS-PAGE, onto a nitrocellulose membrane (Amersham) and immunodetected with the following antibodies: anti-murine TDG Ab (L58, Lab P. Schär), anti-GAPDH Ab (Sigma-Aldrich, G9545; Santa Cruz, sc-25778), anti- β -Actin Ab (Sigma-Aldrich, A5441), anti-p53 Ab (Santa Cruz, sc-393031/sc-126), anti-Phospho-p53 (S15) Ab (Cell Signaling, #9284), anti-Phospho-p53 (S392) Ab (Santa Cruz, sc-51690), anti-53BP1 Ab (Santa Cruz, sc-22760), anti-ARTD1 Ab (Santa Cruz, sc-7150/sc-53643), anti-Pol II Ab (Santa Cruz, sc-899), anti-Phospho-ATM (S1981) Ab (abcam, ab81292), anti-Chk1 Ab (Santa Cruz, sc-8408), anti-Phospho-Chk1 (S317) Ab (Cell Signaling, #2344), anti-PCNA Ab (Santa Cruz, sc-56), anti-

Phospho-RPA32 (S33) Ab (Bethyl, A300-246A), anti-Histone H3 Ab (abcam, ab1791), anti-Phospho-H2A.X (S139) Ab (Biolegend, #613402), anti-Sox2 Ab (Santa Cruz, sc-365823), anti-Tubulin Ab (Sigma-Aldrich, T6199).

Chromatin immunoprecipitation

Cells were treated as described and crosslinked with 1% formaldehyde. Cross-linked cells were lysed in lysis buffer 1 (50 mM Hepes pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100 and 1× protease inhibitor cocktail). After centrifugation, the nuclear pellet was washed once with lysis buffer 2 (10 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA and 1× protease inhibitor cocktail) and lysed in buffer 3 (10 mM Tris-HCl pH 8.0, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.5% sarcosine, 0.1% sodium deoxycholate and 1× protease inhibitor cocktail). The nuclear lysate was sonicated with a Bioruptor 300 (Diagenode) and subjected to immunoprecipitation with Ab coupled to Protein A/G Dynabeads in IP buffer (10 mM Tris-HCl pH 8.0, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate and 1× protease inhibitor cocktail) overnight at 4°C. Beads were washed three times with high salt wash buffer (50 mM Hepes pH 7.6, 0.5 M LiCl, 1 mM EDTA, 1% NP-40, 0.7% sodium deoxycholate and 1× protease inhibitor cocktail) and subsequently once with low salt wash buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl and 1 mM EDTA). Immuno-complexes were eluted with pre-warmed elution buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA and 1% SDS) and reverse cross-linked overnight at 65°C. Then, samples were subjected to RNase A digestion followed by Proteinase K digestion. DNA was extracted with NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL) and quantified by real-time PCR.

ADP-ribosylated peptide enrichment

Mouse ESCs were treated as described and harvested by adding modified RIPA buffer as described before (Martello et al., 2016) and scraping off the cells from the surface of the plates. The collected cell lysate was sonicated and centrifuged prior to proceeding with sample preparation including enrichment of ADP-ribosylated peptides using the Af1521 macrodomain protein for enrichment of ADP-ribosylated proteins based on the product-dependent MS/MS measurement.

Liquid chromatographic/mass spectrometric analysis for ADP-ribosylome

Mass spectrometric analysis was performed on an Orbitrap Fusion mass spectrometer (ThermoFisher Scientific) coupled to a nano EasyLC 1000 liquid chromatograph (ThermoFisher Scientific). In the settings, the template for HCD-Product Preview-ETHcD method was selected for the detection of ADP-ribosylated peptides (Bilan et al., 2017). This method includes data-dependent high-energy dissociation (HCD), followed by electron-transfer and higher-energy dependent (ETHcD) MS/MS, when more than one ADP-ribose fragment peak (136.0623, 250.0940, 348.07091, and 428.0372) was observed in the HCD scan.

MS data analysis

Raw files were deconvoluted to separate HCD and ETHcD spectra and written into a Mascot generic format (MGF) files using Proteome Discoverer v2.1 (ThermoFisher Scientific). These MGF files were further processed as described before (Bilan *et al.*, 2017). The MGF files resulting from measurement following the peptide enriched from mESCs were searched against the UniProtKB mouse database (taxonomy 10090 functional genomic center version 20140715), which included 24,905 Swiss-Prot entries; 34,616 TrEMBL entries; 50,783 decoy hits; and 262 common contaminants. Peptide sequences were identified with previously described search settings (Rosenthal et al., 2015) using Mascot 2.5.1.3 (Matrix Science). Four missed cleavages were allowed after setting enzyme specificity to trypsin. The ADP-ribose variable modification was set to a mass shift of 541.0611, with scoring of the neutral losses equal to 347.0631 and 249.0862. The marker ions at m/z 428.0372, 348.0709, 250.0940, and 136.0623 were ignored for scoring. The residues S, R, and Y were set as variable ADP-ribose acceptor sites and oxidation as a variable modification on M. Carbamidomethylation was set as a fixed modification on C. Peptides with Mascot score >20 and an expectation value <0.05 are considered as correctly assigned peptides. The peptides having a mascot localization score $>95\%$ according to the ETHcD method were considered for the ADP-ribosylation site analyses.

Statistical information

In our mass-spectrometry analysis we only considered the ADPr-peptides having mascot ion score > 20 and an expectation value < 0.05 , as correctly assigned peptides. Only the ADPr-sites assigned with site localization confidence $\geq 95\%$ within the peptides reported in the ETHcD fragmentation method were considered for the ADP-ribosylation site analysis. In our RNASeq analysis genes that were identified as significantly differentially expressed by cuffdiff (Trapnell

et al., 2013) were considered for further analysis after filtering for the genes with ≥ 2 -fold increase/decrease in abundance.

Acknowledgements

We thank Dr. Hubert Rehrauer from the Functional Genomic Center Zurich for his help in RNA-seq and MS measurement and Dr. Florian Geier from the Swiss Institute of Bioinformatics for his input in the analysis of the SSB-seq. We thank Dr. Paolo Cinelli from University Hospital Zurich for the isolation of WT, *Parp1*^{-/-}, and *Parp1*^{fl/fl} mESCs used in this study. We also thank Dr. Faiza Noreen from the DBM (University of Basel) for the help of pyrosequencing and Dr. David Schürmann from the DBM (University of Basel) for proofreading of the manuscript. J.X. is supported by the Mäxi-Foundation and K.G. by the Swiss cancer research foundation (KFS-3740-08-2015). ADP-ribosylation research in the laboratory of M.O.H. is funded by the Canton of Zurich and the Swiss National Science Foundation Grants (SNF 310030_157019, 31003A_176177). The authors declare no conflict of interest.

Author Contributions

J.X. and M.O.H. initiated the project. J.X., M.O.H. and P.S. designed the experiments. J.X. and S.D.S performed the experiments and analyzed the data. K.G. did the MS measurement of ADP-ribosylated protein in mESCs and analyze the data. K.G. analyzed the RNA-seq and mapped the ATAC-seq. S.D.S performed and analyzed the SSB-seq in combination with ATAC and RNA-seq. R.S. did cytotoxicity analysis of Tal with *Tdg*^{-/-} mESCs complemented with TDGwt/TDGvec/TDGcat. J.X., S.D.S, M.O.H. and P.S. prepared the manuscript. All authors critically reviewed the manuscript.

Declaration of Interests

The authors declare no competing interests.

References

- Baranello, L., Kouzine, F., Wojtowicz, D., Cui, K., Przytycka, T.M., Zhao, K., and Levens, D. (2014). DNA break mapping reveals topoisomerase ii activity genome-wide. *Int J Mol Sci* *15*, 13111-13122. 10.3390/ijms150713111.
- Baranello, L., Kouzine, F., Wojtowicz, D., Cui, K., Zhao, K., Przytycka, T.M., Capranico, G., and Levens, D. (2018). Mapping DNA breaks by next-generation sequencing. *Methods Mol Biol* *1672*, 155-166. 10.1007/978-1-4939-7306-4_13.
- Bilan, V., Leutert, M., Nanni, P., Panse, C., and Hottiger, M.O. (2017). Combining higher-energy collision dissociation and electron-transfer/higher-energy collision dissociation fragmentation in a product-dependent manner confidently assigns proteomewide adp-ribose acceptor sites. *Anal Chem* *89*, 1523-1530. 10.1021/acs.analchem.6b03365.
- Brown, J.S., Kaye, S.B., and Yap, T.A. (2016). Parp inhibitors: The race is on. *Br J Cancer* *114*, 713-715. 10.1038/bjc.2016.67.
- Brown, J.S., O'Carrigan, B., Jackson, S.P., and Yap, T.A. (2017). Targeting DNA repair in cancer: Beyond parp inhibitors. *Cancer Discov* *7*, 20-37. 10.1158/2159-8290.CD-16-0860.
- Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005). Specific killing of brca2-deficient tumours with inhibitors of poly(adp-ribose) polymerase. *Nature* *434*, 913-917. 10.1038/nature03443.
- Bulut-Karslioglu, A., De La Rosa-Velázquez, Inti A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galán, C., Winter, Georg E., Engist, B., Gerle, B., et al. (2014). Suv39h-dependent h3k9me3 marks intact retrotransposons and silences line elements in mouse embryonic stem cells. *Mol Cell* *55*, 277-290. 10.1016/j.molcel.2014.05.029.
- Chiou, S.H., Jiang, B.H., Yu, Y.L., Chou, S.J., Tsai, P.H., Chang, W.C., Chen, L.K., Chen, L.H., Chien, Y., and Chiou, G.Y. (2013). Poly(adp-ribose) polymerase 1 regulates nuclear reprogramming and promotes ipsc generation without c-myc. *J Exp Med* *210*, 85-98. 10.1084/jem.20121044.
- Ciccio, A., and Elledge, S.J. (2010). The DNA damage response: Making it safe to play with knives. *Mol Cell* *40*, 179-204. 10.1016/j.molcel.2010.09.019.
- Clifford, B., Beljin, M., Stark, G.R., and Taylor, W.R. (2003). G2 arrest in response to topoisomerase ii inhibitors: The role of p53. *Cancer Res* *63*, 4074-4081.
- 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. 10.1038/nature09672.
- 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. 10.1016/j.cell.2011.06.020.
- Deniz, O., de la Rica, L., Cheng, K.C.L., Spensberger, D., and Branco, M.R. (2018). Setdb1 prevents tet2-dependent activation of iap retroelements in naive embryonic stem cells. *Genome Biol* *19*, 6. 10.1186/s13059-017-1376-y.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). Star: Ultrafast universal rna-seq aligner. *Bioinformatics* *29*, 15-21. 10.1093/bioinformatics/bts635.
- Doerge, C.A., Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A., et al. (2012). Early-stage epigenetic modification during somatic cell reprogramming by parp1 and tet2. *Nature* *488*, 652-655. 10.1038/nature11333.

- Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191. 10.1038/nbt.3437.
- Ehrenfeld, G.M., Shipley, J.B., Heimbrook, D.C., Sugiyama, H., Long, E.C., van Boom, J.H., van der Marel, G.A., Oppenheimer, N.J., and Hecht, S.M. (1987). Copper-dependent cleavage of DNA by bleomycin. *Biochemistry* **26**, 931-942. 10.1021/bi00377a038.
- Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., et al. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917-921. 10.1038/nature03445.
- Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360-369. 10.1016/j.stem.2013.06.002.
- Hardeland, U., Bentele, M., Jiricny, J., and Schar, P. (2000). Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis. *J Biol Chem* **275**, 33449-33456. 10.1074/jbc.M005095200.
- Hatakeyama, M., Opitz, L., Russo, G., Qi, W., Schlapbach, R., and Rehrauer, H. (2016). Sushi: An exquisite recipe for fully documented, reproducible and reusable NGS data analysis. *BMC Bioinformatics* **17**, 228. 10.1186/s12859-016-1104-8.
- Hayashi, K., de Sousa Lopes, S.M.C., Tang, F., Lao, K., and Surani, M.A. (2008). Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3**, 391-401. 10.1016/j.stem.2008.07.027.
- Hecht, S.M. (2000). Bleomycin: New perspectives on the mechanism of action. *J Nat Prod* **63**, 158-168. 10.1021/np990549f.
- Hickson, I., Zhao, Y., Richardson, C.J., Green, S.J., Martin, N.M., Orr, A.I., Reaper, P.M., Jackson, S.P., Curtin, N.J., and Smith, G.C. (2004). Identification and characterization of a novel and specific inhibitor of the ataxia-telangiectasia mutated kinase ATM. *Cancer Res* **64**, 9152-9159. 10.1158/0008-5472.CAN-04-2727.
- Hopkins, T.A., Shi, Y., Rodriguez, L.E., Solomon, L.R., Donawho, C.K., DiGiandomino, E.L., Panchal, S.C., Wilsbacher, J.L., Gao, W., Olson, A.M., et al. (2015). Mechanistic dissection of PARP1 trapping and the impact on in vivo tolerability and efficacy of PARP inhibitors. *Mol Cancer Res* **13**, 1465-1477. 10.1158/1541-7786.MCR-15-0191-T.
- Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W.A., Kang, J., Zepeda-Martinez, J.A., Pape, U.J., Jacobsen, S.E., Peters, B., and Rao, A. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **111**, 1361-1366. 10.1073/pnas.1322921111.
- Ismail, I.H., and Hendzel, M.J. (2008). The gamma-H2AX: Is it just a surrogate marker of double-strand breaks or much more? *Environ Mol Mutagen* **49**, 73-82. 10.1002/em.20358.
- Jacobs, A.L., and Schär, P. (2011). DNA glycosylases: In DNA repair and beyond. *Chromosoma* **121**, 1-20. 10.1007/s00412-011-0347-4.
- Jern, P., and Coffin, J.M. (2008). Effects of retroviruses on host genome function. *Annu Rev Genet* **42**, 709-732. 10.1146/annurev.genet.42.110807.091501.
- Kim, J., and Orkin, S.H. (2011). Embryonic stem cell-specific signatures in cancer: Insights into genomic regulatory networks and implications for medicine. *Genome Med* **3**, 75. 10.1186/gm291.
- Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010). A MYC network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell* **143**, 313-324. 10.1016/j.cell.2010.09.010.

- Kinner, A., Wu, W., Staudt, C., and Iliakis, G. (2008). Gamma-h2ax in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Res* 36, 5678-5694. 10.1093/nar/gkn550.
- Kohli, R.M., and Zhang, Y. (2013). Tet enzymes, tdg and the dynamics of DNA demethylation. *Nature* 502, 472-479. 10.1038/nature12750.
- Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics* 42, 631-634. 10.1038/ng.600.
- Lee, K.H., Li, M., Michalowski, A.M., Zhang, X., Liao, H., Chen, L., Xu, Y., Wu, X., and Huang, J. (2010). A genomewide study identifies the wnt signaling pathway as a major target of p53 in murine embryonic stem cells. *Proc Natl Acad Sci U S A* 107, 69-74. 10.1073/pnas.0909734107.
- Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nat Struct Mol Biol* 20, 311-316. 10.1038/nsmb.2510.
- Liao, Y., Smyth, G.K., and Shi, W. (2013). The subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 41, e108. 10.1093/nar/gkt214.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biol* 15, 550. 10.1186/s13059-014-0550-8.
- Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57-63. 10.1038/nature11244.
- Maksakova, I.A., Romanish, M.T., Gagnier, L., Dunn, C.A., van de Lagemaat, L.N., and Mager, D.L. (2006). Retroviral elements and their hosts: Insertional mutagenesis in the mouse germ line. *PLoS Genet* 2, e2. 10.1371/journal.pgen.0020002.
- Martello, R., Leutert, M., Jungmichel, S., Bilan, V., Larsen, S.C., Young, C., Hottiger, M.O., and Nielsen, M.L. (2016). Proteome-wide identification of the endogenous adp-ribosylome of mammalian cells and tissue. *Nat Commun* 7, 12917. 10.1038/ncomms12917.
- Matsuoka, S., Rotman, G., Ogawa, A., Shiloh, Y., Tamai, K., and Elledge, S.J. (2000). Ataxia telangiectasia-mutated phosphorylates chk2 in vivo and in vitro. *Proc Natl Acad Sci U S A* 97, 10389-10394. 10.1073/pnas.190030497.
- Menear, K.A., Adcock, C., Boulter, R., Cockcroft, X.L., Copsey, L., Cranston, A., Dillon, K.J., Drzewiecki, J., Garman, S., Gomez, S., et al. (2008). 4-[3-(4-cyclopropanecarbonylpiperazine-1-carbonyl)-4-fluorobenzyl]-2h-phthalazin-1-one: A novel bioavailable inhibitor of poly(adp-ribose) polymerase-1. *J Med Chem* 51, 6581-6591. 10.1021/jm8001263.
- Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol* 7, 540-546. 10.1038/nrm1938.
- Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* 10, 105-116. 10.1016/j.devcel.2005.10.017.
- Mouse Genome Sequencing, C., Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520-562. 10.1038/nature01262.
- Mulholland, C.B., Traube, F.R., Ugur, E., Parsa, E., Eckl, E.-M., Schöning, M., Modic, M., Bartoschek, M.D., Stolz, P., Ryan, J., et al. (2020). Distinct and stage-specific contributions of tet1 and tet2 to stepwise cytosine oxidation in the transition from naive to primed pluripotency. *Scientific Reports* 10, 12066. 10.1038/s41598-020-68600-3.

- Murai, J., Huang, S.Y., Das, B.B., Renaud, A., Zhang, Y., Doroshow, J.H., Ji, J., Takeda, S., and Pommier, Y. (2012). Trapping of parp1 and parp2 by clinical parp inhibitors. *Cancer Res* 72, 5588-5599. 10.1158/0008-5472.CAN-12-2753.
- Murai, J., Huang, S.Y., Renaud, A., Zhang, Y., Ji, J., Takeda, S., Morris, J., Teicher, B., Doroshow, J.H., and Pommier, Y. (2014). Stereospecific parp trapping by bmn 673 and comparison with olaparib and rucaparib. *Mol Cancer Ther* 13, 433-443. 10.1158/1535-7163.MCT-13-0803.
- Orkin, S.H., and Hochedlinger, K. (2011). Chromatin connections to pluripotency and cellular reprogramming. *Cell* 145, 835-850. 10.1016/j.cell.2011.05.019.
- Pettitt, S.J., Krastev, D.B., Brandsma, I., Drean, A., Song, F., Aleksandrov, R., Harrell, M.I., Menon, M., Brough, R., Campbell, J., et al. (2018). Genome-wide and high-density crispr-cas9 screens identify point mutations in parp1 causing parp inhibitor resistance. *Nat Commun* 9, 1849. 10.1038/s41467-018-03917-2.
- Ronson, G.E., Piberger, A.L., Higgs, M.R., Olsen, A.L., Stewart, G.S., McHugh, P.J., Petermann, E., and Lakin, N.D. (2018). Parp1 and parp2 stabilise replication forks at base excision repair intermediates through fbh1-dependent rad51 regulation. *Nat Commun* 9, 746. 10.1038/s41467-018-03159-2.
- Roper, S.J., Chrysanthou, S., Senner, C.E., Sienerth, A., Gnan, S., Murray, A., Masutani, M., Latos, P., and Hemberger, M. (2014). Adp-ribosyltransferases parp1 and parp7 safeguard pluripotency of es cells. *Nucleic Acids Res* 42, 8914-8927. 10.1093/nar/gku591.
- Rosenthal, F., Nanni, P., Barkow-Oesterreicher, S., and Hottiger, M.O. (2015). Optimization of Itq-orbitrap mass spectrometer parameters for the identification of adp-ribosylation sites. *J Proteome Res* 14, 4072-4079. 10.1021/acs.jproteome.5b00432.
- Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for crispr screening. *Nat Methods* 11, 783-784. 10.1038/nmeth.3047.
- Schwarz, S.D., Grundbacher, E., Hrovat, A.M., Xu, J., Kuśnierczyk, A., Vågbø, C.B., Schär, P., and Schuermann, D. (2020). Inducible tdg knockout models to study epigenetic regulation. *F1000Research* 9. 10.12688/f1000research.25637.1.
- Shen, Y., Rehman, F.L., Feng, Y., Boshuizen, J., Bajrami, I., Elliott, R., Wang, B., Lord, C.J., Post, L.E., and Ashworth, A. (2013). Bmn 673, a novel and highly potent parp1/2 inhibitor for the treatment of human cancers with DNA repair deficiency. *Clin Cancer Res* 19, 5003-5015. 10.1158/1078-0432.CCR-13-1391.
- Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. *Cell* 138, 722-737. 10.1016/j.cell.2009.07.039.
- Sim, Y.J., Kim, M.S., Nayfeh, A., Yun, Y.J., Kim, S.J., Park, K.T., Kim, C.H., and Kim, K.S. (2017). 2i maintains a naive ground state in escs through two distinct epigenetic mechanisms. *Stem Cell Reports* 8, 1312-1328. 10.1016/j.stemcr.2017.04.001.
- Steinacher, R., Barekati, Z., Botev, P., Kusnierczyk, A., Slupphaug, G., and Schar, P. (2019). Sumoylation coordinates berosome assembly in active DNA demethylation during cell differentiation. *EMBO J* 38. 10.15252/embj.201899242.
- Tovy, A., Spiro, A., McCarthy, R., Shipony, Z., Aylon, Y., Allton, K., Ainbinder, E., Furth, N., Tanay, A., Barton, M., and Oren, M. (2017). P53 is essential for DNA methylation homeostasis in naive embryonic stem cells, and its loss promotes clonal heterogeneity. *Genes Dev* 31, 959-972. 10.1101/gad.299198.117.
- Turk, A.A., and Wisinski, K.B. (2018). Parp inhibitors in breast cancer: Bringing synthetic lethality to the bedside. *Cancer*. 10.1002/cncr.31307.

- Ventura, A., Kirsch, D.G., McLaughlin, M.E., Tuveson, D.A., Grimm, J., Lintault, L., Newman, J., Reczek, E.E., Weissleder, R., and Jacks, T. (2007). Restoration of p53 function leads to tumour regression in vivo. *Nature* *445*, 661-665. 10.1038/nature05541.
- von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N.Q., Salehzadeh-Yazdi, A., Santos, F., Petrini, E., Milagre, I., Yu, M., Xie, Z., et al. (2016). Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. *Mol Cell* *62*, 848-861. 10.1016/j.molcel.2016.04.025.
- Walter, M., Teissandier, A., Perez-Palacios, R., and Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* *5*. 10.7554/eLife.11418.
- Wang, T., Zeng, J., Lowe, C.B., Sellers, R.G., Salama, S.R., Yang, M., Burgess, S.M., Brachmann, R.K., and Haussler, D. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci U S A* *104*, 18613-18618. 10.1073/pnas.0703637104.
- Weber, F.A., Bartolomei, G., Hottiger, M.O., and Cinelli, P. (2013). Artd1/parp1 regulates reprogramming by transcriptional regulation of fgf4 via sox2 adp-ribosylation. *Stem Cells* *31*, 2364-2373. 10.1002/stem.1507.
- Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: Mechanisms, genomics, and biological functions. *Cell* *156*, 45-68. 10.1016/j.cell.2013.12.019.

Figure 1. Tal treatment kills naïve mESCs through PARP1.

- A. Representative pictures show the sensitivity of various cell lines to Tal. (left) WT (Ola129/E14) mESCs grown in 2i or serum, 159 mESCs grown in 2i, and NIH 3T3 cells were treated with 5 nM Tal for 2 days. Scale bar represents 50 μ m. (right) post-mitotic neurons were first treated with 5 nM Tal for 5 days followed by another 4 day treatment with 50 nM Tal. Bone marrow-derived macrophages and myotubes (C2C12) were treated with 50 nM Tal for 6 days. Scale bar represents 20 μ m.
- B. Representative pictures show sensitivity of mESCs to Tal. WT and *Parp1*^{-/-} mESCs were treated with 5 nM Tal for 2 days. Scale bar represents 20 μ m.
- C. (left) Representative pictures show the sensitivity of 2i-cultured mESCs to Tal. *Parp1*^{fl/fl} mESCs either uninduced or induced with 200 nM 4OHT for 3 days, were treated with 5 nM Tal for 2 days. Scale bar represents 50 μ m. (right), WB analysis for PARP1 KO efficiency.
- D. Representative pictures show the sensitivity of mESCs to Olaparib (Ola). WT and *Parp1*^{-/-} mESCs were treated with 10 μ M Ola for 3 days. Scale bar represents 20 μ m.
- E. MTT assay showing the cell viability of WT and *Parp1*^{-/-} (KO) mESCs grown in either 2i or serum upon 5 nM Tal treatment. Data are shown as mean + SD (n=3).
- F. Flow cytometry measuring propidium-iodide (PI) and Annexin V levels after Tal-treatment, showing apoptosis (upper) and cell cycle changes (lower) in WT and *Parp1*^{-/-} (KO) mESCs.

Figure 2. Tal treatment activates p53 signaling in WT mESCs

- A. Heatmap with hierarchical clustering shows differential gene expression of WT and *Parp1*^{-/-} mESCs treated with 5 nM Tal for indicated time periods.
- B. (top) pie chart showing differential gene expression of WT mESCs treated with 5 nM Tal for 6 hrs. (bottom) Venn diagram showing the overlap between Tal-regulated genes and p53 targets, identified by p53-binding (Li et al. 2013).
- C. Gene set enrichment analysis for p53 target genes in the gene expression of WT mESCs treated with 5 nM Tal for 6 hrs, FDR < 0.05.
- D. Representative WB analysis of WT mESCs treated with 5 nM Tal or 10 μ M Merbarone (Mer) for 6 hrs (n=3).
- E. ChIP-qPCR analysis of phospho-p53 in WT and *Parp1*^{-/-} mESCs treated with 50 nM Tal for 6 hrs. Data are shown as mean + SD (n=4 for WT, n=2 for *Parp1*^{-/-} cells).
- F. ChIP-qPCR analysis of H3K4me3 in WT mESCs treated with 5 nM Tal for 4 hrs. Data are shown as mean + SD (n=3).
- G. Relative mRNA expression by qRT-PCR of WT mESCs treated with 5 nM Tal for 6 hrs upon ATM inhibition (10 μ M KU-55933). Data are shown as mean + SD (n=3).
- H. WB analysis of WT mESCs treated with 5 nM for 6 hrs upon ATM/CHK2 or p53 KD (n=3). A: ATM; C: CHK2.
- I. Relative mRNA expression by qRT-PCR in WT mESCs treated with 5 nM Tal for 6 hrs upon ATM and CHK2 or p53 KD. Data are shown as mean + SD (n=3). Asterisks indicate statistical p-values of: *: p \leq 0.05, **: p \leq 0.01, ***: p \leq 0.001, ****: p \leq 0.0001

Figure 3. Tal-induced p53 activation is responsible for gene expression as well as cell death in mESCs

- A. Heatmap with hierarchical clustering showing differential gene expression of WT mESCs treated with 5 nM Tal for 6 hrs upon p53 KD (left). (right) clusters 1, 2, 3, 8, and 9 represent genes whose differential expression upon Tal treatment is p53 dependent.
- B. Double contrast comparing p53-dependent gene expression (sip53 +Tal vs. Tal) with Tal-induced gene expression (Tal vs. DMSO).
- C. Relative mRNA expression by qRT-PCR analysis in WT mESCs treated with 5 nM Tal for 11 hrs upon p53 KD. Data are shown as mean + SD (n=4).
- E. Representative pictures (n=3) show the sensitivity of mESCs to Tal upon p53 KD. WT mESCs transfected with p53 siRNA were treated with 5 nM Tal for 2 days. Scale bar represents 50 μ m.
- F. MTT assay shows the cell viability of WT mESCs treated with 5 nM Tal upon p53 KD. Data are shown as mean + SD (n=3). Asterisks indicate statistical p-values of: *:p \leq 0.05, **: p \leq 0.01, ***: p \leq 0.001

Figure 4. TET1 and TDG are essential contributors to Tal induced cell death in mESCs

- A and B. Schematic illustration and detailed timeline showing the workflow of the CRISPR screening.
- C. Top 25 hits from the screening. Individual sgRNAs were grouped according to their target gene and ranked by the number of reads detected per sgRNA.
- D. Representative pictures showing the sensitivity of mESCs to Tal upon TET KD. WT mESCs transfected with indicated siRNAs were treated with 5 nM Tal for 2 days. Scale bar represents 50 μ m.
- E. MTT assay showing the cell viability of WT mESCs treated as in D. Data are shown as mean + SD (n=4).
- F. Representative pictures show the sensitivity of inducible TDG KO mESCs to Tal. *Tdg^{fl/fl}* mESCs either uninduced or induced with 2 μ M 4OHT for 2 hrs were treated with 5 nM Tal for 24 hrs. Scale bar represents 40 μ m.
- G. Quantitation of viable mESCs treated as in F and shown as mean + SD (n=3). **: p \leq 0.01
- H. (left) PFGE of inducible TDG KO mESCs treated with 5 nM Tal for indicated time. Zeocin (Zeo) serves as a positive control. (right) quantification of DSBs. Data are presented as mean + SD (n=3).
- I: WB analysis of p53 activation in *Tdg^{fl/fl}* mESCs either uninduced or induced with 2 μ M 4OHT for 2 hrs prior to treatment with 5 nM Tal for 24 hrs. (n=2).

Figure 5. TDG acts downstream of p53 to facilitate p53 targeted gene as well as ERV expression

- A Volcano plot depicting differentially expressed genes upon Tal treatment. Red dots indicate significantly deregulated genes. Black dots indicate genes which are occupied by TDG either at the promoter and/or in the gene body. TDG occupancy was retrieved from published ChIP-seq data in mESCs (Neri et al., 2015).
- B. qRT-PCR analysis in *Tdg^{fl/fl}* mESCs treated with 5 nM Tal for 24 hrs. Data are shown as mean + SD (n=3).

- C. ChIP-qPCR analysis of phospho-p53 in *Tdg^{fl/fl}* mESCs treated with 5 nM Tal for 24 hrs. Data are shown as the mean fold enrichment of the percentage of input (+Tal/-Tal) + SD (n=3). A non-targeting region ~120kb upstream of *Chordc1* is used as a negative control.
- D. Expression of all significantly detected (at least 10 counts) LTRs (all, n=57'343) and those having a p53 binding site within the sequence (over p53, n=1'397).
- E. Heat map showing expression levels of the top 50 ERVs deregulated upon 6 hrs of Tal in WT and p53 KD.
- F. Double contrast of gene expression comparing p53-dependent LTR expression (sip53 +Tal vs. Tal) with Tal-induced LTR expression (Tal vs. DMSO).
- G. qRT-PCR analysis in *Tdg^{fl/fl}* mESCs treated with 5 nM Tal for 24 hrs. Data are shown as mean + SD (n=3). Asterisks indicate statistical p-values of: *:p≤0.05, **: p≤0.01, ***: p≤0.001, ****: p≤0.0001

Figure 6. Tal treatment induces SSBs at ERVs in a TDG-dependent manner.

- A and B. Comparison of SSB-seq peak density in *Tdg^{fl/fl}* mESCs with or without 4OHT induction. Distribution of all detected SSB-seq peaks 3 kb around the center of called TDG (Neri et al. 2015) (A) or p53 (Li et al. 2012) (B) peaks from published ChIP-seq data in mESCs. Lighter shadings indicate the 95% confidence interval.
- C. Heatmap showing the distribution of TDG peaks 3 kb around the center of p53 peaks from published ChIP-seq data.
- D and E. Comparison of SSB-seq peak density in *Tdg^{fl/fl}* mESCs with or without Tal treatment, 3 kb around the center of called TDG (D) or p53 (E) peaks from published ChIP-seq data in mESCs. Lighter shadings indicate the 95% confidence interval.
- F. SSB-seq signal as TMM-normalized log₂ counts per million (CPM) at all significantly detected LTRs (minimum 10 counts, n=57'343), LTRs occupied by either p53 alone (n=1'397) or by p53 and TDG together (n=507).
- G. SSB-seq signal at all detected LTRs or LTRs with p53 occupation, in *Tdg^{fl/fl}* ESCs, with or without prior TDG deletion.
- H and I. Comparison of ATAC-seq peak density in *Tdg^{fl/fl}* mESCs with or without 4OHT induction, with and without Tal-treatment. Distribution of all detected ATAC-seq peaks 3 kb around the center of called p53 (H) or TDG (I). Lighter shadings indicate the 95% confidence interval.
- J. Density of condition-specific ATAC-seq peaks over called p53 peaks.
- K and L Scatter plot showing the correlation of ATAC-seq (y-axis) and SSB-seq (x-axis) signal as TMM-normalized log₂ counts per million (CPM) at promoters of all detected genes (TSS +/- 3 kb, n=23'087) with (K) or without (L) Tal treatment. Pearson correlation coefficient is indicated for the linear regression. Genes are color-coded by their absolute expression level.

Figure 7. Tal-mediated ERV de-repression causes a necroptotic and interferon-like response

- A. Relative viability of mESCs treated with *in vitro* synthesized dsRNA as indicated, 2 days after transfection (1 ug/ml medium). Error bars indicate SD of n=6.

B. Relative mRNA expression of canonical p53 targets and necroptosis markers in mESCs treated as in A.

C. Same as A but in HT29 cells.

D. Relative mRNA expression of known interferon-responsive genes in HT29 cells after treatment as in A.

E. Relative mRNA expression of known interferon-responsive genes in mESCs cells after treatment as in A. *Ifi44* was not detectable. All error bars represent the SD of n=3 experiments. Asterisks indicate statistical p-values of: *:p≤0.05, **:p≤0.01, ***:p≤0.001, ****:p≤0.0001

Figure 1

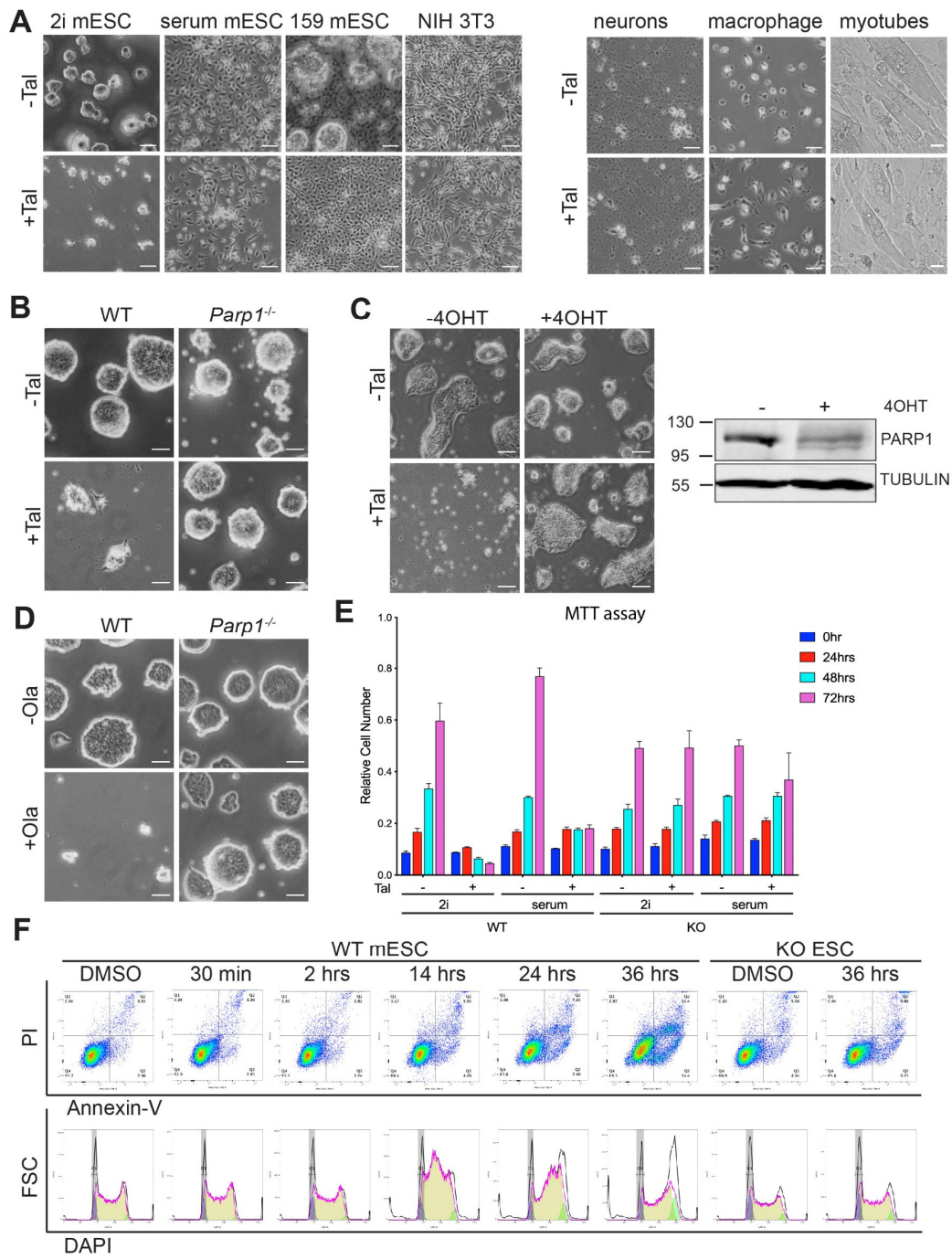


Figure 2

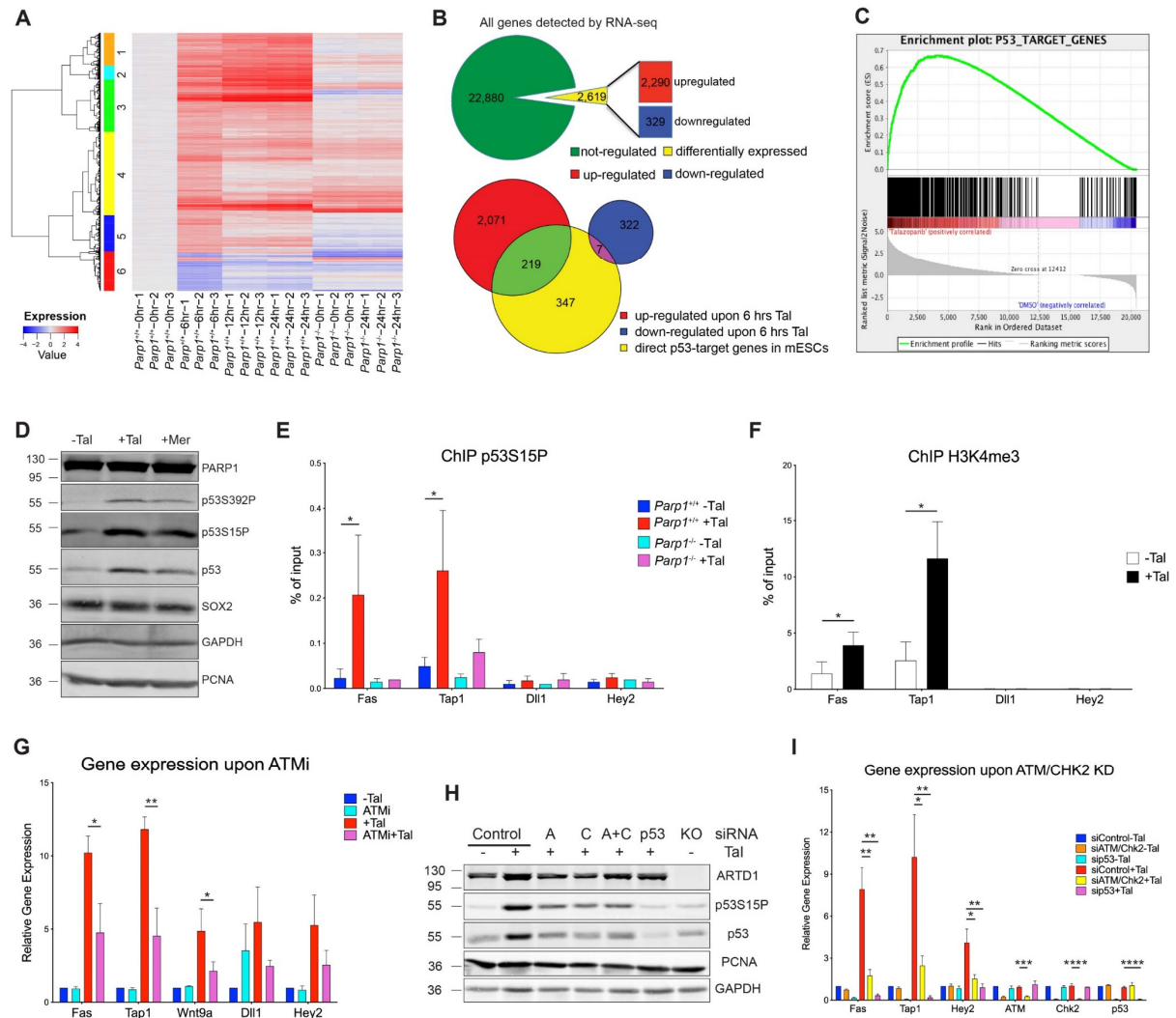


Figure 3

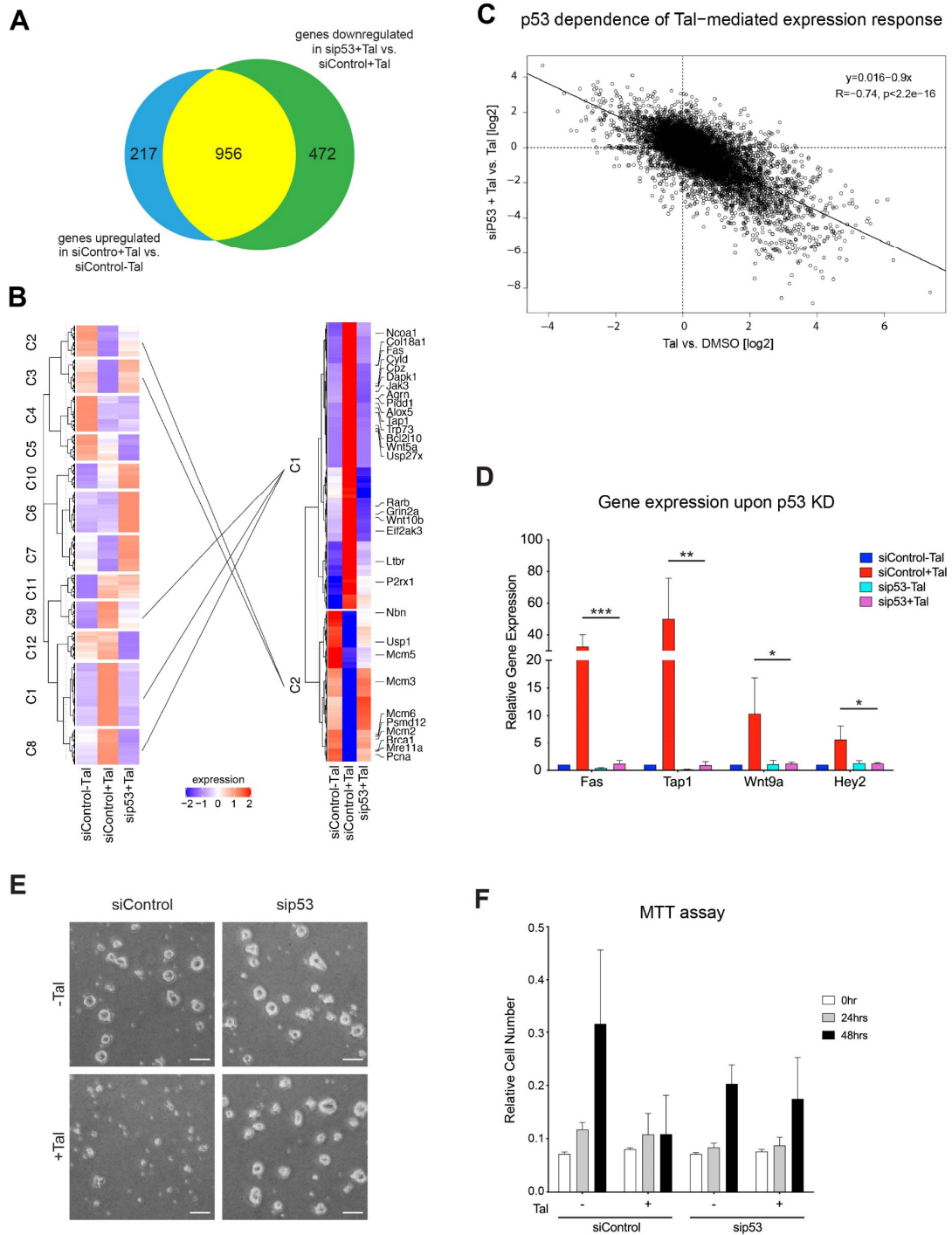


Figure 4

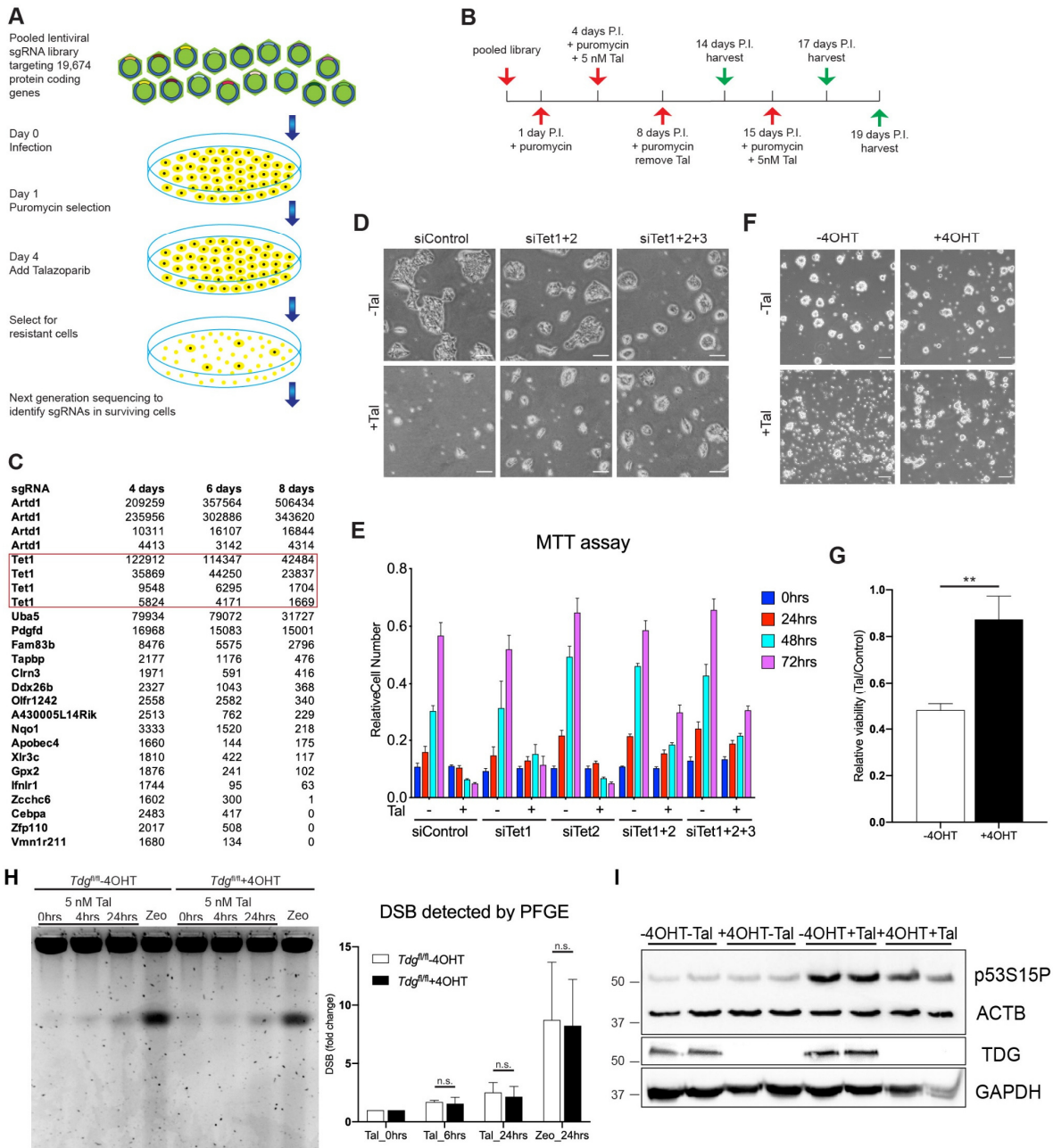


Figure 5

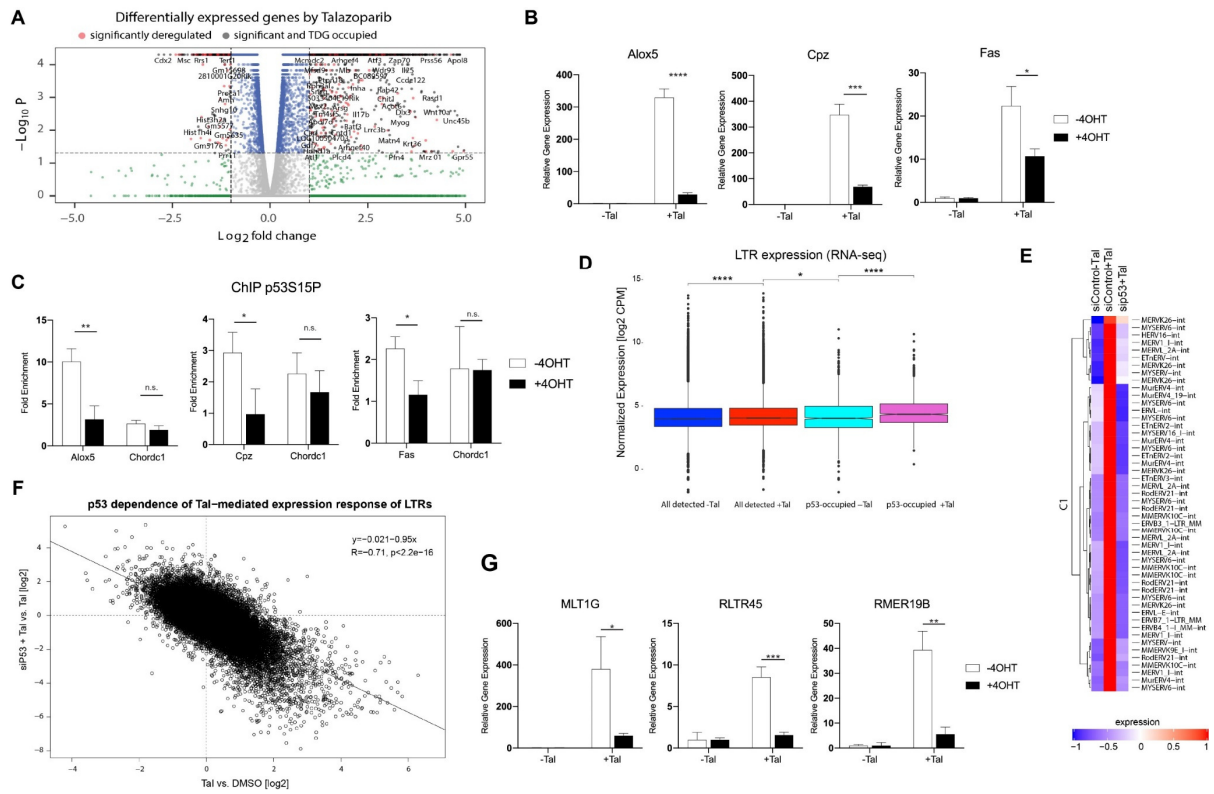


Figure 6

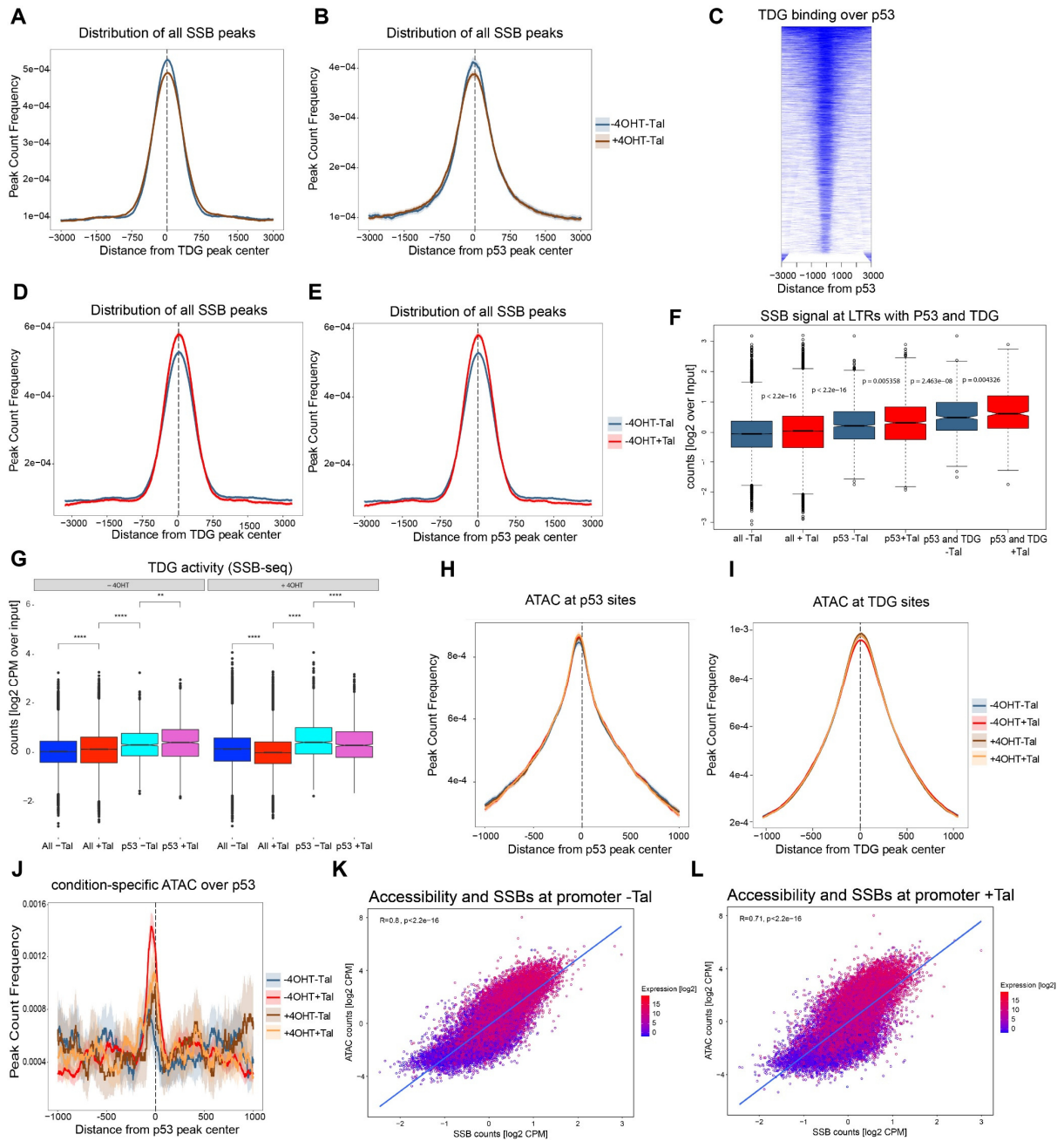
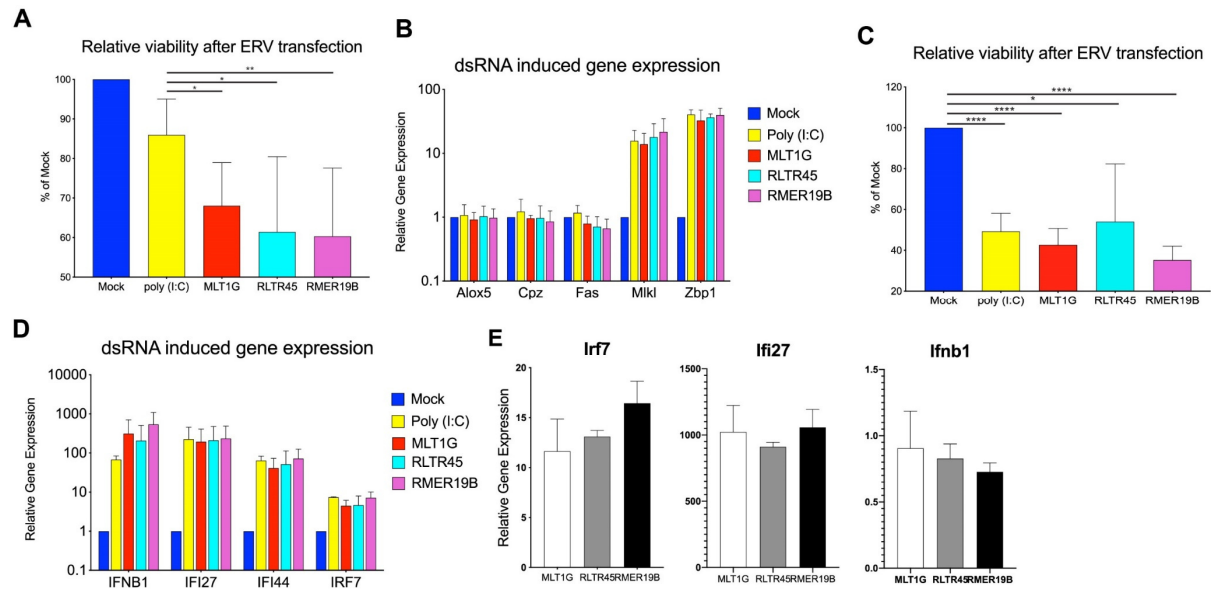


Figure 7



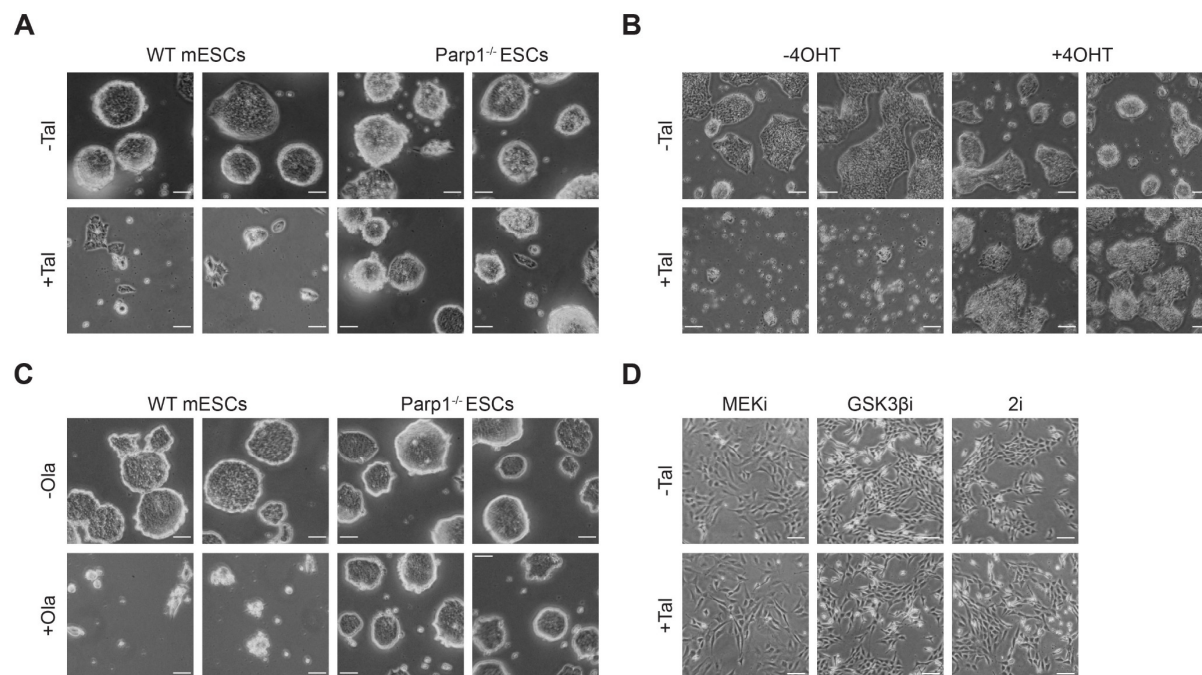


Figure S1. Related to Figure 1; High sensitivity to PARPi is specific to naïve mESCs

A. Representative pictures show the sensitivity of mESCs to Tal. WT and *Parp1*^{-/-} mESCs were treated with 5 nM Tal for 2 days. Scale bar represents 20 μ m.

B. Representative pictures show the sensitivity of inducible *Parp1* KO mESCs to Tal. *Parp1*^{fl/fl} mESCs either uninduced or induced with 200 nM 4OHT for 3 days were treated with 5 nM Tal for 2 days. Scale bar represents 50 μ m.

C. Representative pictures show the sensitivity of mESCs to Olaparib (Ola). WT and *Parp1*^{-/-} mESCs were treated with 10 μ M Ola for 3 days. Scale bar represents 20 μ m.

D. Representative pictures show NIH 3T3 cells incubated with MEKi and GSK3 β i either alone or together with 5nM Tal for 2 days. Scale bar represents 50 μ m.

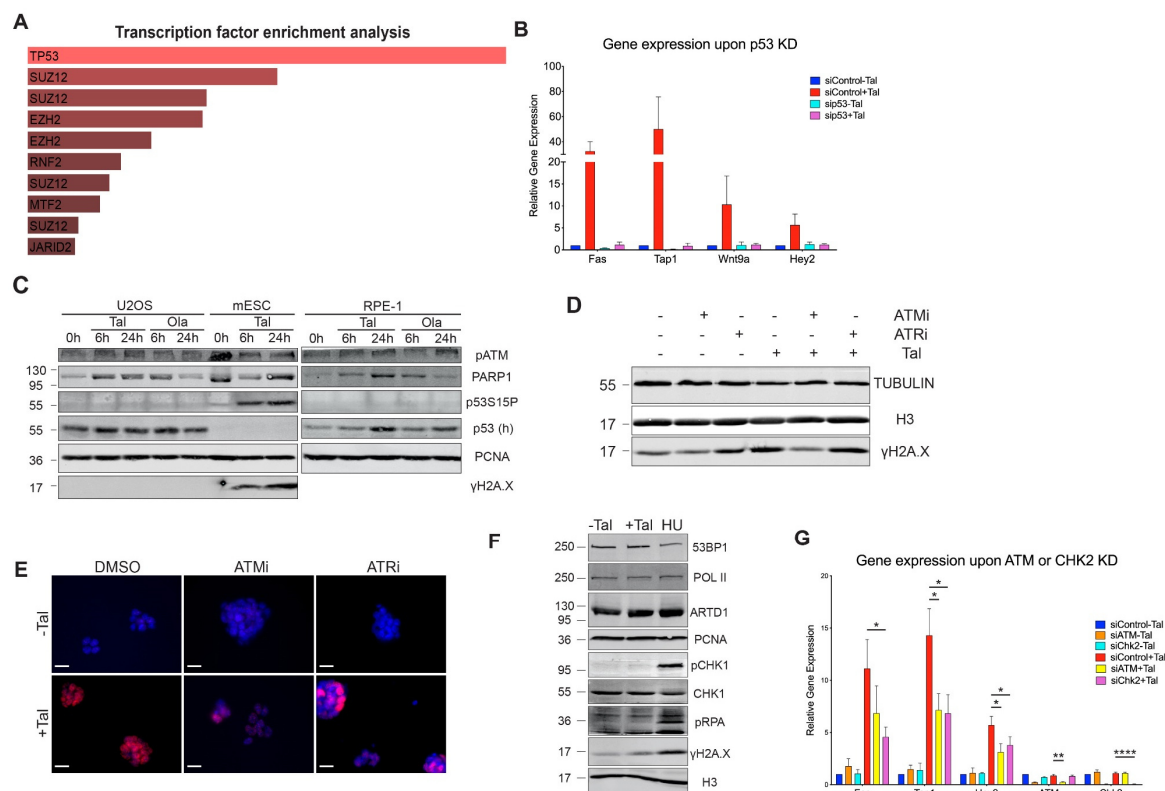


Figure S2. Related to Figure 2; Tal treatment induces p53 phosphorylation by ATM and CHK2.

A. Bar chart showing the enrichment of transcription factor binding motifs among differentially expressed genes upon Tal treatment (50 nM) after 6 hrs.

B. Relative mRNA expression by qRT-PCR analysis in WT mESCs treated with 50 nM Tal for 11 hrs upon p53 KD. Data are shown as mean + SD (n=5).

C. WB analysis of U2OS, human RPE-1 cells and WT mESCs treated with 5 nM Tal or 1 μ M Ola for indicated times (n=3).

D. WB analysis for γ H2A.X in WT mESCs treated with 5 nM Tal for 4 hrs upon ATM (10 μ M KU-55933) or ATR inhibition (1 μ M AZ20) (n=3).

E. Representative images (n=3) show immunofluorescence staining of γ H2A.X (red) and DAPI (blue) in WT mESCs treated with 5 nM Tal for 16 hrs upon ATM (10 μ M KU-55933) or ATR inhibition (1 μ M AZ20). Scale bar represents 20 μ m.

F. WB analysis of WT mESCs treated with 5 nM Tal for 4 hrs or 2 mM hydroxyurea (HU) together with 1 μ M ATR inhibitor (AZ20) for 1 hr (n=3).

G. Relative mRNA expression by qRT-PCR analysis in WT mESCs treated with 5 nM Tal for 6 hrs upon ATM or CHK2 KD. Data are shown as mean + SD (n=3). Asterisks indicate statistical p-values of: *: p \leq 0.05, **: p \leq 0.01, ****: p \leq 0.0001

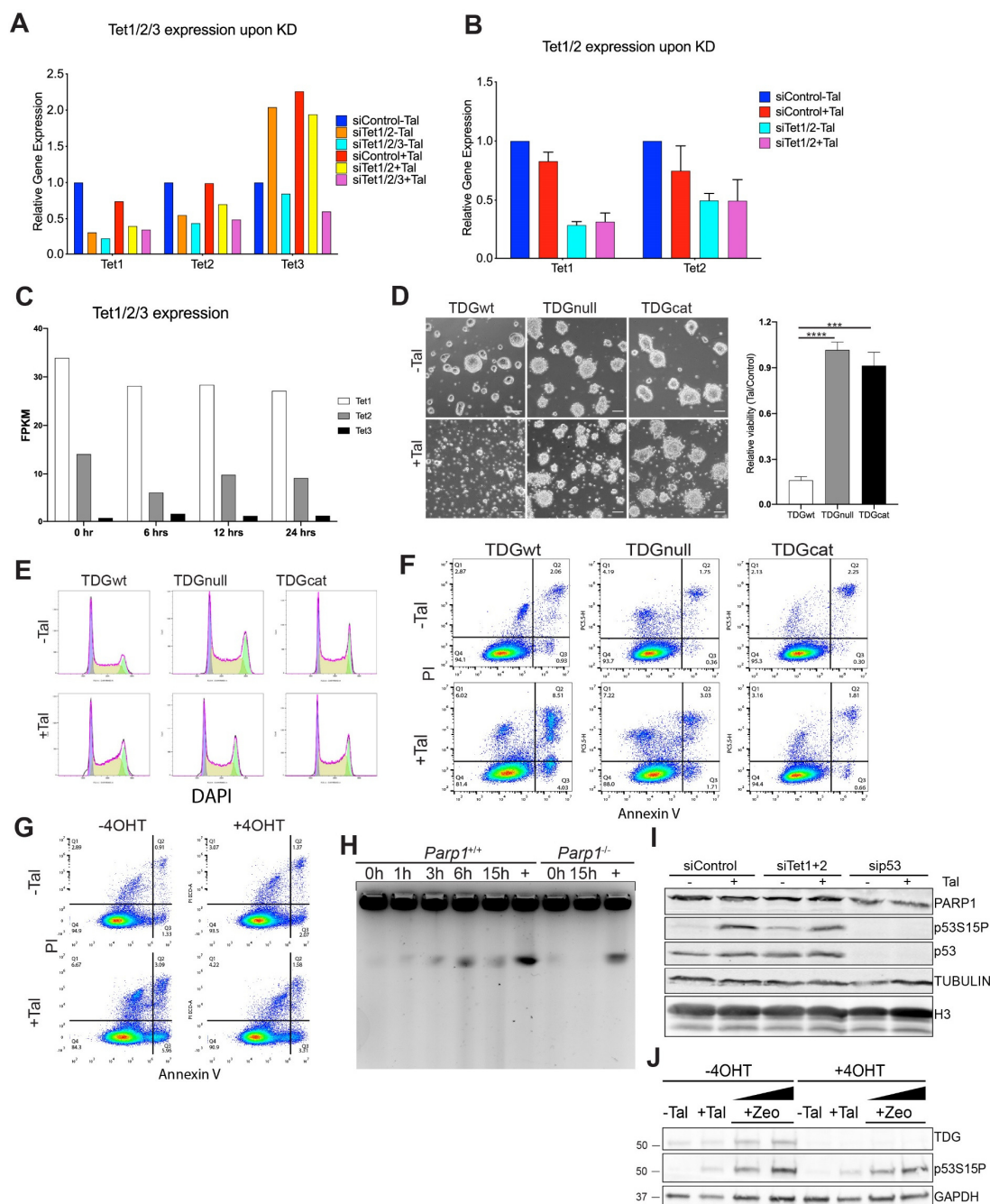


Figure S3. Related to Figure 4; TET and TDG mediated active DNA demethylation contributes to Tal induced cytotoxicity.

A. qRT-PCR analysis for KD efficiency after siTet1/2/3 transfection in WT mESCs.

B. qRT-PCR analysis for KD efficiency after siTet1/2 transfection in WT mESCs. Data are shown as mean + SD (n=3).

C. *Tet1/2/3* mRNA expression by mRNA-seq in WT mESCs treated with 5 nM for indicated time periods

D. (left) Representative pictures showing the sensitivity of TDGwt/null/cat-complemented *Tdg*^{-/-} mESCs to Tal. Cells were treated with 5 nM Tal for 2 days. Scale bar represents 40 μ m. (right), Quantification of viable cells shown as mean + SD (n=3).

E and F. Representative images (n=2) of flow cytometry showing Tal induced cell cycle changes (E) and apoptosis by PI and Annexin V (F) in TDGwt/null/cat-complemented *Tdg*^{-/-} mESCs. Cells were treated with 5 nM Tal for 24 hrs.

G. Representative images (n=2) of flow cytometry show Tal-induced apoptosis by PI and Annexin V in inducible TDG KO mESCs. *Tdg*^{fl/fl} mESCs either uninduced or induced with 2 μ M 4OHT for 2 hrs were treated with 5 nM Tal for 24 hrs.

H. PFGE of WT and *Parp1*^{-/-} mESCs treated with 5nM Tal for indicated time.

I. WB analysis of WT mESCs treated with 5 nM Tal for 6 hrs upon TET or p53 KD (n=3).

J. WB analysis of p53 phosphorylation upon 5nM Tal or 10/20 ug/ml Zeocin (Zeo) in uninduced or induced *Tdg*^{fl/fl} mESCs.

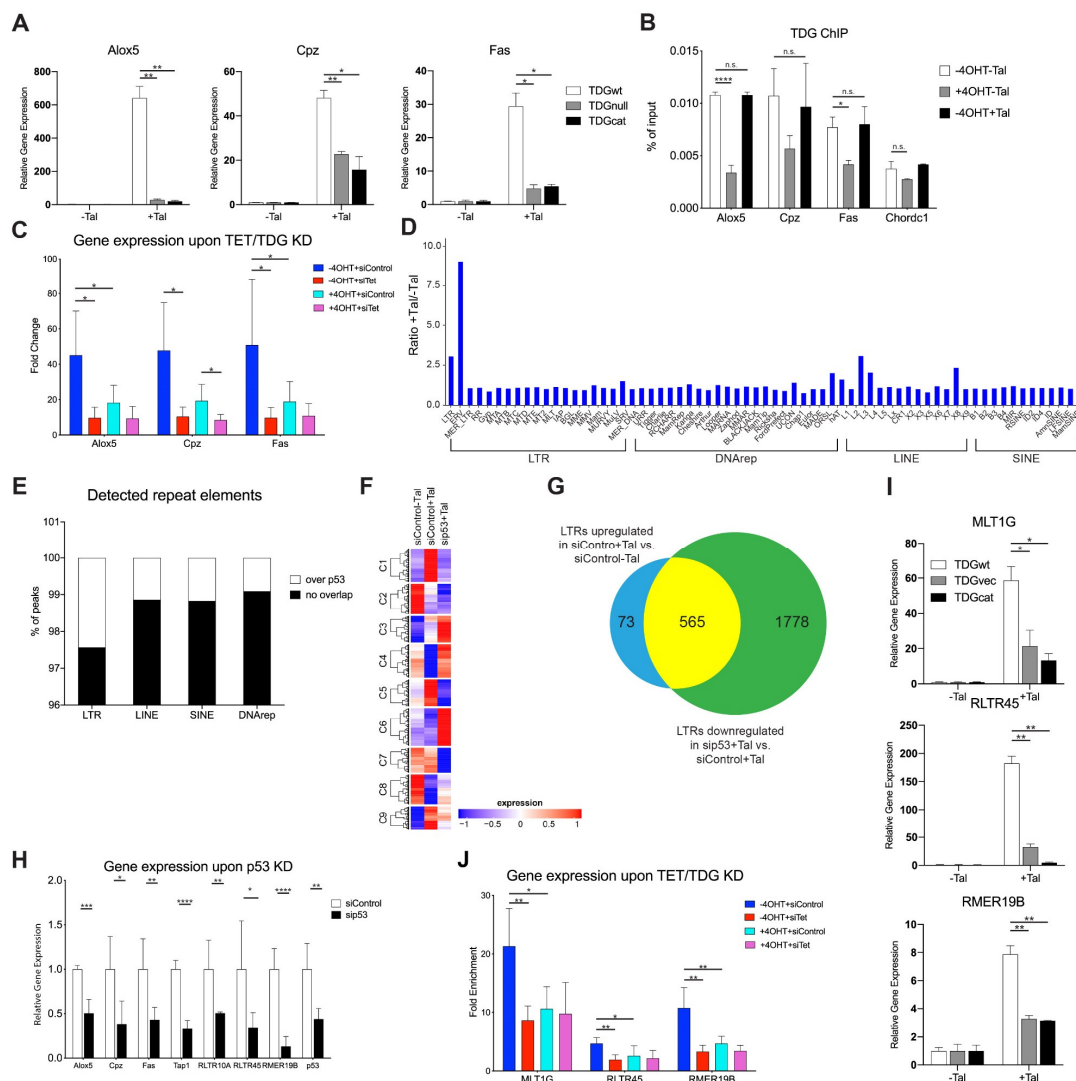


Figure S4. Related to Figure 5; TDG glycosylase activity contributes to Tal-induced gene and ERV activation.

A. Relative mRNA expression by qRT-PCR analysis in TDGwt/null/cat complemented *Tdg*^{-/-} mESCs treated with 5 nM Tal for 24 hrs. Data are shown as mean + SD (n=2).

B. ChIP-qPCR analysis of TDG in *Tdg*^{fl/fl} mESCs treated with 5 nM Tal for 24 hrs. Data are shown as mean + SD (n=3 for uninduced cells, n=2 for induced cells). A non-targeting region ~120kb upstream of *Chordc1* was used as a negative control.

C. Relative mRNA expression by qRT-PCR analysis in *Tdg*^{fl/fl} mESCs treated with 5 nM Tal for 24 hrs with or without prior KD of TET1/2 by siRNA. Data are shown as mean + SD (n=2).

D. Observed vs. expected ratios of repeat element expression in WT mESC treated with 5 nM Tal versus DMSO-treated mESCs.

E. Overview of reliably detected (min. 10 counts) repeat elements in the RNA-seq with a p53 peak within the sequence.

F. Heat map showing differentially expressed ERVs in WT mESCs treated with or without 5 nM Tal for 6 hrs and/or sip53.

G. Venn diagram showing the overlap between Tal-activated LTRs (blue circle) and LTRs whose activation by Tal was downregulated upon p53 depletion (green circle).

H. Relative mRNA expression by qRT-PCR analysis in *Tdg*^{fl/fl} mESCs treated with 5 nM Tal for 24 hrs upon p53 KD. Data are shown as mean of the relative gene expression (+Tal/-Tal) + SD (n=5).

I. Relative RNA expression by qRT-PCR analysis in TDGwt/vec/cat complemented *Tdg*^{-/-} mESCs treated with 5 nM Tal for 24 hrs. Data are shown as mean + SD (n=2).

J. Relative RNA expression by qRT-PCR analysis in *Tdg*^{fl/fl} mESCs treated with 5 nM Tal for 24 hrs with or without additional KD of TET1/2 by siRNA. Data are shown as mean + SD (n=2).

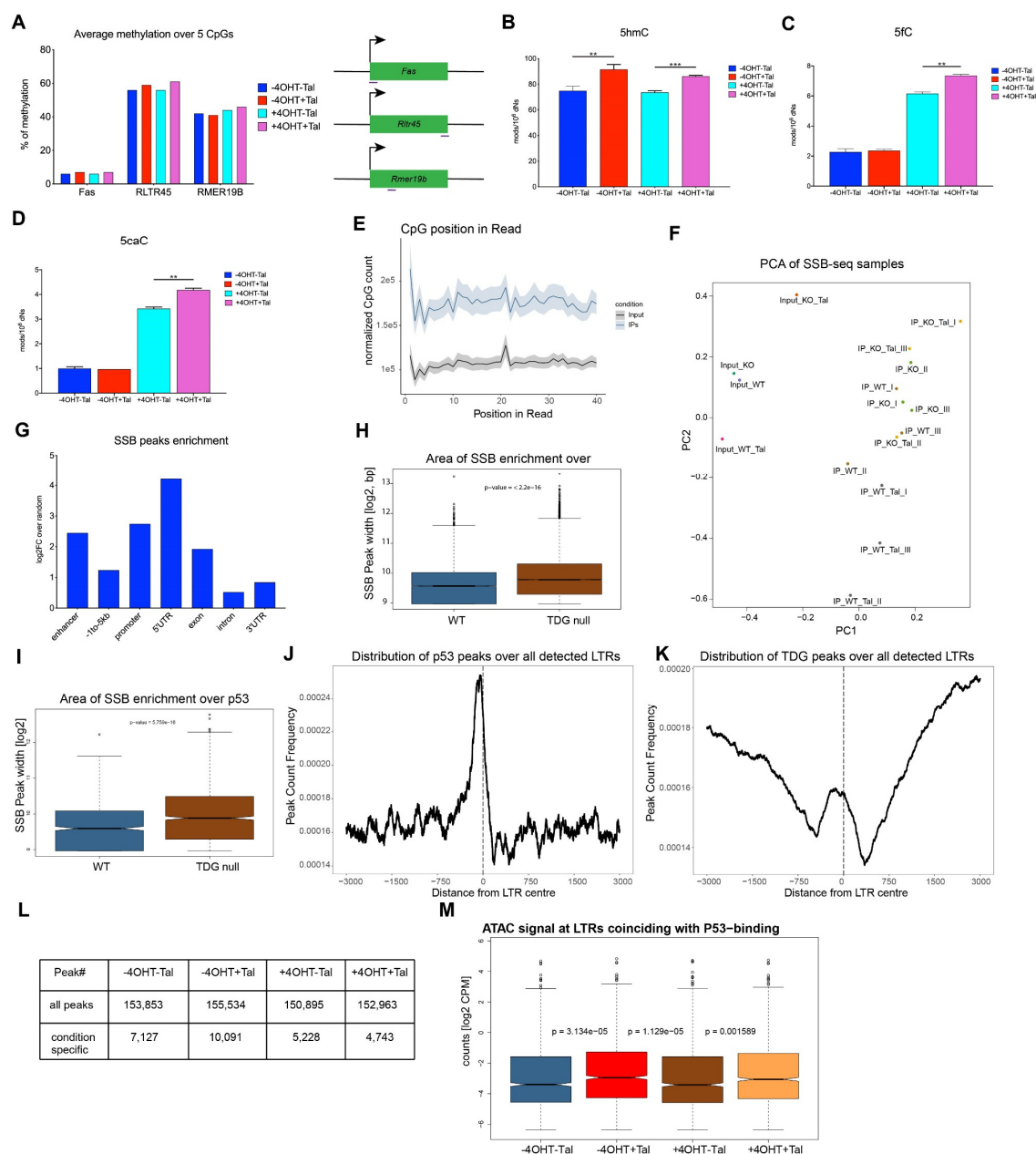


Figure S5. Related to Figure 6; active DNA demethylation-induced SSBs co-localize with p53 and ERVs.

A. mC analysis by bisulfite-assisted pyrosequencing in *Tdg^{fl/fl}* mESCs treated with 5 nM Tal for 24 hrs. Data are shown as the average methylation level (left) over 5 CpGs at indicated loci (right).

B-D. MS analysis of oxidized 5mC derivatives in *Tdg^{fl/fl}* mESCs treated with 5 nM Tal for 24 hrs. Data are shown as mean + SD (n=3).

E. Normalized CpG distribution in IP and input reads from SSB-seq. Lighter shadings indicate standard deviation from all conditions and replicates.

F. PCA analysis of SSB-seq samples.

G. Distribution of called SSB peaks over genic elements. Data is shown as log₂-fold enrichment over a random peak set.

H. and I. Size of SSB-seq peaks in uninduced and induced *Tdg^{fl/fl}* mESCs coinciding with TDG (G) or p53 (H) peaks from published ChIP-seq data in mESCs.

J. Density of called p53 peaks upon doxorubicin from published ChIP-seq data in mESCs over detected LTRs.

K. Density of TDG peaks from published ChIP-seq data in mESCs over detected LTRs.

L. Number of all called ATAC-seq peaks and those that are unique to the condition.

M. Signal intensity of normalized ATAC reads at significantly detected LTRs coinciding with p53 occupation (n=1'397).

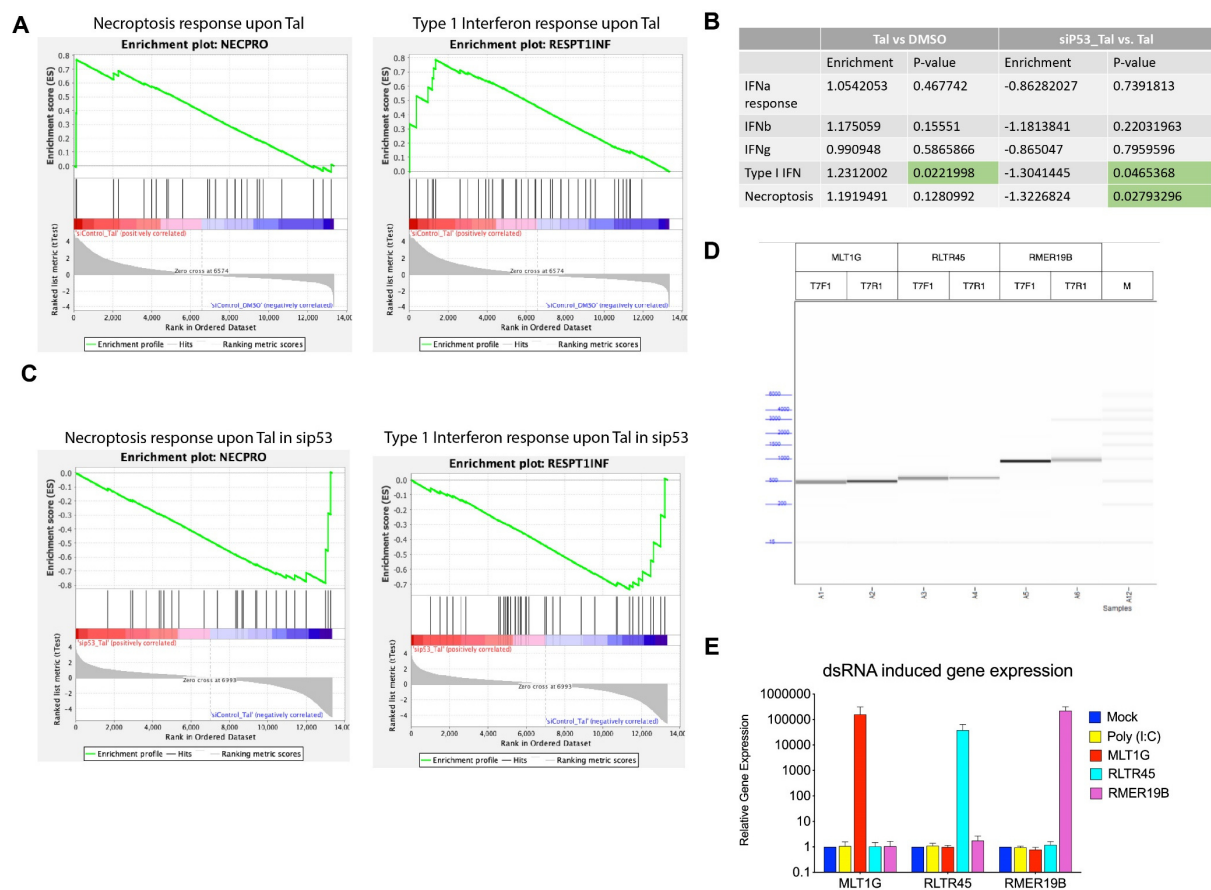


Figure S6. Related to Figure 7; Tal-induced ERV expression contributes to cell death.

A. GSEA of Necroptosis (left) and Type I Interferon response (right) genes, based on total RNA-seq in WT mESCs +/- 5 nM Tal.

B. log₂ enrichment score and p-values of GSEA of indicated pathways

C. GSEA of Necroptosis (left) and Type I Interferon response (right) genes based on total RNA-seq in WT mESCs +/- 5 nM vs. sip53-treated mESCs + 5nM Tal.

D. Fragment analysis of forward and reverse strand from *in vitro* transcribed RNA.

E. qRT-PCR analysis of transfected dsRNAs in WT. Data are shown as mean + SD (n=2).

**Title: Covalent PARylation is an integral part of TDG-BER-mediated active
DNA demethylation in embryonic stem cells**

Authors:

J. Xu^{1,2†}, S.D. Schwarz^{1†}, K. Gunasekera^{2,3}, E. Ferrari², M.O. Hottiger², P. Schär^{1*}, R.
Steinacher^{1*}

Affiliations:

¹Department of Biomedicine, University of Basel; Basel, Switzerland.

²Department of Molecular Mechanisms of Disease; University of Zurich, Switzerland.

³Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Switzerland.

*Corresponding authors. Email: *Primo.Schaer@unibas.ch, *Roland.Steinacher@unibas.ch

† These authors contributed equally to this work

Abstract:

Poly-ADP-ribosylation (PARylation) by PARP1 is implicated in DNA base excision repair (BER) but the underlying molecular mechanism is not fully understood. Here, we biochemically reconstitute the TDG-APE1-XRCC1-POL β -LIG3 complex with PARP1 and revisit the role of PARylation in DNA demethylation associated BER. PARP1 senses BER-induced DNA single strand-breaks (SSBs) and covalently PARylates each protein of the BER complex (BERosome). Notably, PARylation does not facilitate recruitment of BER factors to SSBs, but instead triggers dissociation of the PARylated BERosome from DNA, thereby promoting active DNA demethylation. Consistently, inhibition of PARylation in mouse embryonic stem cells (mESCs) results in reduced locus specific DNA demethylation activity. These results uncover an

overlooked role of PARylation to trigger the release of the BERosome from DNA to ensure epigenetic stability in mESCs.

One-Sentence Summary: PARylation of the DNA demethylation BERosome promotes DNA demethylation activity and repair complex removal from DNA

Main Text:

Poly(ADP-ribose) polymerase 1 (PARP1), also referred to as ADP-ribosyltransferase with diphtheria toxin homology (ARTD1), was suggested to be an important mediator of DNA base excision repair (BER), DNA demethylation and epigenetic reprogramming during early mammalian development (1, 2). BER mediated active DNA demethylation involves ten-eleven translocation (TET) proteins that oxidize 5-methylcytosine (5mC) stepwise to 5-formylcytosine (5fC) and 5-carboxylcytosine (caC), both substrates for the thymine DNA glycosylase (TDG). fC/caC excision by TDG generates abasic sites (AP-sites) that are incised by apurinic/apyrimidinic endonuclease 1 (APE1) to produce a DNA single strand-break (SSB) (3-6). SSBs are sensed by PARP1, activating its poly-ADP-ribosylation (PARylation) activity (7, 8), and then repaired via a pathway involving SSB editing, DNA re-synthesis and DNA ligation (9). PARP1 and PARP2 (ARTD2) were shown to promote SSB repair (SSBR) by facilitating the recruitment of X-ray repair cross-complementing protein 1 (XRCC1) (10). XRCC1 acts as a scaffold in BER/SSBR and interacts with DNA ligase 3 (LIG3), APE1, DNA polymerase β (POL β) and bifunctional polynucleotide kinase 3'-phosphatase (PNKP) to coordinate the individual steps of SSBR (11). Efficient recruitment of XRCC1 to SSBs was shown to involve non-covalent binding of XRCC1 to both, PARP1 generated PAR at the site of the SSB and to DNA (12). This observation led to a current model that PAR chains produced by PARP1 at SSBs function like "lime-twigs", facilitating non-covalent interactions that establish DNA SSBR complexes.

The role of PARP1 in sensing DNA glycosylase-generated SSBs in the context of BER is still under debate, because such "scheduled" SSBs have also been proposed to be passed on between the repair enzymes within the BER complex (BERosome) (5, 13). Thus, in BER, PARP1 activity may be needed to re-engage XRCC1 and SSBR at AP-sites or SSBs that became uncoupled during BER (14). A potential direct role of covalent PARylation of BER factors in the repair process has

not been addressed, yet implicated from proteomic analyses that identified BER proteins as targets of PARylation by PARP1 (15). Here we investigated the molecular mechanism and biological significance of covalent PARylation in SSB and BER mediated site-specific active DNA demethylation.

PARP1-mediated covalent PARylation of BER proteins is triggered by AP-site and SSB DNA repair intermediates

PARP1 was shown to be highly expressed in mESCs and to be involved in the repair of BER oxidative DNA demethylation associated SSBs during early development (1, 2). In line with this, immunoblotting revealed significantly higher PARP1 protein levels in whole-cell extracts of mESCs when compared to mouse embryonic fibroblasts (MEFs), but the PAR signal was very low (fig. S1A). It has been previously demonstrated that H₂O₂ induced oxidative DNA base damage increased BER-associated SSB formation, which in turn activates PARP1 PARylation (16, 17). Consistently, intracellular PAR levels increased in mESCs after H₂O₂ treatment, which was not apparent in the presence of the PARP1 inhibitor talazoparib (18) (fig. S1B). PAR-resin pulldown, followed by proteomic analysis revealed high amounts of covalently PARylated XRCC1 and LIG3 in pluripotent mESCs compared to more differentiated MEFs (19), supporting the notion that covalent BER protein PARylation may participate in maintenance of pluripotency, potentially by regulating BER mediated DNA demethylation. To investigate the mechanism underlying XRCC1 and LIG3 protein PARylation, we first tested their potential interaction with PARP1 (Fig. 1A). “Far-Western” immunoblotting failed to detect robust interactions of PARP1 with XRCC1 and LIG3. Auto-PARylated PARP1, however, physically bound to XRCC1 and LIG3, which is in line with previous observations of non-covalent interactions between the BER scaffold XRCC1 and LIG3 with PAR and auto-PARylated PARP1 (12, 20). No interaction of PARP1 or auto-PARylated

PARP1 with POL β , APE1 or BSA was detected. Although all proteins of the core BER machinery were shown to be PARylated in HeLa cells upon DNA damage induction (15), the molecular transactions that trigger their covalent PARylation have not been addressed. Since PARP1 binding to AP-sites (or SSBs) was shown to activate its auto-PARylation activity (21), we tested whether AP-sites may also trigger PARP1-mediated PARylation of BER enzymes. To this end we prepared a BER AP-site intermediate by digestion of a 60-mer guanine•uracil (G•U) mismatched substrate with uracil DNA glycosylase (UDG) and incubated it with either XRCC1, LIG3, the XRCC1-LIG3 complex, POL β or APE1 with PARP1 and NAD⁺. This caused protein migration shifts of all BER proteins in denaturing SDS-polyacrylamide gel electrophoresis (PAGE) (Fig. 1B, fig. S2A). The higher molecular weight and more dispersed shifts of XRCC1 and LIG3 indicate clear PARylation, the smaller and more discrete shifts observed for POL β and APE1 could be caused by PARylation as well, although mono-ADP-ribosylation (MARylation) cannot be excluded. None of the BER proteins was PARylated when the experiment was repeated with homoduplex DNA, AP-site-containing DNA substrate together PARP inhibitor PJ34 (22), or the catalytic inactive PARP1 (PARP1dead) (Fig. 1B). This demonstrates that covalent PARP1 mediated PARylation of BER proteins is triggered by AP-sites. Next, we assessed PARylation of BER proteins in a reconstituted uracil repair assay. Co-incubation of a G•U DNA substrate with purified UDG, APE1, POL β , XRCC1, LIG3 generates BER intermediates, which are then repaired by the SSBR complex (Fig. 1C). PARP1 strongly PARylated XRCC1, POL β , APE1 and LIG3 to a more modest degree. When we repeated the experiment with homoduplex DNA, PARylation of XRCC1 was reduced and no PARylation of APE1, POL β was observed. Also, no PARylation of BER proteins was observed in the presence of the PARP inhibitor PJ34 or in the absence of DNA (Fig. 1C). Overall, these results demonstrated that covalent PARylation of XRCC1, APE1 and POL β by PARP1, is triggered by the presence of AP-site and SSB BER intermediates.

PARylation of BER proteins results in removal of the SSB complex from DNA

Next, we tested whether PAR synthesis may affect the interactions of PARP1 and BER enzymes with AP-sites or SSBs. To test AP-site DNA binding, a double biotin end-labelled G•U substrate immobilized on streptavidin beads was incubated with UDG, XRCC1, POL β , and PARP1 or PARP1dead and NAD⁺ (150 μ M). Input proteins (I), unbound proteins (F) and bound proteins (E) were separated by SDS-PAGE and analyzed by immunoblotting. PAR synthesis was observed with PARP1, but not with catalytic inactive PARP1 (PARP1dead) (fig. S2B). While PARP1 eluted mainly in the unbound fraction as the auto-PARylated form, unmodified PARP1dead eluted in the streptavidin-bound fraction with the AP-site DNA. The major fraction of XRCC1 was PARylated by PARP1 and failed to bind to AP-site DNA. In contrast, in the presence of PARP1dead, XRCC1 was not PARylated and remained bound to the AP-site DNA. Likewise, POL β was PARylated by PARP1 and failed to bind DNA, whereas PARP1dead did not PARylate POL β , which remained bound to the AP-site DNA (fig. S2B). To test if PARylation mediated the dissociation of the entire BERosome, we assessed the concerted binding of XRCC1, POL β , APE1 and LIG3 to biotinylated SSB DNA heteroduplexes. The XRCC1/POL β /LIG3 complex was detected under physiological (150 mM NaCl) salt concentrations in the streptavidin DNA precipitate when PARP1 but no NAD⁺ was present (Fig. 1D). APE1 binding to SSB DNA was observed only when salt concentrations were kept low (20 mM NaCl). Performing the same experiment in the presence of NAD⁺ resulted in PARylation and release of PARylated XRCC1, PARylated LIG3, PARylated APE1 and PARylated POL β from the SSB DNA duplexes (Fig. 1D and E). When the experiment was performed under very low salt conditions (5 mM NaCl) PARylated XRCC1, PARylated LIG3, PARylated APE1 and PARylated POL β were detected in the streptavidin SSB DNA precipitate (fig. S2C). This demonstrated that PARylated BER proteins were able to bind DNA but the

covalently attachment of PAR chains reduced their DNA affinity and caused their dissociation from DNA at physiological salt concentrations. To investigate whether the reduced DNA affinity of PARylated BER proteins affects their enzymatic activity, we reconstituted a full SSB repair processes starting with an AP-site. We first quantified AP-site cleavage activity of APE1, using a fluormESCent-labeled AP-site substrate generated by digestion of a 60-mer G•U-containing heteroduplex with UDG. PARylated APE1 (50% PARylated) and unmodified APE1 (200 fmol) both fully cleaved AP-sites (5 pmol) (fig. S2D). Next, we incubated POL β and PARylated POL β (50% PARylated) with an AP-site substrate to test its AP lyase activity (fig. S2F) (23). POL β and PARylated POL β cleaved the AP-site substrate with equal efficiency (fig. S2E). Then, we reconstituted the entire BER process. AP-site incision by APE1 produces a nicked AP-site containing a lesion with 3'-hydroxyl and 5'-dRP ends. POL β then incorporates one nucleotide at the 3'-OH and trims the 5'-dRP residue to generate a ligatable substrate for the LIG3 (fig. S2F). APE1 and PARylated APE1 (20 fmol and 200 fmol) both efficiently cleaved the AP-site. POL β (0.1 pmol) incorporated one nucleotide yielding a 24-mer labeled DNA fragment (fig. S2F) but incomplete ligation (~40%) by LIG3. Addition of 1 pmol POL β , however, yielded $\geq 87\%$ of ligatable product DNA, presumably reflecting the rate-limiting dRP-lyase activity of POL β . Although PARylated POL β (~50% PARylated, 0.1 pmol and 1 pmol) also efficiently incorporated one nucleotide, ligation efficiency was slightly reduced compared to the reaction without PARylated POL β (fig. S2F). Finally, we reconstituted BER with PARylated XRCC1(~50% PARylated), POL β (~50% PARylation), LIG3(~10% PARylated). Repair was significantly reduced (-26%; difference of means 26 +/- 4.6%) compared to the unmodified XRCC1/POL β /LIG3 (Fig. 1F). This established that PARylation of APE1, POL β , XRCC1 and LIG3 did not affect their individual enzymatic activities but reduced their cooperative processing due to decreased DNA binding affinity.

Modeling of ADP-ribosylated BER protein-DNA co-crystals

To understand how covalent PARylation of SSBR proteins modulated their interactions with DNA, we grafted MAR and PAR chains to the mapped PARylation acceptor residues of human APE1, POL β , XRCC1 and LIG3 (15) using the Rosetta Comparative Modeling approach. Possible changes in the overall protein conformation induced by MARylation or PARylation were not considered in the modeling. The available APE1/AP-site DNA co-crystal structure (24) is N-terminally truncated, missing the PARylation acceptor site (Ser26) (15). We thus modeled the full-length APE1 which yielded a fold, matching that of the truncated protein (>95%). *In silico* grafting of MAR or PAR chains to Ser26 in the flexible N-terminal redox domain generated a protrusion that may weaken the interaction of APE1 with DNA. MAR or PAR chains modeled to Cys138 (15) protruded from the globular APE1 endonuclease domain (Fig. 2A, fig. S3A and B) (25). POL β consists of two main domains, an N-terminal lyase domain (8-kDa) that mediates DNA contact and removes the 5'-deoxyribose-phosphate (dRP) residue to trim the SSB 3' end for DNA ligation, and a C-terminal (31 kDa) DNA dRP polymerase domain (26) (Fig. 2B). Modeling MAR and PAR to the POL β -DNA co-crystal structure (27), revealed that ADP-ribosyl residues attached to Ser30 and/or Ser44 (15) protruded from the N-terminal domain and may interfere with the function of the helix-hairpin-helix DNA binding motifs of POL β (Fig. 2B; fig. S3C)(26). The model also predicts that POL β PARylation does not affect its interaction with the N-terminal domain of XRCC1 (fig. S3D) (28). Grafting PAR to the mapped acceptor residues (15) near to DNA binding domain DBD of the LIG3-DNA co-crystal (29), implicated that PARylation interfered with DNA binding (29)(fig. S3E). Overall, it seems plausible that after covalent PARylation of SSBR proteins, the attached PAR chains sterically clash with the DNA, which facilitates the dissociation of BER proteins.

PARylation of the TDG-BERosome facilitates active DNA demethylation

Human TDG was reported to be PARylated at Ser85 in its N-terminal domain (15). The TDG N-terminus cooperates with the catalytic core domain in DNA substrate an AP-site binding, and thereby enzymatic turnover and substrate preference of the glycosylase (6, 30). “Far-Western” blotting revealed a robust interaction of full length TDG and C-terminally truncated TDG (TDG Δ C) with PARylated PARP1 but not with unmodified PARP1. This interaction strictly required both, the N-terminus and the catalytic CORE domains of TDG, as neither of these domains showed an interaction with PARylated PARP1 (Fig. 3A). We then tested PARylation of TDG by PARP1 *in vitro* (31, 32). PARP1 PARylated TDG at both the N-terminus and CORE domain (Fig. 3B; fig.S4A). In a base release assay, full-length TDG excised a caC opposite guanine from an oligonucleotide duplex substrate but due to its high AP-site binding activity, was product-inhibited (product/enzyme ratio ≤ 1) (6, 30) (Fig. 3C). PARylation by active PARP1 stimulated caC excision activity of TDG in comparison to reactions with PARP1 dead or no PARP1 (Fig. 3C). As the stimulation beyond a product/enzyme ratio of 1:1 suggested an increased AP-site dissociation of TDG, we assessed whether PARylation affects the concerted binding of TDG and BER proteins to AP-site-containing DNA duplexes. TDG, XRCC1, POL β , APE1 and LIG3 were bound to the AP-site-containing substrate DNA and subsequently PARylated by PARP1. In the presence of PARP1 and NAD⁺, TDG enrichment on the DNA was significantly reduced (5-fold) (Fig. 3D), with PARylated TDG being mainly in the unbound fraction. Exactly the same was observed for all other BER factors in the reaction. These results show that PARylation reduced the DNA binding stability of TDG, XRCC1, LIG3, APE1 and POL β facilitating the dissociation of the BERosome from the DNA. To assess the effect of PARylation on TDG BER-mediated DNA demethylation, we reconstituted the TDG BER process with a DNA oligonucleotide substrate with

a *HpaII* recognition site (CCGG) in a hemi-carboxymethylated configuration (CcaCGG, Fig. 3E) (6). *HpaII* is methylation-sensitive and cannot cleave the CcaCGG sequence (Fig. 3E lanes 7 and 8), but caC repair by TDG BER results in the restoration of a cleavable CCGG site (Fig. 3E, lanes 3-6). caC repair to unmodified Cs by the reconstituted BER reaction with recombinant TDG, APE1, POL β , XRCC1 was significantly higher (>1.8 fold, SD ± 0.3) in presence of active PARP1 compared to PARP1dead (Fig. 3E, lane 5 and 6, fig. S4B). Altogether, this shows that PARylation promotes BERosome disassembly, allowing for enzymatic turnover and caC repair, yet the weakened DNA binding of the PARylated BER enzymes will prevent reformation of the TDG-BERosome, which explains the relatively mild overall stimulation of BER demethylation activity.

PARylation promotes TDG-dependent, DNA demethylation-associated SSBs and the repair of “unscheduled” SSBs in embryonic stem cells

To address the biological function of PARylation in BER-mediated DNA demethylation and general SSB, we measured genome-wide SSB formation in wild type or TDG-depleted mESCs (TDGko) (33), treated or not treated with the PARP1 inhibitor talazoparib. Following *in situ* nick-translation in the presence of digoxigenin-modified nucleotides and subsequent immuno precipitation and high-throughput sequencing (SSB-seq) (34), we identified $>28'000$ genomic locations with significant enrichment of SSBs in the IP over corresponding input in wild-type mESCs (fig. S5A). The distribution of these SSBs was non-random with a pattern showing a strong preference for gene regulatory regions and genic regions (promoters $>16\%$, introns $>28\%$, exons $>15\%$) compared to intergenic regions ($\sim 11\%$) which is in line with previous observations in human cells (34) (fig. S5B). The accumulation of caC DNA demethylation intermediates in TDGko mESCs was used to identify sites of TET/TDG demethylation action (35). Intersection of these caC loci with SSB-enriched genomic regions identified 6'877 DNA demethylation-

associated SSB loci in pluripotent mESCs (fig. S5A). SSB-enriched regions in TDG-proficient mESCs centered around caC excision loci in TDG-deficient mESCs. Following talazoparib treatment of cells, caC-centered SSBs were proportionally reduced and shifted to more distal (>-800 bp and +800) locations (Fig. 4A). Likewise, SSB regions accumulated next to CpG loci previously identified as sites of high TET activity and dynamic DNA methylation change in mESCs and this coincidence was also reduced following talazoparib treatment (Fig. 4B) (36). This showed that PARP inhibition leads to a genome-wide loss of SSB enrichment at sites of targeted TET-TDG-BER mediated mC oxidation and caC repair. Following Cre-mediated TDG excision or PARP1 inhibition by talazoparib treatment, the SSB signal within these regions of overlapping SSB enrichment and caC excision loci, was significantly reduced. PARP1 inhibition by talazoparib in TDGko mESCs did not further reduce the SSB signal at the TDG demethylation sites. These results indicated that PARP1 functions at these sites to promote TDG-dependent SSB formation and, hence, BER (Fig. 4C). Inhibition of PARP1 by talazoparib effected a loss of SSB signal at promoters with TDG caC excision (Fig. 4D). This points towards a role of PARP1 in stimulating TET-TDG-BER-mediated DNA demethylation activity at gene regulatory elements in mESCs. Genes with TDG-associated SSBs in their promoters showed higher transcriptional activity than genes without SSBs or TDG-dependent caC excision in their promoters (Fig. 4E). We therefore reason that promoters with ongoing TET-TDG-dependent active DNA demethylation marked by SSBs reflect an activated gene expression state in mESCs. PARP1 is well known to promote the repair of “unscheduled” SSBs generated in cells during the repair of spontaneously occurring and randomly distributed DNA base damage or topoisomerase activity (37-39). Unlike in regions of active DNA demethylation, where SSB signals decreased following PARP1 inhibition (Fig. 4A, B, C, D), SSB signals increased upon talazoparib treatment at intra- and intergenic simple repeat elements as well as in regions with a high GC-skew outside of regulatory regions, which were

previously associated with increased chromosomal rearrangements and genetic instability (Fig. 4F, fig. S5C). Overall, this demonstrated that PARP1 activity promotes TDG activity resulting in BER-associated “programmed” SSBs and confirms earlier observations that PARP1’s PARylation activity is critical to for the repair of “unscheduled” SSBs at potential fragile sites in the genome of mESCs.

Discussion

It was proposed that PARP1-mediated PARylation, whilst not essential for base excision repair (BER), re-engages the BER machinery to “unscheduled” DNA SSBs to facilitate repair (14). But this model was never tested in a biochemical reconstitution experiment. Here we show that in addition to its reported role in assembling the BER machinery at the site of DNA damage, PARP1 mediated PARylation is an integral component of BER process itself. In this context, SSB BER intermediates induced covalent PARylation of the BER factors themselves, rather than auto-PARylation of PARP1 seeming particularly important. We demonstrate that PARylation of BER factors is not required for their enzymatic activities, but to modulate molecular interactions in the BER process. PARP1 senses AP-sites and SSBs produced during BER and then covalently attaches poly ADP-ribose chains to all key proteins of the BER complex resulting in cleansing the repair factors of the repaired DNA. This also applies during TET-TDG-dependent active DNA demethylation, where PARylation of the TDG, XRCC1, APE1, POL β and LIG3 reduces their DNA affinity and causes dissociation of the BERosome (Fig. 4G). The reduced DNA affinity of the PARylated enzymes will prevent re-formation of a BERosome, implicating that PAR has to be removed from the BER enzymes. Indeed, PARylation is a reversible modification that can be cleaved off the target proteins by hydrolases (e.g. PARG), regenerating the unmodified acceptor protein (40). It is thus plausible that upon disassembly of the BERosome the ribose is removed by

PARG to recycle the BER proteins, which enables re-formation of a BERosome for another BER reaction.

Consistent with a function of PARylation to promote TDG-BER activity *in vitro*, we show that PARP1 is required for the formation of TDG-dependent BER intermediates, detectable as SSBs in living mESCs. These SSBs do not occur randomly but show a strong preference for specific genomic loci such as gene promoters where TET proteins are hyperactive, DNA demethylation is ongoing and transcriptional activity is high. Notably, the pattern of these non-random SSBs change when PARP1 PARylation is inhibited in mESCs, consistent with a direct engagement of PARP1 in TET-TDG-dependent DNA demethylation. Yet, oxidative stress-induced SSBs were also shown to trigger PARylation of XRCC1 and LIG3 in mESCs (15), suggesting that PARylation is a general principle of BER, including global BER of “unscheduled” oxidative base damage as well as targeted BER by TET-generated 5fC and 5caC DNA demethylation intermediates. Consistent with a function in global BER, we show that in unchallenged mESCs PARP1 activity prevents SSB accumulation at sites including simple repeats (e.g. microsatellites) and regions with GC skews. Notably, such loci require efficient BER to eliminate spontaneously arising base damage and SSBs to avoid engagement of genome-destabilizing recombinatorial repair processes (41, 42). We therefore propose that PARP1 has a dual role in promoting (epi)genetic stability. Beyond PARP1’s known role to promote PARylation and recruitment of DNA repair factors to “unscheduled” DNA breaks via non-covalent protein-PAR interactions, PARP1 engages in “programmed” SSBR in the context of BER and covalent attachment of PAR chains to BER proteins triggers the release of the whole BERosome from DNA upon completion of the repair process.

References and Notes

1. M. Wossidlo, J. Arand, V. Sebastiano, K. Lepikhov, M. Boiani, R. Reinhardt, H. Scholer, J. Walter, Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J* **29**, 1877-1888 (2010).
2. S. J. Roper, S. Chrysanthou, C. E. Senner, A. Sienerth, S. Gnan, A. Murray, M. Masutani, P. Latos, M. Hemberger, ADP-ribosyltransferases Parp1 and Parp7 safeguard pluripotency of ES cells. *Nucleic Acids Res* **42**, 8914-8927 (2014).
3. R. M. Kohli, Y. Zhang, TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472-479 (2013).
4. T. R. Waters, P. Gallinari, J. Jiricny, P. F. Swann, Human thymine DNA glycosylase binds to apurinic sites in DNA but is displaced by human apurinic endonuclease 1. *J Biol Chem* **274**, 67-74 (1999).
5. R. Steinacher, Z. Barekati, P. Botev, A. Kusnierczyk, G. Slupphaug, P. Schar, SUMOylation coordinates BERosome assembly in active DNA demethylation during cell differentiation. *EMBO J* **38**, (2019).
6. A. R. Weber, C. Krawczyk, A. B. Robertson, A. Kusnierczyk, C. B. Vagbo, D. Schuermann, A. Klungland, P. Schar, Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism. *Nat Commun* **7**, 10806 (2016).
7. M. S. Satoh, T. Lindahl, Role of poly(ADP-ribose) formation in DNA repair. *Nature* **356**, 356-358 (1992).
8. A. Ray Chaudhuri, A. Nussenzweig, The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nat Rev Mol Cell Biol* **18**, 610-621 (2017).
9. K. W. Caldecott, DNA single-strand break repair. *Exp Cell Res* **329**, 2-8 (2014).

10. H. Hanzlikova, W. Gittens, K. Krejcikova, Z. Zeng, K. W. Caldecott, Overlapping roles for PARP1 and PARP2 in the recruitment of endogenous XRCC1 and PNKP into oxidized chromatin. *Nucleic Acids Res* **45**, 2546-2557 (2017).
11. K. W. Caldecott, C. K. McKeown, J. D. Tucker, S. Ljungquist, L. H. Thompson, An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. *Mol Cell Biol* **14**, 68-76 (1994).
12. L. M. Polo, Y. Xu, P. Hornyak, F. Garces, Z. Zeng, R. Hailstone, S. J. Matthews, K. W. Caldecott, A. W. Oliver, L. H. Pearl, Efficient Single-Strand Break Repair Requires Binding to Both Poly(ADP-Ribose) and DNA by the Central BRCT Domain of XRCC1. *Cell Rep* **26**, 573-581 e575 (2019).
13. S. H. Wilson, T. A. Kunkel, Passing the baton in base excision repair. *Nat Struct Biol* **7**, 176-178 (2000).
14. K. W. Caldecott, Mammalian DNA base excision repair: Dancing in the moonlight. *DNA Repair (Amst)* **93**, 102921 (2020).
15. I. A. Hendriks, S. C. Larsen, M. L. Nielsen, An Advanced Strategy for Comprehensive Profiling of ADP-ribosylation Sites Using Mass Spectrometry-based Proteomics. *Mol Cell Proteomics* **18**, 1010-1026 (2019).
16. R. Wang, C. Li, P. Qiao, Y. Xue, X. Zheng, H. Chen, X. Zeng, W. Liu, I. Boldogh, X. Ba, OGG1-initiated base excision repair exacerbates oxidative stress-induced parthanatos. *Cell Death Dis* **9**, 628 (2018).
17. M. O. Bradley, K. W. Kohn, X-ray induced DNA double strand break production and repair in mammalian cells as measured by neutral filter elution. *Nucleic Acids Res* **7**, 793-804 (1979).

18. Y. Shen, F. L. Rehman, Y. Feng, J. Boshuizen, I. Bajrami, R. Elliott, B. Wang, C. J. Lord, L. E. Post, A. Ashworth, BMN 673, a novel and highly potent PARP1/2 inhibitor for the treatment of human cancers with DNA repair deficiency. *Clin Cancer Res* **19**, 5003-5015 (2013).
19. S. H. Chiou, B. H. Jiang, Y. L. Yu, S. J. Chou, P. H. Tsai, W. C. Chang, L. K. Chen, L. H. Chen, Y. Chien, G. Y. Chiou, Poly(ADP-ribose) polymerase 1 regulates nuclear reprogramming and promotes iPSC generation without c-Myc. *J Exp Med* **210**, 85-98 (2013).
20. J. B. Leppard, Z. Dong, Z. B. Mackey, A. E. Tomkinson, Physical and functional interaction between DNA ligase IIIalpha and poly(ADP-Ribose) polymerase 1 in DNA single-strand break repair. *Mol Cell Biol* **23**, 5919-5927 (2003).
21. S. N. Khodyreva, R. Prasad, E. S. Ilina, M. V. Sukhanova, M. M. Kutuzov, Y. Liu, E. W. Hou, S. H. Wilson, O. I. Lavrik, Apurinic/aprimidinic (AP) site recognition by the 5'-dRP/AP lyase in poly(ADP-ribose) polymerase-1 (PARP-1). *Proc Natl Acad Sci U S A* **107**, 22090-22095 (2010).
22. S. H. Huang, M. Xiong, X. P. Chen, Z. Y. Xiao, Y. F. Zhao, Z. Y. Huang, PJ34, an inhibitor of PARP-1, suppresses cell growth and enhances the suppressive effects of cisplatin in liver cancer cells. *Oncol Rep* **20**, 567-572 (2008).
23. S. Hasan, N. El-Andaloussi, U. Hardeland, P. O. Hassa, C. Burki, R. Imhof, P. Schar, M. O. Hottiger, Acetylation regulates the DNA end-trimming activity of DNA polymerase beta. *Mol Cell* **10**, 1213-1222 (2002).
24. C. D. Mol, T. Izumi, S. Mitra, J. A. Tainer, DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature* **403**, 451-456 (2000).

25. A. M. Whitaker, B. D. Freudenthal, APE1: A skilled nucleic acid surgeon. *DNA Repair (Amst)* **71**, 93-100 (2018).
26. W. A. Beard, S. H. Wilson, Structure and mechanism of DNA polymerase Beta. *Chem Rev* **106**, 361-382 (2006).
27. C. Liptak, M. M. Mahmoud, B. E. Eckenroth, M. V. Moreno, K. East, K. S. Alnajjar, J. Huang, J. B. Towle-Weicksel, S. Doublet, J. P. Loria, J. B. Sweasy, I260Q DNA polymerase beta highlights precatalytic conformational rearrangements critical for fidelity. *Nucleic Acids Res* **46**, 10740-10756 (2018).
28. M. J. Cuneo, R. E. London, Oxidation state of the XRCC1 N-terminal domain regulates DNA polymerase beta binding affinity. *Proc Natl Acad Sci U S A* **107**, 6805-6810 (2010).
29. E. Cotner-Gohara, I. K. Kim, M. Hammel, J. A. Tainer, A. E. Tomkinson, T. Ellenberger, Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states. *Biochemistry* **49**, 6165-6176 (2010).
30. R. Steinacher, P. Schar, Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr Biol* **15**, 616-623 (2005).
31. I. Gibbs-Seymour, P. Fontana, J. G. M. Rack, I. Ahel, HPF1/C4orf27 Is a PARP-1-Interacting Protein that Regulates PARP-1 ADP-Ribosylation Activity. *Mol Cell* **62**, 432-442 (2016).
32. D. M. Leslie Pedrioli, M. Leutert, V. Bilan, K. Nowak, K. Gunasekera, E. Ferrari, R. Imhof, L. Malmstrom, M. O. Hottiger, Comprehensive ADP-ribosylome analysis identifies tyrosine as an ADP-ribose acceptor site. *EMBO Rep* **19**, (2018).
33. S. D. Schwarz, E. Grundbacher, A. M. Hrovat, J. Xu, A. Kusnierczyk, C. B. Vagbo, P. Schar, D. Schuermann, Inducible TDG knockout models to study epigenetic regulation. *Fl000Res* **9**, 1112 (2020).

34. L. Baranello, F. Kouzine, D. Wojtowicz, K. Cui, T. M. Przytycka, K. Zhao, D. Levens, DNA break mapping reveals topoisomerase II activity genome-wide. *Int J Mol Sci* **15**, 13111-13122 (2014).
35. L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H. L. Fung, K. Zhang, Y. Zhang, Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692-706 (2013).
36. P. A. Ginno, D. Gaidatzis, A. Feldmann, L. Hoerner, D. Imanci, L. Burger, F. Zilbermann, A. Peters, F. Edenhofer, S. A. Smallwood, A. R. Krebs, D. Schubeler, A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nat Commun* **11**, 2680 (2020).
37. K. W. Caldecott, Single-strand break repair and genetic disease. *Nat Rev Genet* **9**, 619-631 (2008).
38. T. Lindahl, D. E. Barnes, Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol* **65**, 127-133 (2000).
39. K. W. Caldecott, Protein ADP-ribosylation and the cellular response to DNA strand breaks. *DNA Repair (Amst)* **19**, 108-113 (2014).
40. J. O'Sullivan, M. Tedim Ferreira, J. P. Gagne, A. K. Sharma, M. J. Hendzel, J. Y. Masson, G. G. Poirier, Emerging roles of eraser enzymes in the dynamic control of protein ADP-ribosylation. *Nat Commun* **10**, 1182 (2019).
41. W. Xu, H. Xu, K. Li, Y. Fan, Y. Liu, X. Yang, Q. Sun, The R-loop is a common chromatin feature of the Arabidopsis genome. *Nat Plants* **3**, 704-714 (2017).
42. W. Gan, Z. Guan, J. Liu, T. Gui, K. Shen, J. L. Manley, X. Li, R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev* **25**, 2041-2056 (2011).

43. S. R. Hartono, I. F. Korf, F. Chedin, GC skew is a conserved property of unmethylated CpG island promoters across vertebrates. *Nucleic Acids Res* **43**, 9729-9741 (2015).
44. R. Martello, M. Leutert, S. Jungmichel, V. Bilan, S. C. Larsen, C. Young, M. O. Hottiger, M. L. Nielsen, Proteome-wide identification of the endogenous ADP-ribosylome of mammalian cells and tissue. *Nat Commun* **7**, 12917 (2016).

Acknowledgments:

We thank Dr. Florian Geier, DBM Bioinformatics Core Facility Department of Biomedicine, University of Basel for his support with the analysis of the SSB-seq.

Funding:

The work was supported by the Swiss National Science Foundation (SNSF_138153, 156467) and the Walter Honegger Foundation, Zurich.

Author contributions:

Conceptualization: RS

Methodology: JX, SDS, KG, EF, MOH, PS, RS

Investigation: JX, KG, SDS, EF, RS

Visualization: KG, SDS, RS

Funding acquisition: RS, PS

Project administration: RS

Supervision: RS

Writing – original draft: RS

Writing – review & editing: JX, KG, SDS, EF, MOH, PS, RS

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: The SSB-seq and RNA-Seq data can be found on the gene expression omnibus (GEO) under the accession number GSE166964 or PRJNA743896, respectively.

Supplementary Materials

Materials and Methods

Figs. S1 – S5

Fig. 1. Covalent PARylation of BER proteins results in BERosome removal from DNA.

(A) Scheme of “Far-Western” blot analyses. BSA and recombinant purified human XRCC1 (1 μ g), POL β (1 μ g), APE1 (1 μ g), and LIG3 (1 μ g) were spotted on nitrocellulose (representative Ponceau-stained membrane shown) and probed with PARP1 and PARylated PARP1 (1 μ M) and immuno-stained for PARP1 as depicted. **(B)** Scheme of *in vitro* PARylation reaction.

Homoduplex DNA or AP-site-containing DNA substrate, was incubated with either purified LIG3 (upper panel), the XRCC1-LIG3 complex (upper middle panel), POL β (lower middle panel) or APE1 (lower panel) and purified wild type PARP1_{wt} or catalytic inactive PARP1_{dead} and HPF1 as indicated. BER proteins were separated on SDS-PAGE and analyzed by immunoblot analysis. **(C)** Scheme of *in vitro* PARylation of BER proteins with PARP1 in a reconstituted uracil repair reaction. Uracil DNA glycosylase (UDG) excises uracil to produce an AP-site, which is cleaved by APE1 into a DNA single-strand break (SSB) that is bound by PARP1, which triggers its PARylation activity. POL β incorporates dCTP and XRCC1-LIG3 seal the DNA strand. Biotinylated uracil-containing oligonucleotides were incubated with UDG, APE1, POL β , XRCC1 and LIG3 in equimolar amounts of purified wild type PARP1_{wt} and HPF1. dCTP, ATP and PJ34 PARP1 inhibitor concentrations as indicated. **(D)** Scheme of BER protein pulldown experiments and immunoblot analysis. BER proteins were incubated with DNA immobilized on streptavidin (star) beads and bound proteins eluted in denaturing SDS buffer.

BER proteins in input (I), flow (F), elution (E) were analyzed by immunoblot analysis (E)
 Quantification of pulled BERosome proteins. n=3. Mann-Whitney test, two tailed, ** = P<0.01.
 (F) Schematic of the UDG uracil BER reconstitution reaction. UDG uracil base excision and APE1 cleavage of the AP-site generates a SSB with a 3'-hydroxyl end and a 5'-end carrying the dRP residue, which is processed and repaired by POL β , XRCC1 LIG3. Repair reaction was performed using BER proteins or PARylated BER proteins as indicated. Percentage of total signal of fully repaired product is indicated. Quantification of relative repair efficiencies of BER proteins and PARylated BER proteins (un-PARylated BER activity defined as 100%). n=3. Mann-Whitney test, two tailed, P<0.01.

Fig. 2. Modeling of covalent MARylated APE1 and POL β .

(A) Structure of human APE1, with redox and endonuclease domains indicated. Modeled full length APE1 structure with mono-ADP-ribose (MAR) at the acceptor residues S26 and C138 (15). (B) Structure of human POL β , with 8-kDa and polymerase domains indicated. Modeled full length POL β structure with MAR molecules at the acceptor residues S30, S44, S104 (15). Protein modeling with attached MAR structures was performed using Rosetta Comparative Modeling.

Fig. 3. PARylation promotes BER-mediated DNA demethylation.

(A) Structure of human TDG, with N-terminal, catalytic core, C-terminal truncated (Δ C) and C-terminal domains. Coomassie-stained SDS-PAGE showing purified TDG and TDG domains. “Far-Western” blot analyses to test interactions of PARP1 and PARylated PARP1 with TDG proteins. BSA and recombinant and purified human full size TDG or TDG domains were dot blotted on nitrocellulose (representative Ponceau stained membrane shown) and probed with

PARP1 and PARylated PARP1 and immuno-stained for PARP1. **(B)** *In vitro* PARylation of TDG. Purified full length human TDG was incubated with PARP1 and HPF1 in the presence of [32P]-NAD⁺, double-stranded DNA oligomer and G•U-containing DNA oligonucleotides where indicated. Proteins were separated by SDS-PAGE and analyzed by Coomassie blue (upper panel) and autoradiography (lower panel). Asterisk indicates truncated TDG protein **(C)** The effect of PARP1 PARylation on TDG catalysis in a base release assay. Full size TDG incubated with uracil-containing substrate DNA. Reactions were stopped and the AP-site cleaved by addition of NaOH after indicated time and analyzed using denaturing gel electrophoresis and fluorescent scanning. Dotted line represents turnover threshold where 1 pmol substrate per 1 pmol TDG is processed. Shown are mean values with SD, n=3. Mann-Whitney test, two tailed, *P<0.05, **P<0.01. **(D)** Scheme of TDG-BERosome pulldown experiments and immunoblot analysis. Purified TDG, LIG3, XRCC1, LIG3, POL β and APE1 proteins bound to uracil-containing DNA substrate immobilized on streptavidin (star) incubated with PARP1 and HPF1 as indicated. I input, F flow, E elution. **(E)** Schematic of the reconstitution of the BER as part of DNA demethylation reaction. ds59merDNA substrates: CG/CG *HpaII* methylation-sensitive restriction site, cleavable by *HpaII*, lane 1–2; *HpaII* methylation-sensitive restriction site with a caC (caCG/CG), not cleavable by *HpaII*, lane 7–8, repaired product (CC/CG) cleavable by *HpaII*, lane 3–6. Percentage of repaired product is indicated. ss59merDNA, single-stranded DNA. Quantification of relative caC BER, n=3. Mann-Whitney test, two tailed, *P<0.05.

Fig. 4. PARylation promotes locus-specific BER-mediated DNA demethylation and prevents accumulation of “unscheduled SSB” in mESCs.

(A) Frequency of SSB-enriched regions over caC excision loci. **(B)** Frequency of SSB-enriched regions over genomic loci with above average TET activity (36) in unchallenged (DMSO) and

talazoparib treated (Tal) mESCs. **(C)** Quantification of SSB signal in regions of SSB-enrichment defined in unchallenged wt mESCs overlapping with sites of caC excision, n=6'877 (35) Values are displayed for unchallenged mESCs (TDGwt, blue), talazoparib treated mESCs (TDGwt talazoparib, red), unchallenged TDG knock out mESCs (TDGko) or talazoparib treated TDGko mESCs (TDGko talazoparib). **(D)** Normalized SSB signal in promoters (TSS -2 kb / +500 bp) exhibiting TDG-mediated caC excision (35) in TDG wild type mESCs with or without talazoparib treatment, n=2'520. **** P = 2.6×10^{-6} . **(E)** Normalized expression (log₂ RPKM) of genes without caC excision or SSB regions in the promoter (Genes, no caC, no SSB), of genes only with caC excision in their promoters (Genes caC in promoter) and of genes with caC excision and SSBs in their promoter (Genes caC and SSB in promoter). **** P = $<2.8 \times 10^{-8}$, ***** P = $<2.2 \times 10^{-16}$. **(F)** Normalized SSB signal at regions with GC skews in intergenic and intragenic regions (43) in TDG wild type mESCs untreated or treated with talazoparib. ***** P = $<2.2 \times 10^{-16}$. **(G)** PARP1 PARylation coordinates TDG/BER-mediated active DNA demethylation. The AP-site/SBB produced by TDG BER is sensed by PARP1 and triggers PARylation activity. PARP1 PARylates the BERosome, which stimulates caC-excision and consequential BER activity by dissociation of the repair complex from DNA.

Figure 1

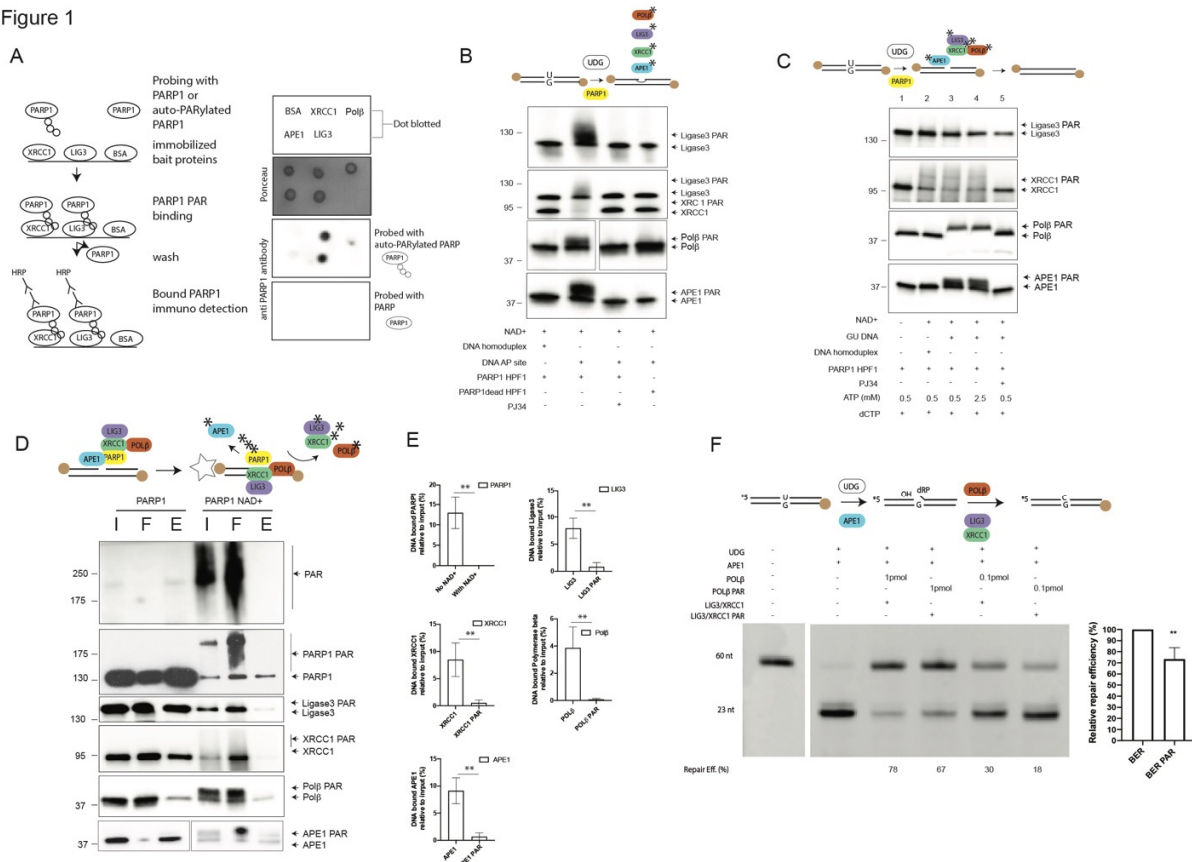


Figure 2

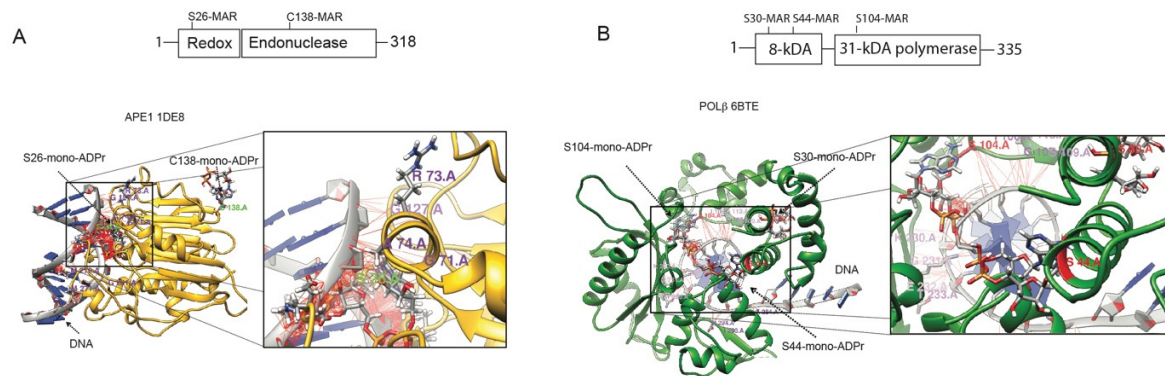


Figure 3

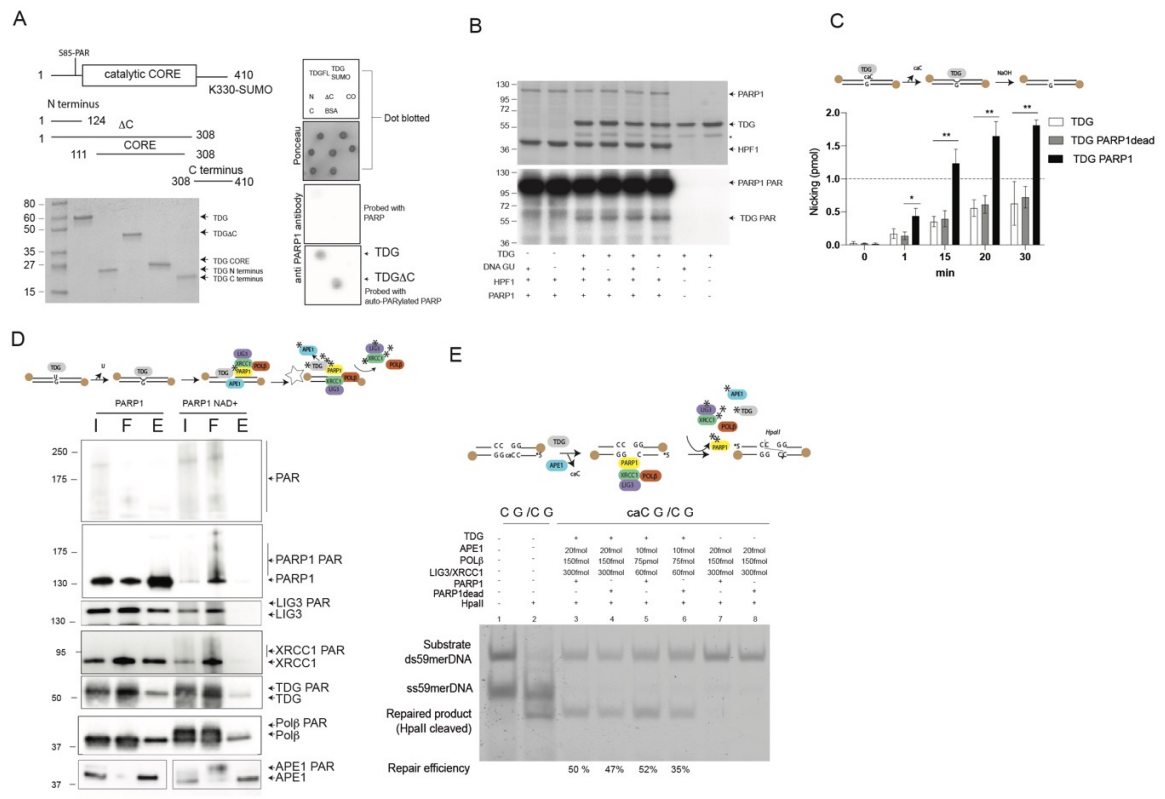
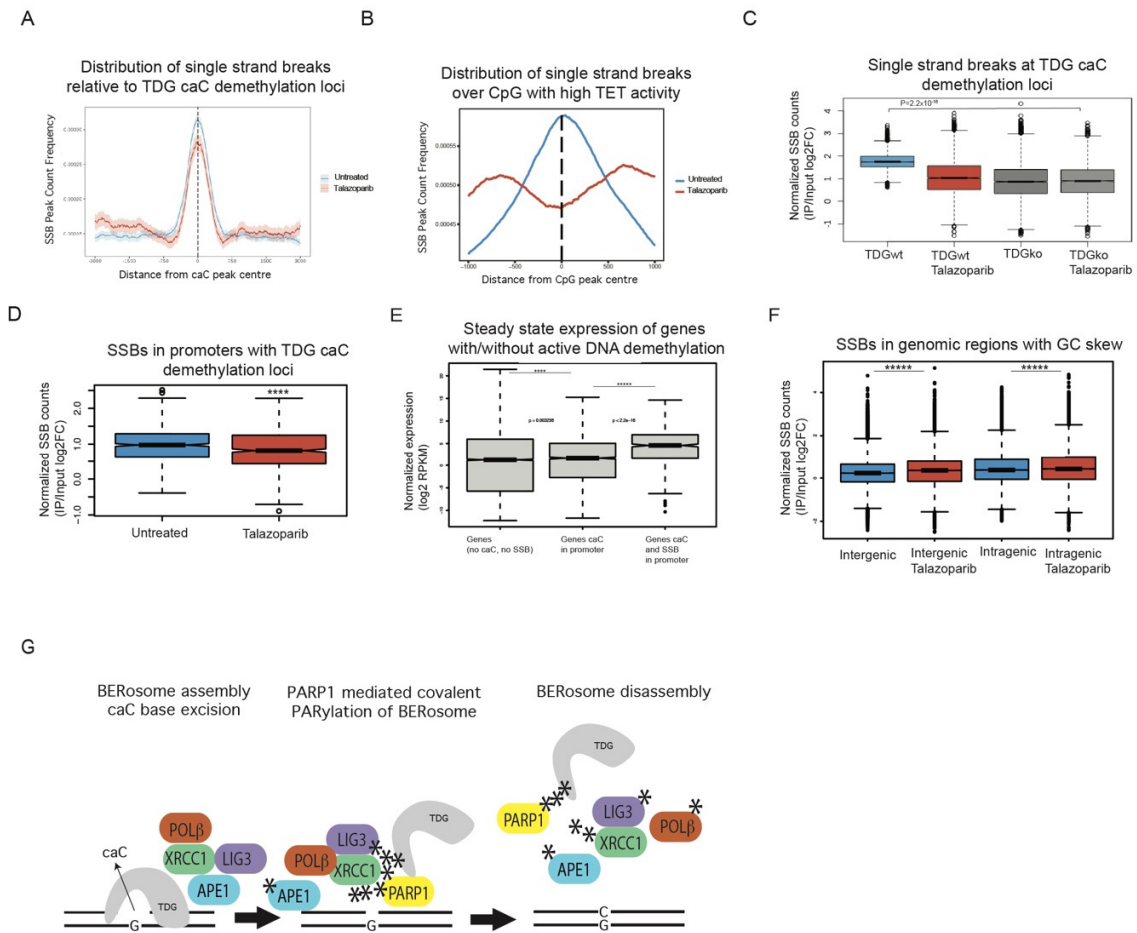


Figure 4



Supplementary Material for

Covalent PARylation is an integral part of TDG-BER-mediated active demethylation in embryonic stem cells

J. Xu^{1,2†}, S.D. Schwarz^{1†}, K. Gunasekera^{2,3}, E. Ferrari², M.O. Hottiger², P. Schär^{1*}, R. Steinacher^{1*}

Correspondence to: *Primo.Schär@unibas.ch, *Roland.Steinacher@unibas.ch

This PDF file includes:

Materials and Methods

Figs. S1 to S5

Materials and Methods:Cell culture

mESCs were grown on feeders at 37°C for two passages in mESC in a humidified atmosphere containing 5% CO₂. Then mESCs were grown without feeders for five passages in 2i medium containing LIF. mESCs were treated with DMSO or 100 nM Talazoparib for 12 hrs. Then the cells were treated with 1 mM H₂O₂/MgCl₂ for 10 minutes to induce PARylation and harvested by adding modified RIPA buffer (44). 10 μM PJ34 and 75 μM tannic acid in modified RIPA buffer prevent PARP1/PARP2 mediated ADP-ribosylation and PARG mediated hydrolysis of poly-ADP chains respectively, during the sample preparation until acetone precipitation.

Immunofluorescent staining

Cells grown on glass cover slips were treated with or without 50 nM Tal for 30 min followed by 10 min treatment with 0.5 mM H₂O₂ and fixed with methanol/acetic acid (3:1) for 5 min on ice. After fixation, cells were blocked by 5% milk and 0.05% Tween 20 in PBS. After blocking, cells were first stained with anti-PAR Ab (10H) followed by Cy3-conjugated goat anti-mouse secondary Ab (Jackson ImmunoResearch). Finally, cells were stained with Hoechst 33258 stain solution and mounted with Vectashield mounting solution.

Bioinformatic modeling of BER protein structures

Rosetta comparative modeling algorithm with multi-templates approach was used to generate a full protein structure from the truncated crystal structure of human BER proteins. The crystal structures used in the modeling approach for XRCC1 (3k77, 3qvg, 3k75), LIG3 (3l2p, 6wh1, 3pc7, 3l2p), polB (3lqc, 1bpx, 4nlm) and APE1 (1de8, 1e9n, 6w2p). Pymol and Chimera were used to add mono-ADP-ribose to the reported acceptor sites and structural alignment again with the selected crystal structures for visualization purposes and image creation. Mono- and poly-ADP-ribose chains were grafted to the *in vivo* mapped ADP-ribose acceptor sites of human BER proteins (15).

Protein interaction and “Far-Western” analysis

“Far-Western” analysis was performed by spotting purified human POLβ (0.5 μg), APE1 (0.5 μg), LIGIII (0.2 μg), XRCC1 (0.2 μg), TDG full size or TDG domains, TDG SUMO (0.5 μg) and BSA (1 μg) on nitrocellulose membrane (Protran, Amersham). PARP1 was incubated prior to probing in PARylation buffer for 5 min at 37°C (50 mM Tris pH 8.0, 0.5 mg/ml BSA, 1 mM DTT, MgCl₂ 2 mM) in presence or absence of NAD⁺ (1 mM), with homoduplex DNA (25 nM). Membranes were then incubated with purified PARP1 or PARylated PARP1 (1 μM) in incubation buffer (50 mM Tris pH 8.0, 10% glycerol, 150 mM NaCl, 0.5 mg/ml BSA, 0.1% NP-40, 1 mM DTT) at 4°C for 4 h. Unbound PARP1 protein was removed by brief washing and membranes were probed with polyclonal anti-PARP1 antibody (ab32138 Abcam (rabbit) 1:1000 dilution) in 5% non-fat dry milk dissolved in TBS (100 mM Tris-HCl pH 8, 150 mM NaCl) with 0.1% Tween-20 (Sigma) and analysed by chemiluminescence detection using PXi imaging system (Syngene).

PARylation reaction

BER proteins (0.2 μg) were PARylated separately or as a BER complex (0.2 μg each protein) on DNA (10 pmol) with PARP1wt or PARP dead (0.1 μg) and HPF1 (0.1 μg) as indicated. PARylation reaction was performed in PARylation buffer (50 mM Tris pH 8.0, 0.5 mg/ml BSA, 1 mM DTT, 2 mM MgCl₂) in presence or absence of NAD⁺ (150 μM - 1 mM), for 15 min at 37°C. AP-site substrate was generated by digestion of uracil-containing homoduplex oligonucleotides with uracil DNA glycosylase (UDG, 5 units New England BioLabs) in PARylation buffer, 5 min at 37°C. The reaction volume was 20 μl. To generate SSB substrate, the AP substrate was incubated with APE1 in PARylation buffer for 5 min at 37°C. *In vitro* PARylation of TDG with radioactive labelled NAD⁺. Purified full-length human TDG (0.4 μM) was incubated with PARP1

(0.4 μ M) and HPF1 (2 μ M) in the presence of [32 P]-NAD, double-stranded DNA oligomer (200 nM) and U-containing DNA oligonucleotides (400 nM) were indicated. Reaction was stopped with 4x SDS loading buffer 0.2 M Tris pH 8.0, 0.4 M DTT, 277 mM SDS, 4.3 M glycerol and heating at 99°C for 1 min.

PARylation repair complex dissociation assays

AP-site or SSB-containing double-stranded DNA oligonucleotide substrate, generated by digestion of an uracil-containing biotinylated duplex with UDG alone or UDG and APE1 respectively, 5 min at 37°C in PARylation buffer. 0.5 μ M of the substrate DNA was incubated with human POL β , APE1, XRCC1 or LIGIII protein (each 0.5 μ M) as well as UDG (5 units, New England BioLabs) in PARylation buffer (50 mM Tris pH 8.0, 0.5 mg/ml BSA, 1 mM DTT, MgCl₂ 2 mM) in presence or absence of NAD⁺ (1 mM), for 15 min at 37°C. Subsequently equimolar human PARP1 or catalytic inactive PARP1 (PARP1dead) and HPF1 (0.2 μ M) were added and incubated for 15 min at 30°C while shaking at 750 rpm. Pre-equilibrated 20 μ l magnetic streptavidin beads were added per reaction and incubated for 20 min at 20°C, shaking at 750 rpm. Flow through (F) was collected. After washing three times with 250 μ l high salt wash buffer (50 mM Tris pH 8.0, 1 mM DTT, 2 mM MgCl₂, 150 mM NaCl) or low salt wash buffer (50 mM Tris pH 8.0, 1 mM DTT, 2 mM MgCl₂, 20 mM NaCl) at 4°C, 10 μ l of 2 \times SDS loading dye was added and the samples denatured at 99°C. Released proteins (elution, E) were separated on 15%-4% SDS-polyacrylamide gels (Mini Protean TGX, BIO-RAD precast gels) and transferred to nitrocellulose membranes (Protran, Amersham) by electroblotting. Blots were then incubated with rabbit polyclonal anti-TDG 141 antibody (raised against recombinant full-length hTDG), dilution 1:20,000; rabbit polyclonal anti-XRCC1 antibody (Sigma-Aldrich, X0629), dilution 1:1,000; rabbit polyclonal anti-POL β antibody (Acris, AM00275PU-N), dilution 1:1,000; rabbit polyclonal anti-PAR (Trevigen 4336-BPC-100), dilution 1:1000; mouse polyclonal anti-LIG3 (Genetex (6G9) GTX70147), dilution 1:1000; mouse anti-APEX1 (Invitrogen MA1-440 (13B 8E5C2)), dilution 1:1000; rabbit anti-PARP antibody (ab32138 Abcam), dilution 1:1000. All antibodies were diluted in non-fat dry milk TBS (100 mM Tris-HCl pH 8, 150 mM NaCl), 0.1% Tween-20 (Sigma). Analysis was done by chemiluminescence detection (WesternBright ECL, Advansta) on film (FujiFilm) or PXi imaging system, Syngene.

Plasmids

Construction of the overexpression plasmids of human full-length his-tagged TDG and TDG domains, his-tagged POL β , APE1, XRCC1, LIG3 were generated by PCR cDNA amplification with adaptor oligonucleotides fused to suitable restriction sites (5, 6, 30).

Protein purification

Human BER proteins, POL β , LIG3, XRCC1, TDG, TDG domains, APE1, PARP1, PARP1dead were expressed and purified as described in (6, 30, 44). The protein expression plasmids were electroporated into E. coli BL21 (DE3) cells. Starter cultures were grown overnight and diluted with LB broth medium to an OD600 of 0.1 and grown at 30°C to an OD600 of 0.8 under selection with 100 mg/L of ampicillin. Protein expression was induced by addition of IPTG. TDG, XRCC1 and LIG3: 250 μ M IPTG, 25°C for 4 h, APE1: 500 μ M IPTG, 25°C for 6 h, POL β : 250 μ M IPTG, 25 °C for 3.5 h. After harvesting the cells by centrifugation (GSA, Sorvall, 4000 rcf, 30 min, 4°C), the pellet was resuspended in lysis buffer (50 mM Na-phosphate buffer pH 7.5, 300 mM NaCl, 20% glycerol, 0.1% Tween-20, 1 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF) and sonicated (Bioruptor, Diagenode) to obtain protein fractions. Pre-cleared 6-His tagged BER proteins (centrifugated at >30,000 rcf, 60 min, 4°C) were loaded onto a disposable 15 ml column (Bio-RAD) packed with Ni²⁺-nitrilotriacetic acid (Ni-NTA)-agarose (Qiagen) and washed with 100 ml lysis buffer containing 30 mM imidazole. Elution of bound proteins was performed

with 5 ml elution buffer (lysis buffer with 300-500 mM imidazole), and proteins dialyzed and loaded on columns using Äkta Explorer 10 (GE Healthcare) HPLC system for further purification as indicated. TDG full size was dialyzed (50 mM Na-phosphate pH 8.0, 50 mM NaCl, 20% glycerol, 1 mM DTT, 1 mM PMSF) and loaded on a 5 ml HiTrap Heparin HP column (GE Healthcare), flow rate 1 ml min⁻¹, washed with 10 ml dialysis buffer and the bound proteins eluted with a gradient of 50-800 mM NaCl in 50 ml. After dialysis (50 mM Na-phosphate pH 8.5, 20 mM NaCl, 10% glycerol, 1 mM DTT, 1 mM PMSF) proteins were loaded (flow rate of 1 ml min⁻¹) on a 1ml HiTrap Q HP (GE Healthcare) column, followed by washing (10 ml dialysis buffer). Bound proteins were eluted using a gradient of 20-500 mM NaCl in 20 ml and homogeneous protein fractions were pooled, dialyzed (50 mM Na-phosphate pH 8.0, 50 mM NaCl, 10% glycerol, 1 mM DTT, 1 mM PMSF) and snap frozen in liquid nitrogen. TDG and TDG domains were further purified by anion exchange chromatography (RESOURCE Q and S GE Healthcare; 20 mM Tris-HCl pH 8.5, 5% glycerol, 1 mM DTT, 0.1% PMSF, 5 mM NaCl) and eluted with a linear gradient of 0.005–1 M NaCl. Ni-NTA-purified APE1 and POL β were further purified by cation exchange chromatography (RESOURCE S, GE Healthcare; 25 mM Na-phosphate pH 6.9, 5% glycerol, 1 mM DTT, 0.1% PMSF, 5 mM NaCl) and eluted with a linear gradient of 0.005–1 M NaCl. Ni-NTA-purified XRCC1 was purified using Heparin (HiTrap Heparin HP; 25 mM Na-phosphate pH 7, 0.02–1.5 M NaCl, 5% glycerol, 1 mM DTT, 0.1% PMSF) and an anion exchange chromatography (RESOURCE Q; 50 mM Bicine-NaOH pH 8.8, 20% glycerol, 1 mM DTT, 0.1% PMSF, 0.025–1 M NaCl). SUMOylated TDG was generated according to (5) and the SUMO conjugates purified by affinity chromatography by 6-His tag purification over Ni²⁺-nitrilotriacetic acid (Ni-NTA)-agarose (Qiagen) and subsequently by fast protein liquid chromatography (FPLC) using a 1 ml HiTrap Q HP column (GE Healthcare) (50 mM Na-phosphate pH 8, 5% glycerol, 1 mM DTT, 0.1% PMSF, 0.005-1M NaCl). Pure homogeneous protein fractions were pooled and snap frozen for storage at -80°C.

BER reconstitution

Reconstitution of BER reactions were carried out in 20 μ l reaction volumes in reaction buffer (50 mM Tris-HCl pH 8.9, 1 mM DTT, 0.1 mg/ml BSA, 1 mM ATP, 2 mM MgCl₂, 200 μ M dCTP) with TDG (1 pmol), APE (20 fmol, 200 fmol), POL β (0.1 pmol, 1 pmol), XRCC1 (1 pmol), LIGIII (1 pmol) with DNA substrates as indicated. (2 pmol; upper strand 5'-TAGACATTGCCCTCGACGACCCGCCGCGCGCXGGCCACCCGCACCTAGACGAATCCG-3' where X = C was annealed to 1 pmol lower strand 5'-CGGAATTCGTCTAGGTGCGGGTGGCXGGCGCGGCGGGTTCGTCGAGGGCAATGTCTA-3' where X = C/caC/U) as indicated. Reactions were incubated at 37°C for 1, 15, 20, 30 min as indicated, and the DNA was ethanol precipitated O/N at -20°C. DNA was resuspended and digested with HpaII (1 U, NEB) in 1 \times CutSmart buffer (NEB) at 37°C for 1 h. Then DNA was ethanol precipitated O/N at -20°C and resuspended on glycerol loading buffer (0.5 \times TBE, 50% glycerol). After separation on an 8% native PAGE, the fluorescently labeled DNA was detected using the blue fluorescence mode of the Typhoon 9400 (GE Healthcare) or PXi imaging system, Syngene and analyzed quantitatively by Fiji ImageJ.

Single-strand break detection and sequencing

The detection of endogenous SSBs was performed based on the protocol (34). Genomic DNA was extracted with the Genomic-Tip 100/G Kit (Qiagen) from freshly treated mESCs. 50 μ g of gDNA were dissolved in a total of 145 μ l solution including 15 μ l Pol I buffer (10 \times , NEB buffer 2) and 15 μ l of nick labelling mix (10 \times : 20 μ M DIG-dUTPs, 200 μ M dATP/dCTP/dGTP, 117 μ M ddATP/ddCTP/ddGTP). The mixture was incubated with 5 μ l of *E. coli* DNA Pol I (NEB) for one minute at 16 °C, sample by sample and immediately quenched by the addition of EDTA to a final

concentration of 50 μ M to be put on ice. The labeled DNA was purified by precipitation using 2.5 M ammonium acetate and two volumes of 100% EtOH. The DNA was resuspended in H₂O and fragmented with a Bioruptor Plus (Diagenode) for 10 cycles of 30 sec on/off following a restriction digest with 20 U of each EcoRI, HindIII and XbaI overnight. The fragmented DNA was again purified using DNA clean & concentrator kit (Zymo) according to the manual. Equal amounts of precleared DNA were subjected to immunoprecipitation over night with 2 μ g of anti-DIG antibody (Roche) at 4°C and the immunocomplexes were recovered with 40 μ l of preblocked (tRNA, BSA) Protein G Dynabeads (Invitrogen) for 2 hours at 4 °C. The beads were washed once with PBS, three times with NP-40 Buffer (20 mM Tris-HCl, pH=8.0, 137 mM NaCl, 10% Glycerol, 2 mM EDTA, 1% NP-40), and twice with TE (10 mM Tris-HCl, pH=8.0, 1 mM EDTA). The immunocomplexes were eluted twice with 50 μ l of TE containing 0.5% SDS prior to digestion with 50 μ g/ml Proteinase K at 52 °C for 2 hours. The DNA was then purified with the ChIP Clean & Concentrator Kit (Zymo) and subjected to library preparation.

Library preparation and high-throughput sequencing

Libraries of Input and IP samples (10 ng DNA each) were prepared using the KAPA HyperPrep Kit (Roche) following the manufacturers protocol. Subsequent paired-end sequencing (75 cycles) was performed on an Illumina MiSeq system at the genomics facility Basel to an average depth of 50 mio reads per sample.

Bioinformatic Processing

Reads were aligned to the mouse genome (mm10 UCSC version) with bowtie2 (version 2.3.4.2) and extra options "--maxins 2000 --no-mixed --no-discordant --local --mm". Duplicates were marked with picard tools (version 2.9.2) and resulting BAM files were filtered by removing reads falling into ENCODE blacklist regions (version 2014 + manual removal of high coverage regions) using Rsamtools (R version 3.6, Bioconductor version 3.10).

Peaks were called across all replicates of a sample group using HOMER (version 4.11) and "findPeaks" with specifying commands "-style histone -o" with the corresponding input as a control. Signals of SSBs outside of significant enrichment (peaks) were counted as log₂ fold enrichment of IP over Input by using "bamCount" (bamsignals package V1.22.0) and normalization by trimmed mean of M-values according to the library size using edgeR (V3.13). Graphs were created using R-studio version 4.0.3 and the R Package "ChIPseeker" (V 1.26).

TotalRNA-seq analysis

TotalRNA-seq reads were mapped to mm10 mouse genome using STAR (version 2.6). Gene expression values were extracted using HTSeq (version 0.12.4) from BAM files (sorted by Coord) generated by STAR aligner. HTSeq output files of replicates were used as input files to compute differential gene expression using the generalized linear model implemented in the Bioconductor package DESeq2 (version 1.26.0).

Accession

The SSB-seq and RNA-Seq data can be found on the gene expression omnibus (GEO) under the accession number GSE166964 or PRJNA743896, respectively.

Figure S1

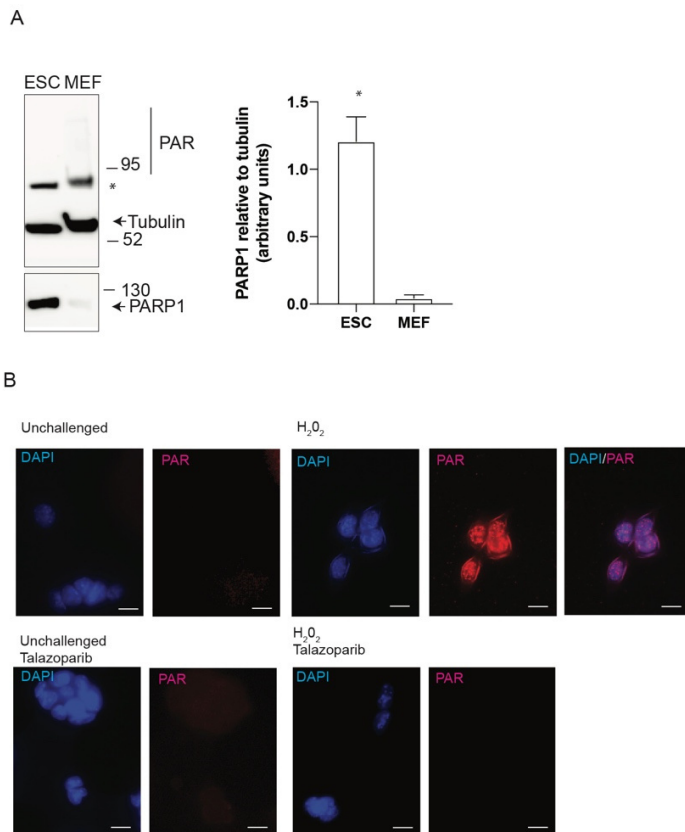


Fig. S1.
Characterization of the PARylome in mouse embryonic stem cells. (A) Immunoblot of whole-cell extracts of mouse embryonic stem cells (mESCs) and fibroblasts (MEFs) showing PARP1, poly-ADP-ribose and tubulin as loading control. Quantification of PARP1 relative to tubulin. $n = 4$. Mann-Whitney test, two tailed, $p < 0.05$. * = unidentified band. (B) Immunofluorescent staining showing DAPI and PAR signals in mESCs either unchallenged or challenged with H₂O₂ (0.5 μ M) with and without talazoparib (100 nM) treatment as indicated. Scale bar = 10 μ m.

Figure S2

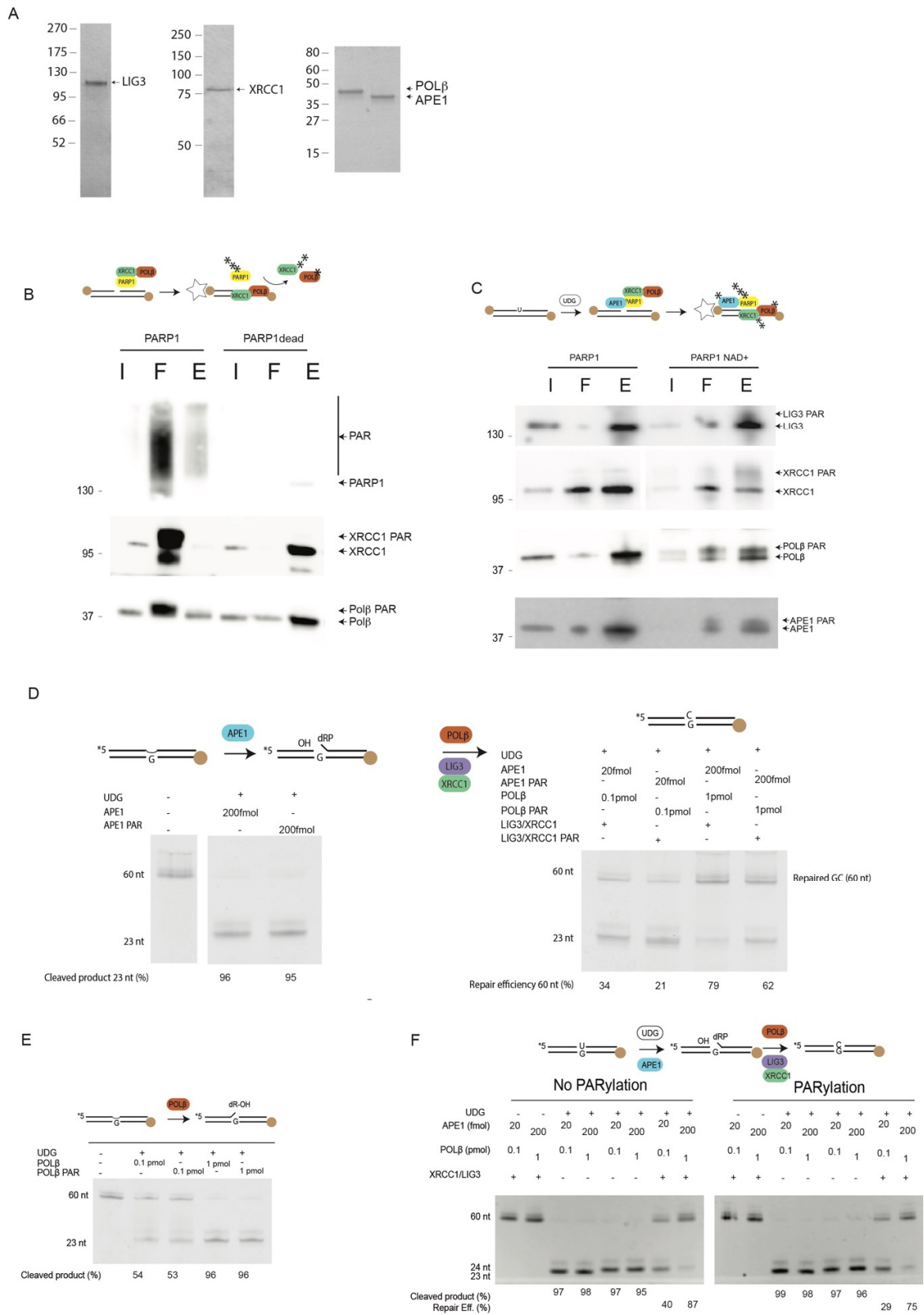


Fig. S2.

PARylated PARP1 interacts with XRCC1, LIG3 and PARP1 covalently PARylates BER proteins. **(A)** Coomassie-stained SDS-PAGE showing recombinant and purified BER proteins as indicated. **(B)** Scheme of XRCC1-POL β pulldown experiments and immunoblot analysis. Purified XRCC1, POL β proteins (0.5 μ M) bound to AP-site containing DNA (1 μ M) immobilized on streptavidin beads (star) incubated with PARP1 (0.5 μ M) and HPF1 (0.5 μ M) as indicated. I: input, F: flow through, E: elution. **(C)** Scheme of BER protein pulldown experiment and immunoblot analysis. Purified UDG, XRCC1, LIG3, POL β and APE1 proteins (0.5 μ M) bound to uracil-containing DNA substrate (1 μ M) immobilized on streptavidin beads (star), were incubated with PARP1 (0.5 μ M) and HPF1 (0.5 μ M) and the immobilized protein DNA complex washed with low salt buffer. I input, F flow, E elution. **(D)** Scheme of AP-site incision assay (left panel) and repair reaction (right panel). A 60mer G•U-containing substrate (100 nM) was digested with UDG to generate an AP-site. AP-site cleavage efficiency (cleaved product %) is shown (left). AP-site repair by APE1, APE1 PARylated, POL β , POL β PARylated, XRCC1, XRCC1 PARylated, LIG3, LIG3 PARylated is shown as repair efficiency (%). **(E)** Scheme of AP-site POL β lyase assay. AP-site cleavage efficiency (cleaved product %) by POL β , POL β PARylated is shown. **(F)** Scheme of G•U BER repair assay. Reconstitution of the G•U BER reaction with UDG, APE1, POL β and PARylated POL β , XRCC1 and LIG3. AP-site cleavage efficiency (cleaved product %) and repair efficiency (Repair eff. %) is shown.

Figure S3

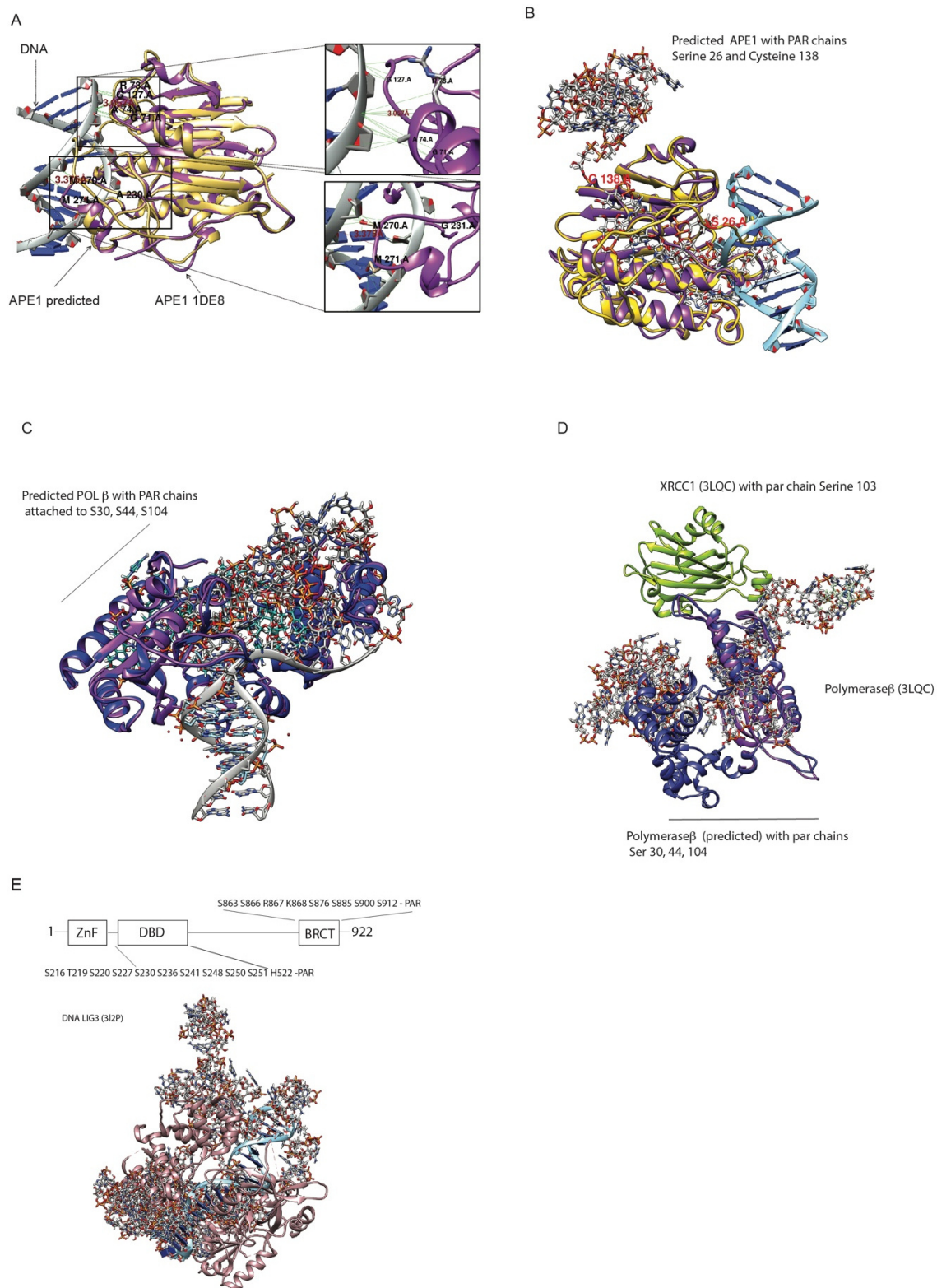
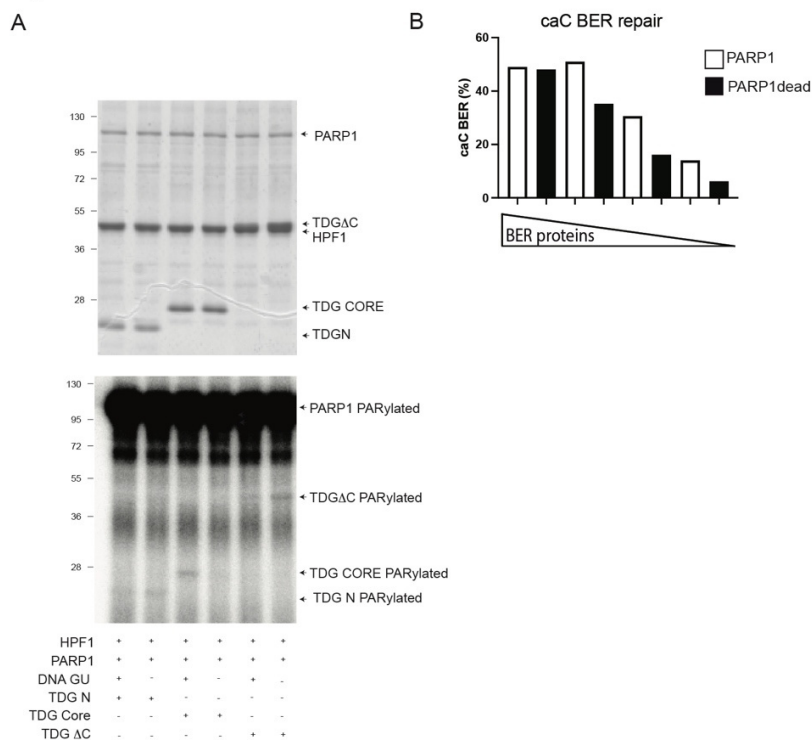


Fig. S3.

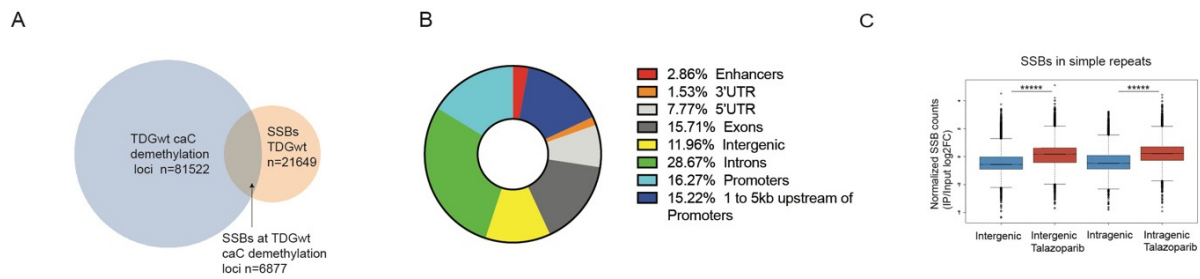
Modeling of covalently PARylated BER proteins. (A), (B) and (C) Protein structures of PARylated human APE1, POL β , showing the overlap (>95%) of the modeled and the published protein structures (1DE8, 8ICZ). Magnifications show amino acid residues involved in DNA contacts. Hydrogen bonds are shown in green. PAR chains were modeled to the APE1 and POL β structures. PARylated serines are indicated. **(D)** Modeled PARylated structure of POL β bound to XRCC1 (3LQC). **(E)** Structure of human DNA ligase 3 (LIG3) with zinc finger domain (ZnF) and the DNA binding domain (DBD) indicated. Modeled PARylated LIG3 structure (predicted). The ADP-ribosylation acceptor sites of the BER proteins were mapped in (15).

Figure S4

**Fig. S4.**

PARP1 PARylation promotes BER mediated DNA demethylation. (A) *In vitro* PARylation of TDG. Purified domains human TDG domains (0.4 μ M) were incubated with PARP1 (0.4 μ M) and HPF1 (2 μ M) in the presence of [32 P]-NAD, homoduplex DNA oligomer (200 nM) and G•U-containing DNA oligonucleotides (400 nM) where indicated. Samples were separated by SDS-PAGE and analyzed by Coomassie blue (upper panel) and autoradiography (lower panel). (B) Quantification of reconstituted caC excision and repair reaction with recombinant TDG, APE1, POL β , XRCC1 and LIG3 and PARP1 or catalytic dead PARP1 (PARP1dead).

Figure S5

**Fig. S5.**

PARylation promotes locus specific BER mediated DNA demethylation and prevents accumulation of “unscheduled” SSBs in mESCs. (A) Venn diagram of the SSBs-enriched regions coinciding with caC excision loci in unchallenged wild type (TDGwt) mESCs. **(B)** Chart with relative representation of SSB-enriched regions detected in wild type mESCs at indicated genomic locations **(C)** Normalized SSB signal at simple repeats (defined by RepeatMasker via UCSC) in intergenic and intragenic regions in wild type mESCs untreated or treated with talazoparib. **** P = $<2.2 \times 10^{-16}$

Active DNA Demethylation Evicts H2A.Z to Facilitate Pause Release of RNA Polymerase 2

Simon D Schwarz¹, Jianming Xu¹, Kapila Gunasekera², Zeinab Barekati¹, Florian Geier^{1,3}, David Schuermann,¹ Primo Schär^{1,4}

¹ Department of Biomedicine, University of Basel, Switzerland

² Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Switzerland

³ Swiss Institute of Bioinformatics, Basel, Switzerland

⁴ Corresponding author

Keywords: Active DNA demethylation, TDG, Single-Strand Breaks, H2A.Z, Pause Release

Abstract

TET/TDG-mediated active DNA demethylation is processive in gene regulatory regions as well as in gene bodies. Exactly how active DNA demethylation contributes to and modulates transcriptional processes, particularly when it occurs in gene bodies, is poorly understood at the molecular level. We used an inducible TDG KO system in mouse embryonic stem cells (mESC) and analysed nascent RNA transcription by precision run-on sequencing (PRO-seq). We show that the loss of TDG gives rise to a genome-wide increase of RNA polymerase II pausing downstream of the transcription start site (TSS) in mESCs. We find that pausing of RNA polymerase II correlates with reduced active DNA demethylation and SSB formation at sites of high H2A.Z occupation. The data suggest a role of active DNA demethylation-associated BER in the eviction of H2A.Z downstream the TSS. We thus present a novel role for DNA repair-mediated active DNA demethylation in transcriptional control not only at level of initiation but also in adjusting nucleosomal properties for efficient elongation.

Introduction

CpG methylation as an epigenetic mark is an established principle of gene control in vertebrates and has implications for developmental and cancer biology (Christman, 2002). While the mark is faithfully copied during DNA replication and hence inherited from mother to daughter cell it has long been known that it is in principle a reversible modification (Métivier et al., 2008; Parry et al., 2020). Its true dynamic properties, however, became known only about ten years ago when proteins of the ten-eleven translocation (TET) family and the base excision repair (BER) machinery were shown to be able actively demethylate DNA without the need of cell division (Ito et al., 2011; Maiti and Drohat, 2011; Tahiliani et al., 2009). These observations introduced BER, and more precisely the thymine DNA

glycosylase (TDG) (Cortázar et al., 2011; Cortellino et al., 2011) as a DNA repair enzyme that contributes to transcriptional control in a targeted manner. While phenotypes of altered gene expression have been described upon the depletion of TETs or TDG, it has never been shown that the process of demethylation is indeed generating apurinic/apyrimidinic sites (AP-sites) or single-strand breaks (SSBs) as predicted by the BER model. Neither have these DNA lesions been investigated whether they themselves may have a role in gene regulation. As DNA base damage occurs frequently but randomly throughout the genome, SSBs generation as a consequence of such damage, were initially expected to occur equally randomly. Yet, it was reported that SSBs preferentially arise in genic and regulatory regions and correlate with transcriptional activity but no mechanistic cause or consequence was investigated (Baranello et al., 2014; Cao et al., 2019; Wu et al., 2021). We could recently show that SSBs associate with regions of active DNA demethylation in response to stress (Xu and Schwarz et al. manuscript in preparation), but whether these SSBs are a mere by-product of DNA demethylation and/or transcription, or whether they might have an active role in gene control has not been addressed.

DNA demethylation at promoters was shown to increase accessibility and facilitate TF binding (Calvanese et al., 2012). Oxidised derivatives of 5mC and the active DNA demethylation machinery are, however, also observed to associate with genic regions immediately downstream the transcription start site (TSS) and across the entire gene body (Raiber et al., 2012; Shen et al., 2013) but very little functional evidence is available. A recent study links the presence of fC/caC to CTCF-occupation and splicing (Nanan et al., 2019) or accumulation of fC/caC was reported to cause a mild retardation of RNAP2 transcription in TDG-depleted mESCs (Wang et al., 2015). Furthermore, DNA methylation has been described to affect transcription factor binding but also higher-order chromatin organisation by interacting with proteins that post-translationally modify histone proteins (reviewed in Du et al., 2015; Zhang et al., 2015). Histone modifications can in turn influence DNA methylation patterns and the binding of transcription factors. The combinatorial interpretation of this so-established “epigenetic code” by a variety of reader proteins finally directs the activity and output of the transcription machinery.

To figure out how modulation of the DNA demethylation machinery might affect transcriptional elongation in addition to transcriptional initiation at promoters we performed a precision run-on sequencing (PRO-seq) to capture newly emerging RNA. Combination of this data with previously obtained and published datasets on H2A.Z variants (Hu et al., 2013) and TET activity (Ginno et al., 2020) allowed us to identify a role of TDG and active DNA demethylation in the release of paused RNA polymerase 2 (RNAP2) by the modulation of histone variant H2A.Z abundance downstream of the transcription start site.

Material and Methods

Cell culture

The embryonic stem cells (ESCs) were derived from male C57BL/6JRj mice and depleted of endogenous *Tdg* expression. They carry an artificial *Tdg* minigene that fully reconstitutes molecular and phenotypical characteristics of endogenous TDG. It is flanked by *loxP* sites to allow a tamoxifen-inducible *Tdg* knockout, which is achieved within 24 h after 2 h of induction with 5 μ M of 4-Hydroxytamoxifen. For more detailed information please refer to (Schwarz et al., 2020). ESCs were treated with 5 nM of Talazoparib for 24 h in every experiment and *Tdg* was depleted freshly before each experiment.

Chromatin immunoprecipitation and protein analysis

H2A.Z ChIP was performed in ESCs depleted of endogenous *Tdg* expression but reconstituted with a *Tdg* minigene or an empty “vector” control as described before (Steinacher et al., 2019) using 3 μ g of the H2A.Z antibody (Abcam, ab4174) and incubation with 25 μ g of chromatin.

Western blotting followed by immunodetection was performed as described before (Schwarz et al., 2020) with whole-cell extracts obtained by lysis with a NP-40 lysis buffer including phosphatase inhibitors (PhosSTOP, Roche). Antibodies were incubated with the membrane in a solution of 5% milk powder diluted in TBS with 0.1% Tween-20 for 1 h at RT. Antibodies against acetylated H2A.Z (1:5'000, Sigma-Aldrich, ABE1363), GAPDH (1:20'000, Sigma-Aldrich, G9545), RNAP2 Ser2 phos (1:5'000, Abcam, ab5095), TDG (1:10'000, own production, L58) were then probed with corresponding HRP-conjugated secondary antibody for 1 h at RT.

Single strand break and high throughput sequencing

The detection of endogenous single-strand breaks (SSB) was performed based on the protocol published by Baranello et al. in 2014 (Baranello et al., 2014, 2018) and described in Ming and Schwarz et al. (manuscript submitted).

In brief, genomic DNA was extracted with the Genomic-Tip 100/G Kit (Qiagen) from freshly treated mESCs. 50 μ g of gDNA were subjected to nick labelling using 5 μ l of E.Coli Pol I (NEB) in presence of 2 μ M DIG-dUTPs. The labelled DNA was purified and fragmented and immunoprecipitation was performed over night with 2 μ g of anti-DIG antibody (Roche). The immune complexes were recovered with 40 μ l of preblocked Protein G Dynabeads (Invitrogen), and purified DNA was subjected to library construction for NGS.

Precision run-on and library preparation

PRO-Seq library construction and initial data analysis was performed by the Nascent Transcriptomics Core at Harvard Medical School, Boston, MA as following: For each sample, 1 million permeabilised

cells were used for nuclear run-on. For normalization, 50,000 permeabilised *Drosophila* S2 cells were added to each sample of 1 million cells. Nuclear run-on assays and library preparation were then performed essentially as described (Elrod et al., 2019) with modifications noted below.

The nuclear run-on buffer was prepared as a 4x stock (20mM Tris (pH 8), 20mM MgCl₂, 2mM DTT, 600mM KCl, 40uM biotin-11-NTPs (Perkin Elmer), 20 U SupraseIN (Thermo)). The 4X stock was mixed 1:1 with 2% sarkosyl (Sigma) to yield 2x complete run-on mix. The 3' adapter (RNA oligo: 5'-GAUCGUCGGACUGUAGAACUCUGAAC-3'InvdT) was pre-adenylated prior to use (5' DNA adenylation kit, NEB, according to the manufacturer's instructions). Adenylated oligo was purified by ethanol precipitation, resuspended in water and modification quality was confirmed by electrophoresis on a 15% TBE-Urea gel (Novex). The adapter concentration was adjusted to 10uM and 1uL was used for each ligation using T4 RNA ligase 2, truncated KQ (NEB) with 15% PEG-8000 in the reaction mix. Reactions were incubated at 16°C overnight. Each ligation reaction was then diluted with 180uL of betaine binding buffer (1.42g of betaine brought to 10mL with binding buffer and sterile filtered) containing 1uL of 100uM blocking oligo (TCCGACGATCCCACGTTCCCGTGG/3InvdT/). The blocking oligo is complementary to and overhangs the 3' end of the 3' adapter and reduces 5'/3' adapter dimer formation during the 5' adapter ligation step. The blocking oligo has modifications at its 3' end, to prevent extension by reverse transcriptase. Blocking oligo was also included in the final wash solution prior to the 5' adapter ligation and in the ligation reaction itself, both at 3-fold molar excess over the initial amount of 3' adapter (i.e. 1uL of 30uM per reaction). The 5' adapter ligation was performed as in Elrod et al. except PEG-8000 was increased to 15%.

After reverse transcription, cDNA was immediately amplified for 5-cycles ("preCR" step). A PCR cocktail consisting of NEBNext Ultra II Q5 master mix (NEB) and Illumina TruSeq PCR primers (RP-1, common; and RPI-X, indexing) was added to each cDNA and amplification was done according to manufacturer's suggested 2-step cycling conditions for NGS applications. To determine optimal library amplification, the preCRs were thawed on ice and serially diluted. Each dilution was mixed with a PCR cocktail (Q5 DNA polymerase with high GC enhancer) containing 10uM of Illumina universal P5 and P7 primers and amplified 15-cycles. The reactions were analysed on a 2.2% agarose gel (SeaKem) cast with SYBR gold (ThermoFisher) and optimal number of cycles for final amplification were determined. preCRs were transferred from ice to a pre-heated thermal cycler block and amplified for the appropriate number of additional cycles to reach the total determined in the test amplification. Pooled libraries were then sequenced using Illumina NovaSeq S1 (PE 50) platforms.

Bioinformatic processing

Generation and initial bioinformatic processing of the SSB and ATAC dataset is described in (Xu and Schwarz et al., manuscript in preparation)

PRO-seq FASTQ read pairs were trimmed to 41bp per mate, and read pairs with a minimum average base quality score of 20 were retained. Read pairs were then further trimmed using cutadapt 1.14 to remove adapter sequences and low-quality 3' bases (--match-read-wildcards -m 20 -q 10). R1 reads were then aligned to the spike genome index (dm3) using Bowtie 1.2.2 (-v 2 -p 6 --best --un), with those reads not mapping to the spike genome serving as input to the reference, i.e. primary, genome alignment step (using Bowtie 1.2.2 options -v 2 --best). Reads mapping to the mm10 reference genome were then sorted, via samtools 1.3.1 (-n).

Custom annotation of dominant genes and TSS was performed using the 5' ends from all PRO-seq reads to identify active transcription start sites. A single, dominant TSS was assigned to each gene, selecting the TSS with the highest number of reads within a +/-150 interval relative to each annotated TSS. Active genes were determined by in combination with total RNA-seq of 2i-cultivated C57Bl/6J mESCs from (Williams et al., 2015).

For pause index calculation, loci < 500bp were filtered out and counting windows from TSS/+150 and +250 to +2250 (or TES, whichever was shorter) were defined. The pause index results from the ratio of read densities in the TSS proximal window over the more distal window (TSS proximal reads/TSS window size) / (gene body reads/gene body window size).

Statistic differences were assayed with a Welch two sample t-test or a Wilcoxon rank sum test with continuity correction where (log2 transformed) subsets were used that included zeros or infinities.

All calculations and visualisations were performed with R (Version 4.0.3) with the use of following packages and their dependencies: ggplot2_3.3.2, ggpubr_0.4.0, ChIPseeker_1.26.0, dplyr_1.0.2, QuasR_1.30.0, bowtie_1.30.0, edgeR_3.32.0, eulerr_6.1.0, GenomicRanges_1.42.0.

Accession

Datasets of SSB-seq, ATAC-seq can be found on the Gene Expression Omnibus with the accession number GSE166964 and the total RNA-seq as part of the BioProject PRJNA743896.

Results

TDG is required for efficient Pause Release of RNAP2

To investigate the role of active DNA demethylation in transcriptional elongation, we performed a precision run-on sequencing (PRO-seq) to capture newly emerging RNA and the corresponding position of engaged RNAP2 (Mahat et al., 2016). We performed the experiment with TDG proficient and deficient mESCs each in two biological replicas and exposed them to DMSO or 5 nM Talazoparib (Tal) for 24 h (Supplementary Figure S1A&B), of which the latter showed a strong induction of stress-related

genes (Xu and Schwarz et al., manuscript in preparation). The absolute expression levels in WT mESCs detected in the PRO-seq were in good agreement with previous totalRNA-seq data ($R=0.69$, Figure S1C&D) allowing us to directly compare the two datasets (Wang et al., 2014). As reported before, overall expression levels between unchallenged WT and TDGnull mESCs did barely vary and showed only 6 transcripts to be significantly deregulated upon TDG depletion (Figure 1A, Cortázar et al. 2011). We therefore focussed our analysis on the genes induced by the Tal-mediated stress response and detected that in the five minutes of run-on reaction, 250 genes exhibit a productive gene body expression that is increased by a \log_2 factor ≥ 1.5 , of which 80 genes are common with the 1'569 upregulated genes detected in the total RNA-seq after 24 h of treatment (Figure S1E). To then examine transcriptional initiation vs. elongation, we used the "pause index" (PI). The PI represents the amount of promoter-proximally paused RNA polymerase II (RNAP2) versus RNAP2 engaged in productive elongation and is defined by the ratio of the promoter-proximal read density over the transcript density further downstream in the gene body (Figure 1B, a over b)(Min et al., 2011).

This analysis revealed that the 1'569 Tal-induced genes detected in the total RNA display a significantly higher PI in mESCs depleted of TDG (Figure 1C), which was notably also visible on the level of all detected transcripts (Figure S1F). As an increased PI can either result from higher transcriptional initiation, or reduced productive transcriptional elongation, we separately measured the read densities at the location of RNAP2 pausing, representing initiation (as in Figure 1B, a), and over the whole gene body to assess elongation. Measuring the Tal-upregulated genes from the total RNA-seq we observed the same level of PRO-seq reads at the TSS and a trend towards reduced level of reads in the gene body when TDG was depleted (Figure 1B/D). Analysing then the set of 250 significantly upregulated genes from the PRO-seq, we observed a much stronger and significant reduction of reads in the gene body of TDG-depleted mESCs (Figure 1E/F, S1H). This indicates that in this scenario, TDG is responsible for the facilitated release into productive elongation rather than for transcriptional initiation, and explains why no alterations of DNA methylation or accessibility could be observed at the promoter of these genes despite the changed transcription (Xu and Schwarz et al., manuscript in preparation). This is also underlined by the observation that highly-paused genes, display a significantly higher upregulation upon Tal in TDG proficient mESCs, which is not distinguishable in mESCs depleted of TDG (Figure 1G).

H2A.Z Accumulation in TDG Deficient mESCs Affects Pause Release

To understand how TDG activity might promote transcriptional elongation, we investigated the non-canonical histone variant H2A.Z. H2A.Z was described nearly 40 years ago to be associated preferentially with active genes and later shown to correlate with chromatin accessibility and being enriched at enhancers and in particular around the TSS of transcriptionally active genes (reviewed in Giaimo et al., 2019). While H2A.Z seems to support transcriptional initiation at promoters, its presence downstream of the TSS was recently shown to correlate with lower transcription (Lashgari et al., 2017). Consistently, degran-mediated depletion of H2A.Z was reported to increase transcriptional elongation (Mylonas et al., 2021). These observations indicate that H2A.Z downstream the TSS needs to be exchanged if transcriptional elongation is to be maximized, a concept that is supported by studies in plants and yeast (Nguyen and Cheong, 2018; Ranjan et al., 2020; Tramantano et al., 2016). How this exchange is achieved is not entirely clear, but one possibility involves the acetylation of H2A.Z (Hu et al., 2013; Valdés-Mora et al., 2012). Notably, it was also shown that nucleosome occupation was increased upon knockdown of H2A.Z, particularly at regions occupied by the histone acetyl-transferase p300 (Hu et al., 2013). Acetylation of H2A.Z might thereby facilitate its exchange by nucleosome eviction by an unknown chromatin remodelling process facilitates its displacement by the passage of RNAP2 (Hodges et al., 2009; Kulaeva et al., 2013).

To address a potential role of TDG-mediated active DNA demethylation in H2A.Z exchange, we analysed our PRO-seq data with published CHIP-seq data on H2A.Z and its acetylated version H2A.Zac in mESCs (Figure S2A, Hu et al., 2013). This revealed that in WT mESCs, genes with H2A.Z at their TSS show increased transcriptional activity at the TSS by a median factor of ~ 2 when compared to genes with no significant enrichment of H2A.Z. Genes with significant H2A.Zac enrichment at the TSS show an even higher transcriptional activity with a $\sim 10x$ increase compared to promoters without any H2A.Z (Figure 2A). This increased transcriptional activity at the TSS is, however, not accompanied by increased transcription further downstream; the same genes with unmodified H2A.Z at the TSS do not show higher levels of elongating transcripts and this those with H2A.Zac enrichment do so only by a factor ~ 2.7 (Figure 2B). This translates into an increased PI for genes occupied by (non)acetylated H2A.Z (Figure 2C), which recapitulates previous observations and supports the potential role of H2A.Z in RNAP2 pausing.

When we examined the same gene sets in mESCs depleted of TDG, we observed that the genes occupied by acetylated H2A.Zac show a significant further increase of the PI, which indicates a defect in transcriptional pause release in the absence of TDG (Figure 2D). We then examined our own H2A.Z CHIP-seq data obtained from WT and TDG-depleted mESCs and observed that the number of regions with significant H2A.Z enrichment increased drastically upon loss of TDG (88'779 vs 55'988, Figure 2E). The depletion of TDG hence lead to a wide-spread increase of H2A.Z at TSS and an accompanying

increase of RNAP2 pausing at the 3'125 genes with a significant increase of H2A.Z enrichment (i.e. ChIP-peaks, Figure 2F&G). While the increase of PI in genes with H2A.Z accumulation is significant, we have to consider that these mESCs depleted of TDG generally show only minute changes in transcription when cultivated in 2i-medium (Figure 1A, Cortázar et al., 2011; Steinacher et al., 2019), and no large effects are expected. Concerning the accessibility of regions that are either enriched in H2A.Z or H2A.Zac, we could confirm by ATAC-seq (Xu and Schwarz et al., manuscript in preparation) that regions with a high abundance of H2A.Zac display a higher accessibility than those with H2A.Z (Figure S2B)(Hu et al., 2013; Valdés-Mora et al., 2017).

More interestingly, the accessibility in regions enriched with H2A.Zac in WT mESCs, is reduced in cells that are depleted of TDG (Figure S2C), indicating a TDG-dependent effect on accessibility via the acetylation of H2A.Z. All in all, we can recapitulate the beneficial effect of H2A.Z (variants) at TSSs for transcriptional initiation and show for the first time that accumulation of H2A.Z in TDG-depleted mESCs leads to an increase of RNAP2 pausing. This infers a role of TDG in the modulation of H2A.Z (variants) to facilitate transcriptional elongation into the gene body, where H2A.Z is detrimental (Figure S3A, Coleman-Derr and Zilberman, 2012).

H2A.Z positioning coincides with regions of active DNA demethylation and SSBs

TDG is known to interact with the histone acetyl transferase complex CBP/p300 and was shown to cooperatively increase transcriptional activation by acetylation of histone 3 tails (Henry et al., 2016; Léger et al., 2014). Data on TDG-mediated acetylation of H2A variants are missing, so far. We however realised that acetylation might not be the only way on how H2A.Z displacement by RNAP2 or chaperones might be achieved for productive transcription.

It was previously shown that DNA single strand-breaks (SSBs) at gene promoters correlate with transcriptional activity (Baranello et al., 2014) and we showed recently by genome-wide mapping of SSBs, that at least 40% of detectable SSBs in mESCs overlap with sites of active DNA demethylation and contribute to gene control (Xu and Schwarz et al., manuscript submitted). Analysing the distribution of SSBs at the TSS of genes, we observed two peaks of enrichment, one upstream of the TSS, likely representing active DNA demethylation in core promoters (Cooper, 2005), and one downstream of the TSS (Figure 3A). The enrichment of SSBs in both areas (and to a lesser degree also in the gene body) positively correlates with transcriptional activity (Figure 3A and S3C). The SSB enrichment downstream of the TSS coincides with the area of maximum enrichment of H2A.Z and its acetylated form at around +75 bp (Figure 3B). Analysis of regions with high H2A.Zac enrichment also revealed a higher level of SSB occurrence than those with unmodified H2A.Z (Figure S3D). In line with previous observations, stating that H2A.Z can facilitate BER (Li and Delaney, 2019), regions enriched in SSBs (i.e. ChIP peaks) do generally associate with H2A.Z and H2A.Zac positioning, and SSBs at these sites are reduced in

mESCs without TDG (Figure 3C).

To address the potential role of TDG and active DNA demethylation in H2A.Z dynamics, we combined our data with the recently published dataset of 5mCs undergoing TET-dependent oxidation at a high rate (Ginno et al., 2020). The authors showed that high TET-activity on CpGs located upstream of a TSS correlates with low expression, while higher expression was much better correlating with high TET activity downstream of the TSS and in the gene body. This is in line with the observation that genes with TDG-dependent excision of fC/caC in their gene bodies display higher levels of expression than those genes without oxmC generation (Figure S3E, Shen et al. 2013) and we observed that SSBs associate with these dynamic CpGs in a TET-activity dependent manner (Figure S3F). Including H2A.Z in these analyses revealed that high TET-activity also correlates very well with the abundance of H2A.Z (Figure 3D), but slightly less so with H2A.Zac (Figure S3G), indicating different dynamics of H2A.Z acetylation upon TET oxidation. Furthermore, we observed that SSBs around TSSs increase in dependence of H2A.Z(ac) abundance (Figure 3E), which was indicative of increased transcriptional levels (Figure 2A/B). Together, this establishes a tight connection between elevated transcriptional activity and active DNA demethylation downstream of the TSS which is accompanied by increased levels of H2A.Zac.

Based on these observations, we analysed SSBs occurrence at the TSS and the gene body of genes with regard to expression. Analysis over all genes in unchallenged mESCs, showed a higher level of SSBs in TDG-depleted cells that was accompanied by a decrease of PRO-seq reads at the TSS (Figure 3F, S1G). No significant difference of transcription or SSBs was observed in the gene body of both genotypes. As a higher level of SSBs around TSSs was observed to be beneficial for transcription in WT ESCs (Figure 3A, Xu and Schwarz et al. manuscript in preparation), it seems that this is not the case in TDG-depleted cells. Since TDG-depletion in mESCs kept in a pluripotent ground state barely alters the overall expression profile of genes (Cortázar et al., 2011), we analysed genes that become upregulated in ESCs upon the treatment with Tal, but show a reduced transcriptional elongation and a consequential increase of PI in the absence of TDG (Figure 1C-F). We measured that in both gene subsets (defined by the total RNA-seq or the PRO-seq), SSBs were reduced in the 150 bp window after the TSS, where RNAP2 is pausing and H2A.Z is most abundant, but also further on in the gene body (Figure 3F). These results indicate the involvement active DNA demethylation represented by high TET-activity and TDG-generated SSBs localising to H2A.Z(ac)-enriched regions downstream of the TSS to facilitate RNAP2 pause release.

Discussion

Regulation of transcription at the level of RNAP2 pause release has lately emerged as an efficient means for rapid responses to signalling cues, making use of the assembled and poised transcription

machinery (Adelman and Lis, 2012; Liu et al., 2015). The best described models to achieve pause release include acetylation of TSS-proximal histones by histone acetyl transferases like p300 (Vaid et al., 2020) and/or the regulation of negative elongation factors, either by dissociation through phosphorylation or acetylation, or by attraction to enhancer RNAs to displace negative elongation factors (Jonkers and Lis, 2015; Schaukowitch et al., 2014; Williams et al., 2015). The histone variant H2A.Z has been of particular interest in transcriptional pause release as its deposition is focussed to gene promoters and a narrow window downstream of the TSS (Giaino et al., 2019) and its dissociation from chromatin correlates with productive elongation (Mylonas et al., 2021). How eviction of H2A.Z is achieved in the context of transcription is, however, not entirely clear. Based on the data presented here, we provide a model proposing that active DNA demethylation is part of this process. We show that high TET activity and SSB formation correlate coincide with occurrence of (acetylated) H2A.Z and that depletion of TDG leads to a drastic increase of H2A.Z at the TSSs. The role of TDG in this scenario might be multifunctional. On one hand, it is interacting with the histone acetylase p300 and thereby contributes to histone acetylation and facilitates transcription (Henry et al., 2016; Kulaeva et al., 2013; Léger et al., 2014). TDG might therefore act on H2A.Z eviction via acetylation as this modification reduces association of histones to the DNA. On the other hand, TDG's enzymatic activity is causing abasic sites that are further processed into SSBs, the subsequent repair of which causes chromatin conformational disturbances, which will facilitate the eviction of single histones or even whole octamers (Madders and Parsons, 2020; Odell et al., 2013). A similar scenario was recently described where histone H1 eviction as well as PARP1-dependent local protein modifications was facilitated in the presence of TDG (Barekati et al., manuscript in preparation). These processes are not mutually exclusive and might very well act in combination. While acetylation was already shown to increase transcriptional efficiency (Vaid et al., 2020), the impact of SSBs has not yet been demonstrated.

Both these hypotheses demand further investigation that specifically address H2A.Z acetylation and its eviction in dependence of TDG. Using a CRISPR-guided CAS9-nickase to induce SSBs at candidate TSSs in a controlled manner could for example reveal the impact of SSB-generation on the abundance of histone variants or nucleosomes. To increase the observed effect, these experiments should be carried out in challenged cells and it might be advisable to choose a cell type that is not as hypomethylated and generally accessible as naïve mESCs. Nonetheless, the obtained results and the observations by others that DNA methylation and H2A.Z abundance are antagonistic (Kobor and Lorincz, 2009; Zilberman et al., 2008), strongly support the mechanistic coupling of active DNA demethylation and regulation of H2A.Z abundance.

In conclusion, the data presented here suggest a model for transcriptional pause release (Figure 4F) whereby active DNA demethylation plays an active regulatory role. While acetylation of H2A.Z might facilitate a passive displacement by Ser2-phosphorylated RNAP2, generation of SSBs through targeted

active DNA demethylation close to the +1 nucleosome might evict the entire nucleosome (Madders and Parsons, 2020; Odell et al., 2013) and thereby provide a larger window of passage for paused RNAP2 with strong transcriptional bursts (Larsson et al., 2018).

Acknowledgements

We would like to thank the Nascent Transcriptomics Core at Harvard Medical School, Boston, MA for performing the PRO-seq library construction. Special thanks go to Seth Goldman and Karen Adelman for the kind assistance with data analysis and helpful explanations and discussions. We would also like to thank Roland Steinacher for his discussion and feedback to the manuscript.

References

- Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* *13*, 720–731.
- Baranello, L., Kouzine, F., Wojtowicz, D., Cui, K., Przytycka, T., Zhao, K., and Levens, D. (2014). DNA Break Mapping Reveals Topoisomerase II Activity Genome-Wide. *Int. J. Mol. Sci.* *15*, 13111–13122.
- Calvanese, V., Fernández, A.F., Urdinguio, R.G., Suárez-Alvarez, B., Mangas, C., Pérez-García, V., Bueno, C., Montes, R., Ramos-Mejía, V., Martínez-Cambor, P., et al. (2012). A promoter DNA demethylation landscape of human hematopoietic differentiation. *Nucleic Acids Res.* *40*, 116–131.
- Cao, H., Salazar-García, L., Gao, F., Wahlestedt, T., Wu, C.L., Han, X., Cai, Y., Xu, D., Wang, F., Tang, L., et al. (2019). Novel approach reveals genomic landscapes of single-strand DNA breaks with nucleotide resolution in human cells. *Nat. Commun.* *10*, 1–14.
- Christman, J.K. (2002). 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* *21*, 5483–5495.
- Coleman-Derr, D., and Zilberman, D. (2012). Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS Genet.* *8*, e1002988–e1002988.
- Cooper, S.J. (2005). Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* *16*, 1–10.
- Cortázar, 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.
- 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.
- Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015). DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* *16*, 519–532.
- Elrod, N.D., Henriques, T., Huang, K.L., Tatomer, D.C., Wilusz, J.E., Wagner, E.J., and Adelman, K. (2019). The Integrator Complex Attenuates Promoter-Proximal Transcription at Protein-Coding Genes. *Mol. Cell* *76*, 738-752.e7.
- Gaiimo, B.D., Ferrante, F., Herchenröther, A., Hake, S.B., and Borggrefe, T. (2019). The histone variant H2A.Z in gene regulation. *Epigenetics and Chromatin* *12*, 1–22.
- Ginno, P.A., Gaidatzis, D., Feldmann, A., Hoerner, L., Imanci, D., Burger, L., Zilbermann, F., Peters, A.H.F.M., Edenhofer, F., Smallwood, S.A., et al. (2020). A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nat. Commun.* *11*, 2680.
- Henry, R.A., Mancuso, P., Kuo, Y.-M., Tricarico, R., Tini, M., Cole, P.A., Bellacosa, A., and Andrews, A.J. (2016). Interaction with the DNA Repair Protein Thymine DNA Glycosylase Regulates Histone Acetylation by p300.
- Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* *325*, 626–628.
- Hu, G., Cui, K., Northrup, D., Liu, C., Wang, C., Tang, Q., Ge, K., Levens, D., Crane-Robinson, C., and Zhao, K. (2013). H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* *12*, 180–192.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* *333*,

1300–1303.

Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* *16*, 167–177.

Kobor, M.S., and Lorincz, M.C. (2009). H2A.Z and DNA methylation: irreconcilable differences. *Trends Biochem. Sci.* *34*, 158–161.

Kulaeva, O.I., Hsieh, F.K., Chang, H.W., Luse, D.S., and Studitsky, V.M. (2013). Mechanism of transcription through a nucleosome by RNA polymerase II. *Biochim. Biophys. Acta - Gene Regul. Mech.* *1829*, 76–83.

Larsson, A.J.M., Johnsson, P., Hagemann-jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, Å., Rivera, C.M., Ren, B., and Sandberg, R. (2018). Genomic encoding of transcriptional burst kinetics. *Nature*.

Lashgari, A., Millau, J.-F., Jacques, P.-É., and Gaudreau, L. (2017). Global inhibition of transcription causes an increase in histone H2A.Z incorporation within gene bodies. *Nucleic Acids Res.* *45*, 12715–12722.

Léger, H., Smet-Nocca, C., Attmane-Elakeb, A., Morley-Fletcher, S., Benecke, A.G., and Eilebrecht, S. (2014). A TDG/CBP/RAR α ternary complex mediates the retinoic acid-dependent expression of DNA methylation-sensitive genes. *Genomics, Proteomics Bioinforma.* *12*, 8–18.

Li, C., and Delaney, S. (2019). Histone H2A Variants Enhance the Initiation of Base Excision Repair in Nucleosomes. *ACS Chem. Biol.* *14*, 1041–1050.

Liu, X., Kraus, W.L., and Bai, X. (2015). Ready, pause, go: regulation of RNA polymerase II pausing and release by cellular signaling pathways. *Trends Biochem. Sci.* *40*, 516–525.

Madders, E.C.E.T., and Parsons, J.L. (2020). Base Excision Repair in Chromatin and the Requirement for Chromatin Remodelling BT - Mechanisms of Genome Protection and Repair. D.O. Zharkov, ed. (Cham: Springer International Publishing), pp. 59–75.

Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J., and Lis, J.T. (2016). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* *11*, 1455–1476.

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.

Métivier, R., Gallais, R., Tiffoche, C., Le Péron, 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.

Min, I.M., Waterfall, J.J., Core, L.J., Munroe, R.J., Schimenti, J., and Lis, J.T. (2011). Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells. *Genes Dev.* *25*, 742–754.

Mylonas, C., Lee, C., Auld, A.L., Cisse, I.I., and Boyer, L.A. (2021). A dual role for H2A.Z.1 in modulating the dynamics of RNA polymerase II initiation and elongation. *Nat. Struct. Mol. Biol.* *28*, 435–442.

Nanan, K.K., Sturgill, D.M., Prigge, M.F., Thenoz, M., Dillman, A.A., Mandler, M.D., and Oberdoerffer, S. (2019). TET-Catalyzed 5-Carboxylcytosine Promotes CTCF Binding to Suboptimal Sequences Genome-wide. *iScience* *19*, 326–339.

Nguyen, N.H., and Cheong, J.-J. (2018). H2A.Z-containing nucleosomes are evicted to activate AtMYB44 transcription in response to salt stress. *Biochem. Biophys. Res. Commun.* *499*, 1039–1043.

- Odell, I.D., Wallace, S.S., and Pederson, D.S. (2013). Rules of engagement for base excision repair in chromatin. *J. Cell. Physiol.* **228**, 258–266.
- Parry, A., Rulands, S., and Reik, W. (2020). Active turnover of DNA methylation during cell fate decisions. *Nat. Rev. Genet.*
- Raiber, E.-A., Beraldi, D., Ficz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.* **13**, R69.
- Ranjan, A., Nguyen, V.Q., Liu, S., Wisniewski, J., Kim, J.M., Tang, X., Mizuguchi, G., Elalaoui, E., Nickels, T.J., Jou, V., et al. (2020). Live-cell single particle imaging reveals the role of RNA polymerase II in histone H2A.Z eviction. *Elife* **9**.
- Schaukowitch, K., Joo, J.-Y., Liu, X., Watts, J.K., Martinez, C., and Kim, T.-K. (2014). Enhancer RNA Facilitates NELF Release from Immediate Early Genes. *Mol. Cell* **56**, 29–42.
- Schwarz, S.D., Grundbacher, E., Hrovat, A.M., Xu, J., Kuśnierczyk, A., Vågbø, C.B., Schär, P., Schuermann, D., Sczepanski, J., Torchia, J., et al. (2020). Inducible TDG knockout models to study epigenetic regulation [version 2; peer review: 3 approved].
- 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.
- Steinacher, R., Barekati, Z., Botev, P., Kuśnierczyk, A., Slupphaug, G., and Schär, P. (2019). SUMOylation coordinates BERosome assembly in active DNA demethylation during cell differentiation. *EMBO J.* **38**, e99242.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* (80-). **324**, 930–935.
- Tramantano, M., Sun, L., Au, C., Labuz, D., Liu, Z., Chou, M., Shen, C., and Luk, E. (2016). Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex. *Elife* **5**.
- Vaid, R., Wen, J., and Mannervik, M. (2020). Release of promoter-proximal paused Pol II in response to histone deacetylase inhibition. *Nucleic Acids Res.* **48**, 4877–4890.
- Valdés-Mora, F., Song, J.Z., Statham, A.L., Strbenac, D., Robinson, M.D., Nair, S.S., Patterson, K.I., Tremethick, D.J., Stirzaker, C., and Clark, S.J. (2012). Acetylation of H2A.Z is a key epigenetic modification associated with gene deregulation and epigenetic remodeling in cancer. *Genome Res.* **22**, 307–321.
- Valdés-Mora, F., Gould, C.M., Colino-Sanguino, Y., Qu, W., Song, J.Z., Taylor, K.M., Buske, F.A., Statham, A.L., Nair, S.S., Armstrong, N.J., et al. (2017). Acetylated histone variant H2A.Z is involved in the activation of neo-enhancers in prostate cancer. *Nat. Commun.* **8**, 1346.
- Wang, I.X., Core, L.J., Kwak, H., Brady, L., Bruzel, A., McDaniel, L., Richards, A.L., Wu, M., Grunseich, C., Lis, J.T., et al. (2014). RNA-DNA differences are generated in human cells within seconds after RNA exits polymerase II. *Cell Rep.* **6**, 906–915.
- Wang, L., Zhou, Y., Xu, L., Xiao, R., Lu, X., Chen, L., Chong, J., Li, H., He, C., Fu, X.-D., et al. (2015). Molecular basis for 5-carboxycytosine recognition by RNA polymerase II elongation complex. *Nature*.
- Williams, L.H., Fromm, G., Gokey, N.G., Henriques, T., Muse, G.W., Burkholder, A., Fargo, D.C., Hu, G., and Adelman, K. (2015). Pausing of RNA Polymerase II Regulates Mammalian Developmental

Potential through Control of Signaling Networks. *Mol. Cell* 58, 311–322.

Wu, W., Hill, S.E., Nathan, W.J., Paiano, J., Callen, E., Wang, D., Shinoda, K., van Wietmarschen, N., Colón-Mercado, J.M., Zong, D., et al. (2021). Neuronal enhancers are hotspots for DNA single-strand break repair. *Nature*.

Xu, J., Schwarz, S.D., Gunasekera, K., Ferrari, E., Hottiger, M.O., Schär, P., and Steinacher, R. Covalent PARylation is an integral part of TDG-BER-mediated active DNA demethylation in embryonic stem cells.

Xu, J., Schwarz, S.D., Gunasekera, K., Steinacher, R., Hottiger, M.O., and Schär, P. PARP inhibition induces cytotoxicity in mESCs by activating endogenous retroviruses.

Zhang, T., Cooper, S., and Brockdorff, N. (2015). The interplay of histone modifications – writers that read. *EMBO Rep.* 16, 1467–1481.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* 456, 125–129.

Figure Legends and Figures

Figure 1 TDG depletion leads to a decreased pause release of RNAP2

A Volcano plot of transcripts detected by PRO-seq in TDGnull vs WT mESCs. Large dots are genes, smaller dots are predicted transcripts. **B** Metagene analysis of Tal-responsive genes (n=1'568) from total RNA-seq. a: 150 bp window immediately downstream the TSS, containing transcripts created by paused RNAP2. b: window of 2 kb downstream of TSS as a proxy for productive elongation. **C** Pause index (PI) of Tal-upregulated genes from total RNA-seq. **D** PRO-seq reads of genes in C, at RNAP2 pause site (TSS) and in the total gene body (GB). **E** Pause index (PI) of genes upregulated by Tal from PRO-seq. **F** PRO-seq reads of genes in E, at RNAP2 pause site (TSS) and in the total gene body (GB). **G** Tal-mediated upregulation of genes from E, stratified by their PI. Asterisks indicate statistical p-values of: *: p≤0.05, **:p≤0.01, ***:p≤0.001, ****:p≤0.0001

Figure 2 H2A.Z accumulation in TDG deficient ESCs affects pause release

A PRO-seq expression of WT mESCs at TSS that exhibit no H2A.Z occupation (n= 6'431), only unmodified H2A.Z (n=1'582) or acetylated H2A.Z (n=12'710). Fold change (FC) comparison to genes without H2A.Z at TSS are indicated. **B** WT PRO-seq expression in GB of genes from A **C** Pause index of genes in **D** PI of genes from A in WT and mESCs depleted of TDG. **E** Number of H2A.Z enriched regions (i.e. ChIP-seq peaks) in TDG WT and TDG null mESCs **F** H2A.Z levels at TSS of all detected transcripts in PRO-seq. **G** PI of genes that display a gain of H2A.Z peaks upon TDG depletion (n=3'125). Asterisks indicate statistical p-values of: *: p≤0.05, **:p≤0.01, ***:p≤0.001, ****:p≤0.0001

Figure 3 Sites of active DNA demethylation correlate with H2A.Z enrichment and transcription

A Distribution of SSB reads over TSS, stratified by the total RNA-seq expression levels of genes. Each tile contains the same number of genes (n=4'618). **B** Peak distribution of H2A.Z(ac) around the TSS **C** Density plot of SSB peaks across regions of H2A.Z(left) and H2A.Zac(right) enrichments. Lighter shadings indicate the 95% confidence interval **D** Density plot of dynamic CpGs, stratified by TET-activity, over H2A.Z **E** Distribution of SSB reads over TSSs based on H2A.Z(ac) abundance **F** SSB counts at TSS or GBs of all detected genes in WT and TDGnull mESCs **G** SSB counts at TSSs or GBs of genes that become upregulated by Tal in total RNA-seq (left) or PRO-seq (right). Asterisks indicate statistical p-values of: *:p≤0.05, **:p≤0.01, ***:p≤0.001, ****:p≤0.0001

Figure 4 Working model for active DNA demethylation associated RNAP2 pause release

Left: Transcriptional initiation is happening and facilitated by permissive chromatin marks (H3K4me3) and DNase hypersensitivity (DHS) at the promoter. Productive elongation by RNAP2 is hindered because of H2A.Z occupancy downstream of the TSS. Right: TET/TDG-mediated active DNA demethylation evicts H2A.Z, possibly with the help of histone acetylation by p300, and facilitates pause release of RNAP2.

Figure 1

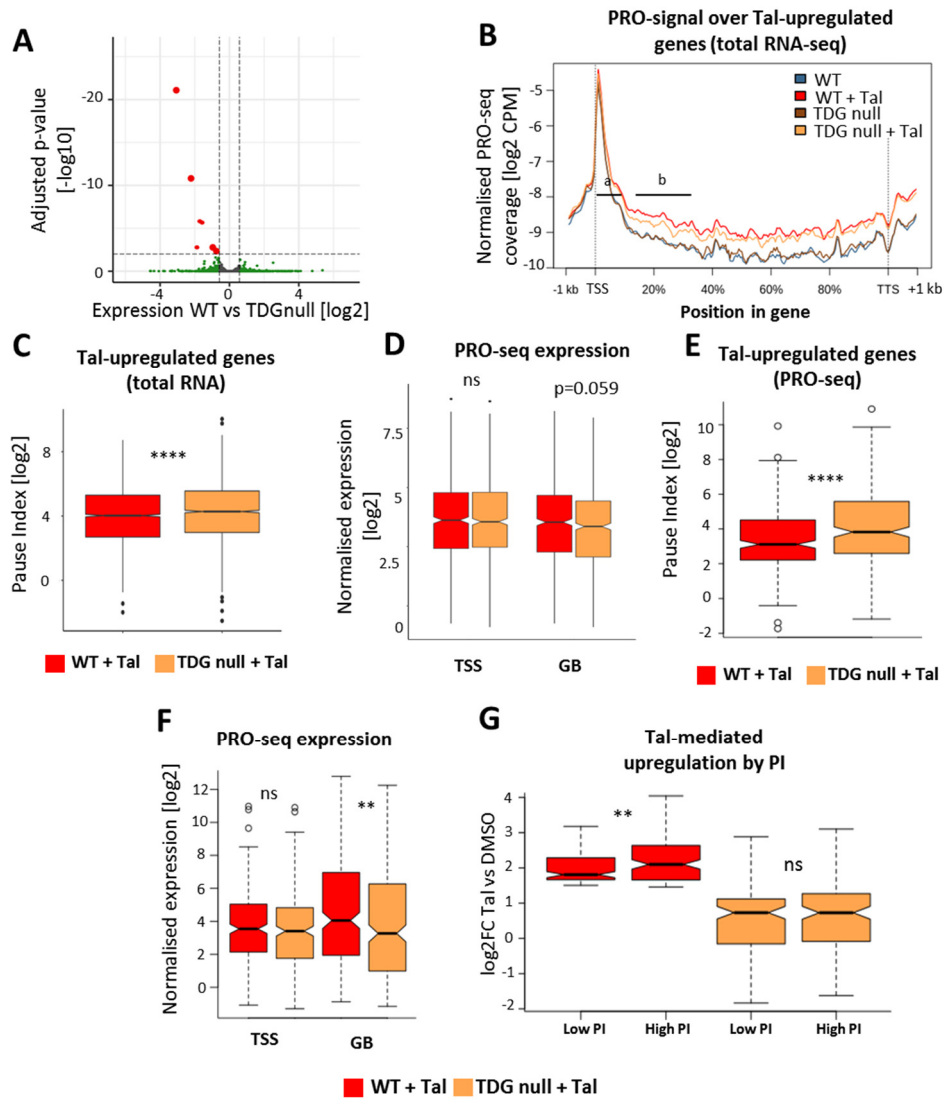


Figure 2

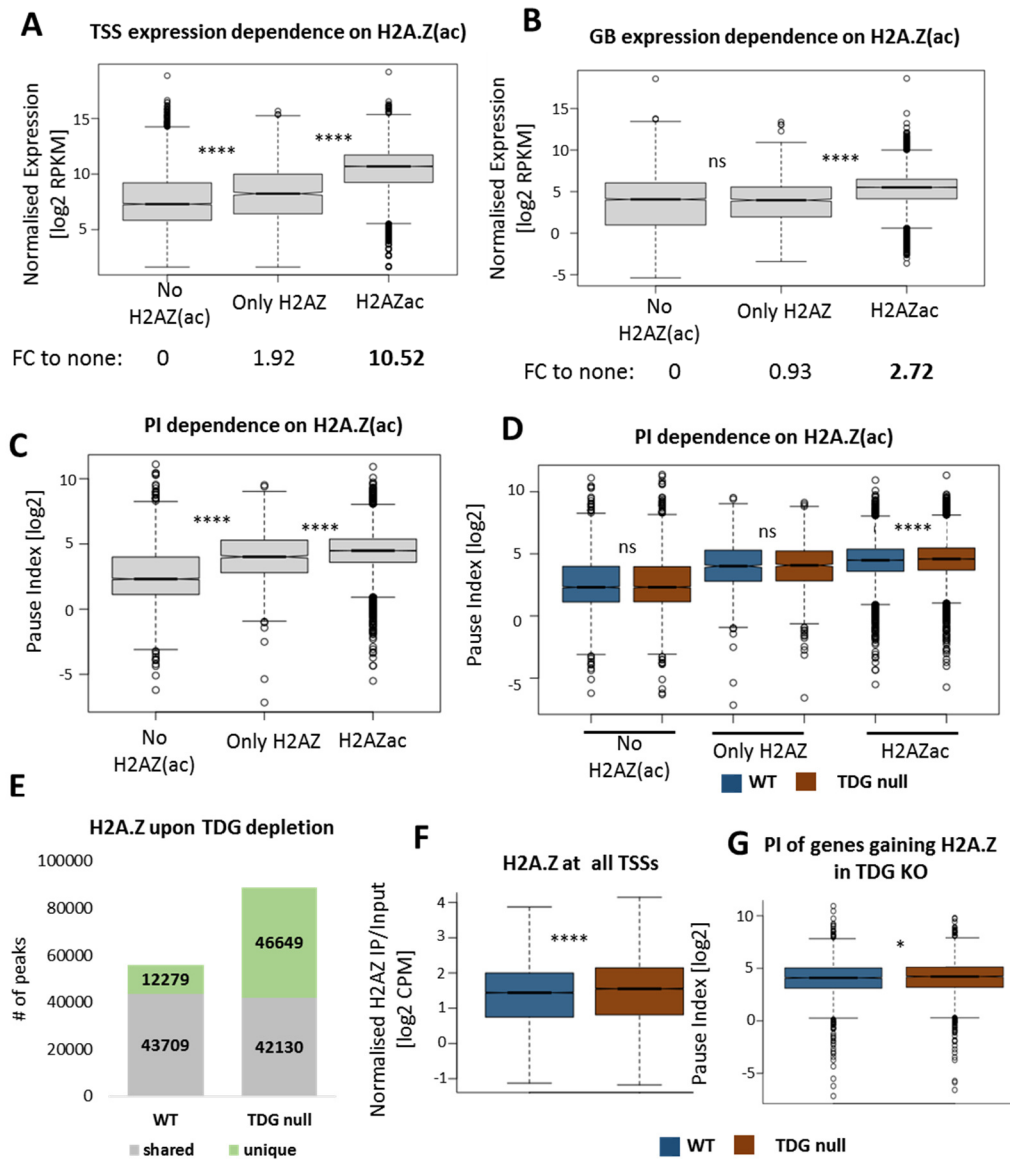


Figure 3

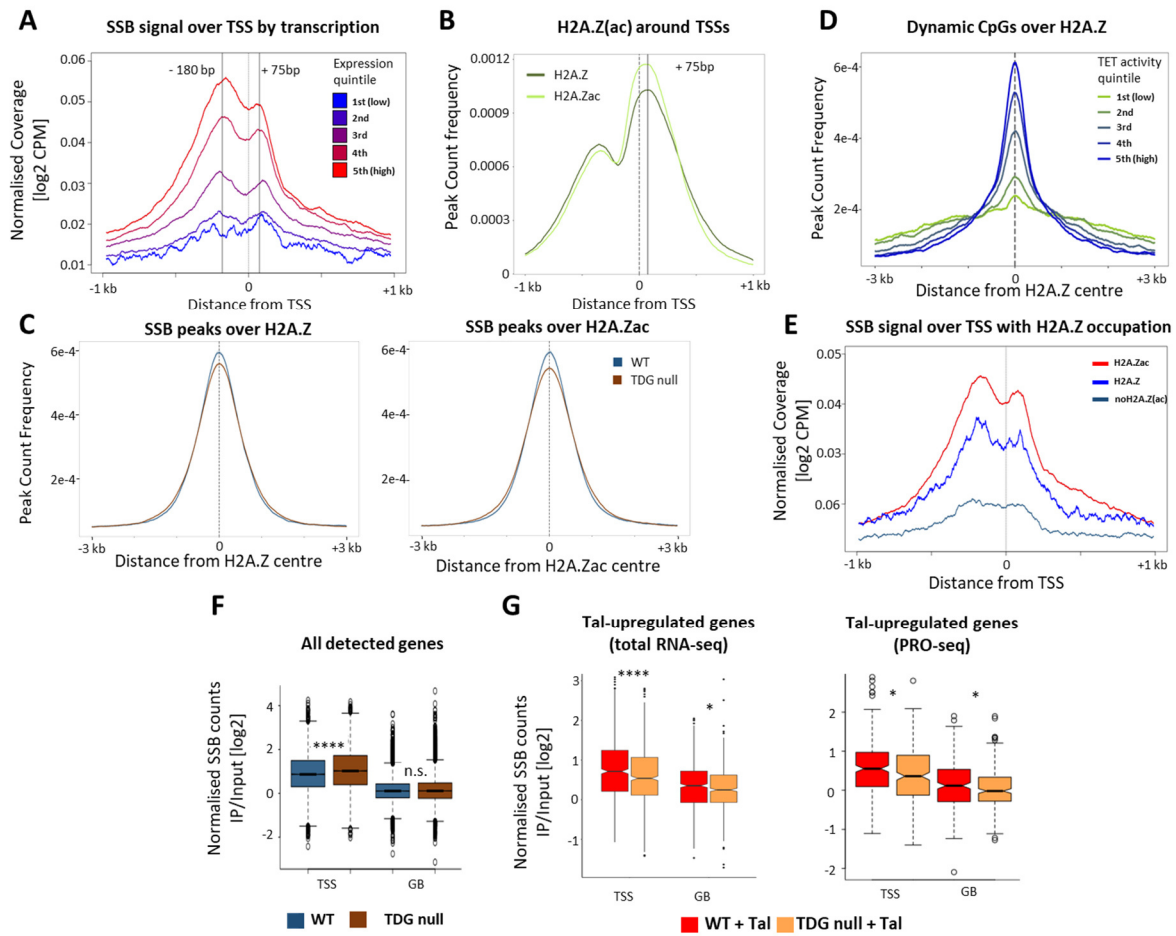
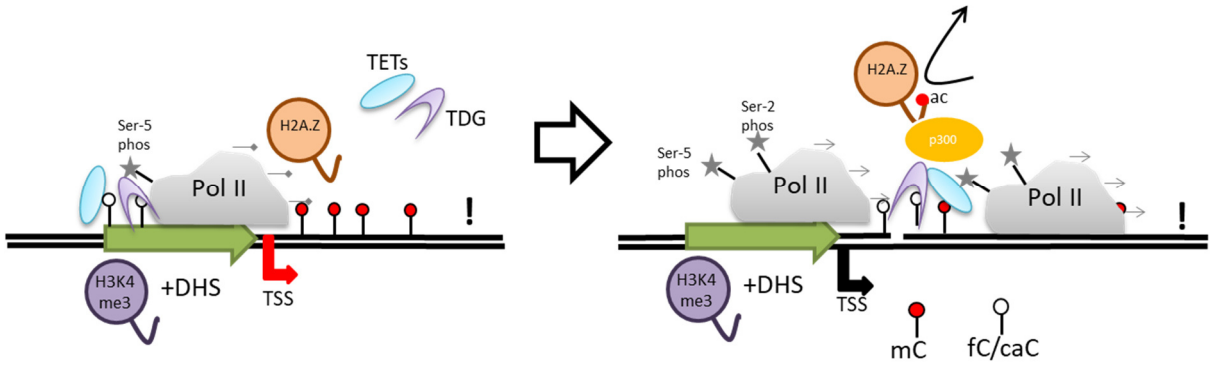
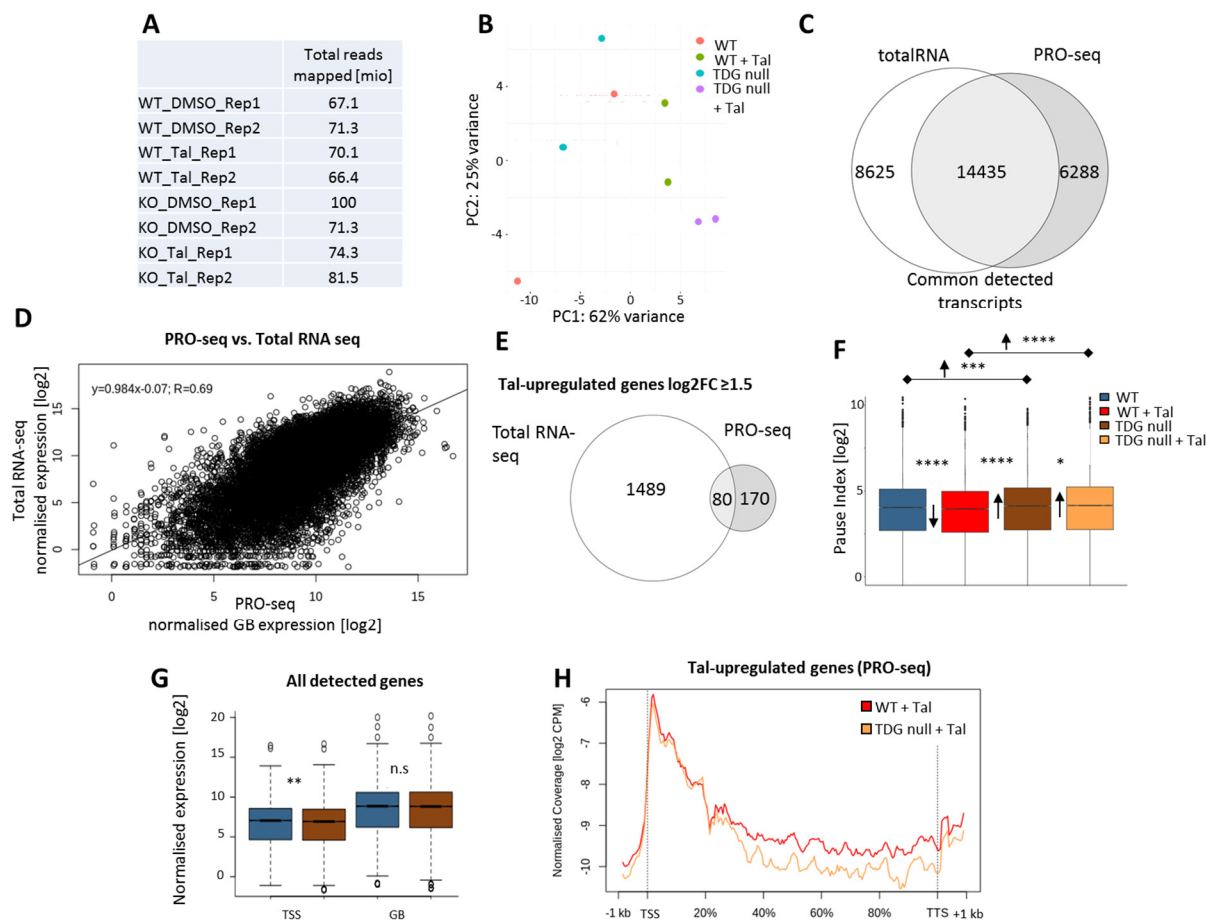


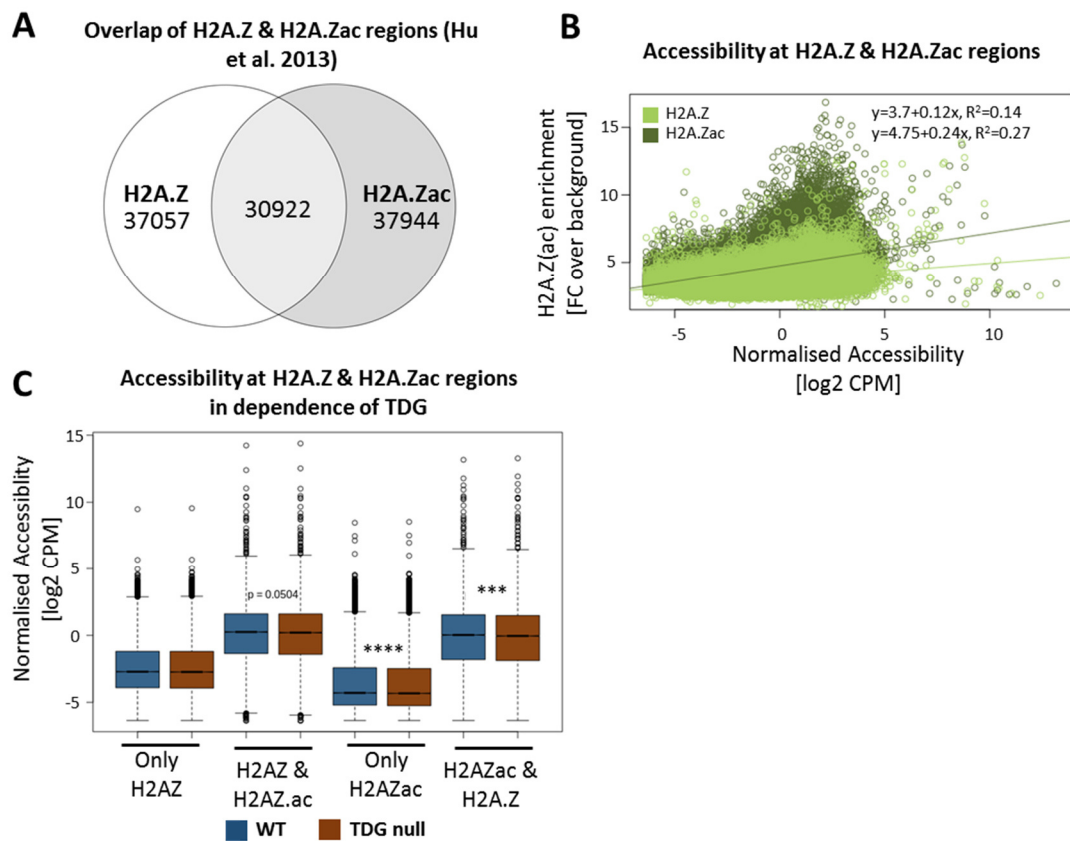
Figure 4



Supplementary Figures

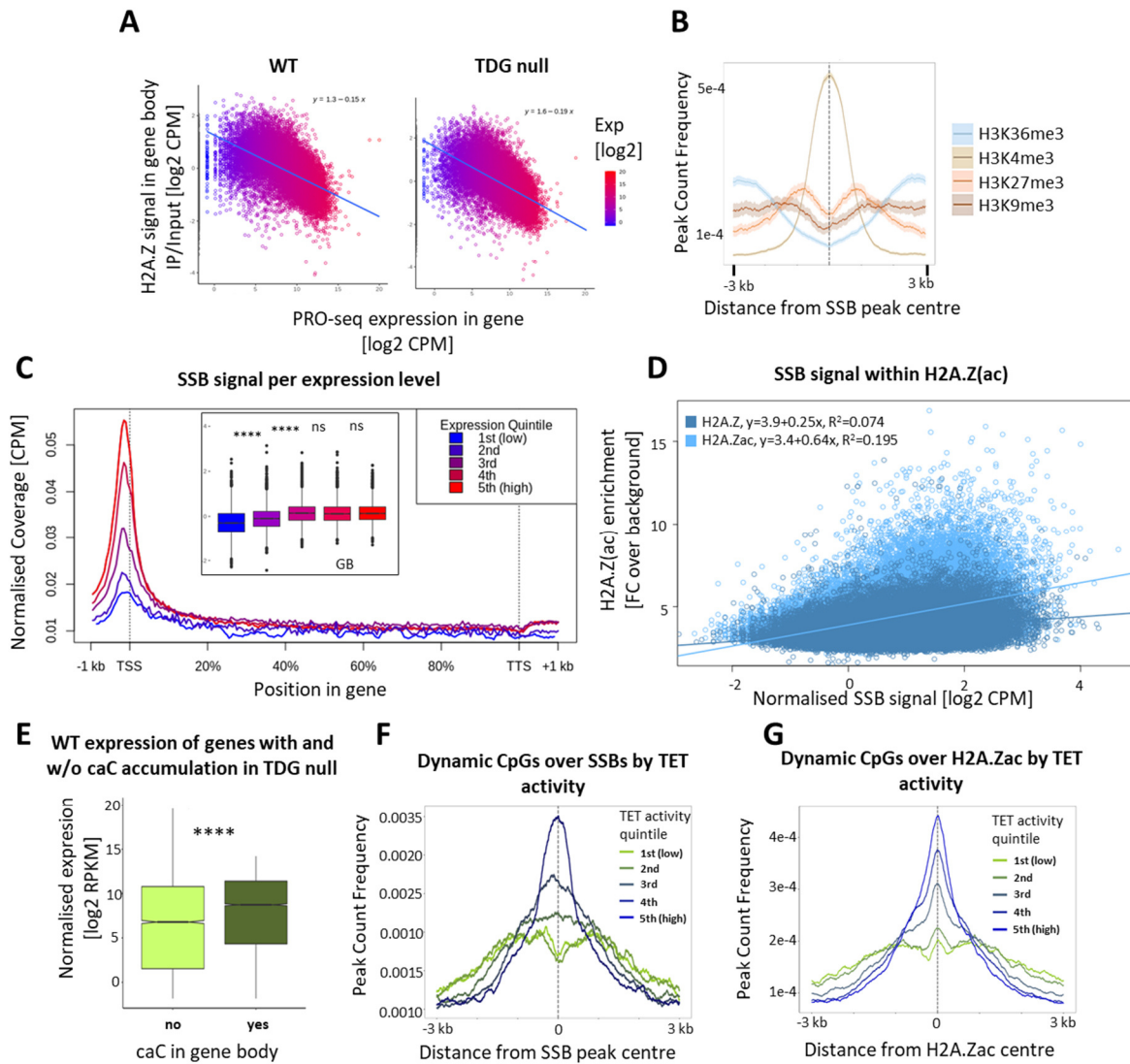
**Supplementary Figure S1 TDG depletion leads to a decreased pause release of RNAP2**

A Number of successfully mapped reads in PRO-seq per individual sample **B** PCA analysis of PRO-seq samples **C** Venn diagram of detected transcripts in total RNA-seq and PRO-seq **D** Correlation of expression levels of common detected transcripts in total RNA-seq and PRO-seq. **E** Venn diagram of upregulated genes ($\log_2FC \geq 1.5$) in totalRNA-seq and PRO-seq upon 5 nM Tal for 24 h. **F** PI of all genes detected in PRO-seq **G** Normalised PRO-seq reads at transcription start site (TSS) and in the gene body (GB) of all detected genes **H** Metagene analysis of Tal-upregulated genes defined from PRO-. Asterisks indicate statistical p-values of: *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$



Supplementary Figure S2 H2A.Z accumulation in TDG deficient ESCs affects pause release

A Venn diagram with the number of H2A.Z and H2A.Zac-enriched regions from Hu et al. 2013. **B** Accessibility (ATAC-seq) within regions of H2A.Z enrichment (light green) and H2A.Zac enrichment (dark green) **C** Accessibility (ATAC-seq) within H2A.Z(ac) enriched regions in WT and TDG null ESCs. Asterisks indicate statistical p-values of: *:p≤0.05, **:p≤0.01, ***:p≤0.001, ****:p≤0.0001







Supplementary Figure S3 Sites of active DNA demethylation correlate with H2A.Z variants and transcription

A Corellation of H2A.Z abundance and GB transcription in WT and TDG null ESCs **B** Density plot of indicated histone variants (Marks et al 2012) over SSBs in WT ESCs **C** Metagenome analysis of SSBs stratified by the expression. Inset: Quantification of SSBs in gene body **D** SSB signal within regions of H2A.Z enrichment (dark blue) and H2A.Zac enrichment (light blue) **E** Expression of genes that display caC accumulation in their gene body upon the depletion of TDG («yes») or not («no»). **F** Density plot of dynamic CpGs over SSBs in WT ESCs **G** Densitiy plot of dynamic CpGs over H2A.Zac by Tet activity. Asterisks indicate statistical p-values of: *:p≤0.05, **:p≤0.01, ***:p≤0.001, ****:p≤0.0001



METHOD ARTICLE

REVISED Inducible TDG knockout models to study epigenetic regulation [version 2; peer review: 3 approved]

Simon D. Schwarz ¹, Eliane Grundbacher¹, Alexandra M. Hrovat ¹, Jianming Xu¹, Anna Kuśnierczyk², Cathrine B. Vågbø², Primo Schär ¹, David Schuermann ¹

¹Department of Biomedicine, University of Basel, Basel, 4058, Switzerland

²Proteomics and Modomics Experimental Core Facility (PROMEC), Norwegian University of Science and Technology, Trondheim, 7491, Norway

V2 First published: 09 Sep 2020, 9:1112
<https://doi.org/10.12688/f1000research.25637.1>

Latest published: 19 Oct 2020, 9:1112
<https://doi.org/10.12688/f1000research.25637.2>

Abstract

Mechanistic and functional studies by gene disruption or editing approaches often suffer from confounding effects like compensatory cellular adaptations generated by clonal selection. These issues become particularly relevant when studying factors directly involved in genetic or epigenetic maintenance. To provide a genetic tool for functional and mechanistic investigation of DNA-repair mediated active DNA demethylation, we generated experimental models in mice and murine embryonic stem cells (ESCs) based on a minigene of the thymine-DNA glycosylase (TDG). The *loxP*-flanked *miniTdg* is rapidly and reliably excised in mice and ESCs by tamoxifen-induced Cre activation, depleting TDG to undetectable levels within 24 hours. We describe the functionality of the engineered *miniTdg* in mouse and ESCs (TDGiKO ESCs) and validate the pluripotency and differentiation potential of TDGiKO ESCs as well as the phenotype of induced TDG depletion. The controlled and rapid depletion of TDG allows for a precise manipulation at any point in time of multistep experimental procedures as presented here for neuronal differentiation *in vitro*. Thus, we provide a tested and well-controlled genetic tool for the functional and mechanistic investigation of TDG in active DNA (de)methylation and/or DNA repair with minimal interference from adaptive effects and clonal selection.

Keywords

Embryonic Stem Cells, TDG, Active DNA Demethylation, Base Excision Repair, Neuronal Differentiation, Minigene, Cre/loxP, Tamoxifen

Open Peer Review

Reviewer Status 

	Invited Reviewers		
	1	2	3
version 2 (revision) 19 Oct 2020			 report
version 1 09 Sep 2020	 report	 report	 report

1. **Jonathan Sczepanski**, Texas A&M University, College Station, USA

2. **Joseph Torchia**, Western University, London, Canada

3. **Sebastian Bultmann** , Ludwig-Maximilians-Universität München, Munich, Germany

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding authors: Primo Schär (primo.schaer@unibas.ch), David Schuermann (david.schuermann@unibas.ch)

Author roles: **Schwarz SD:** Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Grundbacher E:** Investigation, Methodology, Writing – Review & Editing; **Hrovat AM:** Investigation, Writing – Review & Editing; **Xu J:** Investigation, Writing – Review & Editing; **Kuśnierczyk A:** Investigation, Resources, Writing – Review & Editing; **Vågbø CB:** Investigation, Resources, Writing – Review & Editing; **Schär P:** Conceptualization, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Review & Editing; **Schuermann D:** Conceptualization, Formal Analysis, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Swiss National Science Foundation, grant 156467.

Copyright: © 2020 Schwarz SD *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Schwarz SD, Grundbacher E, Hrovat AM *et al.* **Inducible TDG knockout models to study epigenetic regulation [version 2; peer review: 3 approved]** F1000Research 2020, 9:1112 <https://doi.org/10.12688/f1000research.25637.2>

First published: 09 Sep 2020, 9:1112 <https://doi.org/10.12688/f1000research.25637.1>

REVISED Amendments from Version 1

In response to the reviewers' comments and suggestions, the article has been revised as follows:

The integration of the *miniTdg* transgene in a random genomic location was clarified.

Indicators of statistical significance and extended description of the statistical approaches were added to all figures and respective legends.

Off-target analyses for the two ESC clones generated by CRISPR/Cas9 were added to the *Extended data*.

Clarifications and text improvement were made according to the comments and questions raised by the reviewers, referring to the generation and use of the engineered system in cultured cells and animals as well as to technical details.

Typos and ambiguous phrasing were corrected throughout the article.

Any further responses from the reviewers can be found at the end of the article

Introduction

To allow for the differentiation to a multitude of cell types during the development of multicellular organisms, stem cells need to respond to a variety of extrinsic and intrinsic developmental cues and integrate these into epigenetically stabilized gene expression patterns during lineage specification. The underlying mechanisms to ensure this necessary epigenetic plasticity are therefore active and highly dynamic in embryonic stem cells (ESC). One of those mechanisms includes the thymine DNA-glycosylase (TDG) and the downstream factors of DNA base excision repair (BER), which were shown to be involved in the dynamic, locus-specific regulation of DNA methylation. Following the activity of dioxygenases of the ten-eleven-translocation (TET) family that iteratively oxidize 5-methylcytosine (5-mC) to formyl- and carboxylcytosine (5-fC, 5-caC), TDG-induced BER provides a mechanism of active DNA demethylation and, thereby, impact gene regulation (Cortázar *et al.*, 2011; Cortellino *et al.*, 2011; He *et al.*, 2011; Ito *et al.*, 2011; Schuermann *et al.*, 2016; Shen *et al.*, 2013; Weber *et al.*, 2016). Studying the precise function of TDG (and BER) in epigenetic regulation during differentiation of stem cells, however, is challenging due to its multiple and interwoven interactions with other proteins like transcription factors and epigenetic modifiers (Henry *et al.*, 2016; Jacobs & Schär, 2012; Léger *et al.*, 2014), the deregulation of which may cause a plethora of confounding effects. Also, genetic depletion of TDG itself is not compatible with embryonic development (Cortázar *et al.*, 2011; Cortellino *et al.*, 2011). To reduce and circumvent such confounders in functional studies of TDG (and BER) in murine ESCs, we established and validated a *Tdg* minigene, introduced it into mice and derived a series of different ESC variants. The minigene carries, among other features, a *loxP*-flanked coding region of *Tdg*, allowing for a controlled and fast, 4-hydroxytamoxifen (OHT)-inducible depletion of TDG in the background of a homozygous disruption of the endogenous *Tdg*. By eliminating phenotypic divergence between TDG proficient and deficient ESCs, this versatile model facilitates precise

functional and mechanistic investigations into TDG-mediated active DNA demethylation. This novel *Tdg* minigene introduced into mice as well as ESCs will facilitate future research in the field of active DNA demethylation *in vivo* and *in vitro*.

Methods**Ethics statement and Animal Work**

All animal work was carried out in accordance with the Swiss Animal Welfare Act and the guidelines of the Swiss Federal Veterinary Office (SFVO) or with the UK Animals (Scientific Procedures) Act. Housing, breeding and experimentation of mice has been performed with the approval of the Cantonal Veterinary Office of Basel-Stadt (Licenses 10062-H, 1912) or was covered by a project license to Wolf Reik (80/1896) further regulated by the Babraham Institute Animal Welfare, Experimentation, and Ethics Committee. Mice were housed under specific pathogen free condition in individually ventilated cages under 12 h light/dark cycles at 22 ±2°C and 35–65% relative humidity. Animal and colony health was checked daily and quarterly, respectively. Sterilized diet (Kliba-Nafag Extrudate 3436) and water was provided *ad libitum*.

Tdg^{tm1Psch} mice (<http://www.informatics.jax.org/allele/MGI:5487834>) were generated from E14 ESC (RRID:CVCL_C320) and backcrossed for more than 10 generations with female C57BL/6JRj mice, obtained at 7–8 weeks of age from Janvier Labs. A *Tdg* expression cassette (pTCO2-mTDGi.0, Addgene Plasmid #149429) was introduced as single copy into C57BL/6JRj mice by pronuclear injection and crossed into *Tdg^{tm1Psch}* mice. Genotyping was done by PCR with the HOT FIREPol (Solis BioDyne) on diluted crude DNA preparations (diluted by a factor of five with 10 mM Tris-HCl pH8) by boiling toe clips with 25 mM NaOH/0.2 mM EDTA for 1 h followed by neutralization with 40 mM Tris-HCl. Reactions were performed according to the manufacturer's recommendation with 35 PCR cycles at 95°C for 30 s, 60°C for 30 s, and at 72°C for 60 s (for details about all used primers and annealing temperatures in PCR reactions, see *Extended data* Table S1) (Schwarz, 2020).

Generation of ESCs

ESCs were derived from male blastocysts from timed matings of male *miniTdg^{tg/tg}//Tdg^{-/-}* mice with 2 super-ovulated female *Tdg^{+/-}* mice and cultured on murine immortalized feeder cells in 2i medium with leukemia inhibitory factor LIF (Merck-Millipore) (Ying *et al.*, 2008).

To introduce the inducible Cre recombinase into the ROSA26 locus, 30 µg of the targeting vector (pROSA26-ERT2CreERT2, Addgene Plasmid #149436) was electroporated into 15×10⁶ mESCs using a gene pulser Xcell (Bio-Rad) at 240 V and 475 µF. Transgenic cells were selected with 8 and 5 µg/ml blasticidin for a week each, before colonies were picked and then amplified without selection pressure. From blasticidin-resistant colonies, genomic DNA was extracted with the "QIamp DNA mini" kit (Qiagen) and screened for a targeted integration at the 3' junction of the ROSA26 locus by PCR with the Phusion polymerase (NEB) (see *Extended data*, Table S1) (Schwarz, 2020), applying a general three-step cycling protocol: initial denaturation 95°C for 30 s, 35 times 95°C for 10 s, annealing for

20 s, 72°C for 15–80 s. PCR reactions were analyzed by gel electrophoresis followed by image acquisition with a U:Genius 3 system (Version 3.0.12.0, Syngene).

To evoke a disruption of the *Neo^R* reading frame by CRISPR/Cas9, two guide RNAs((Neo^R)3: 5'-GCCGATCCCATATTGGC TGCAGG-3', (Neo^R)4: 5'-GAAGGCGATGCGCTGCGAATCG G-3', IDT) were designed and transfected with *TransIT-X2* (Mirus Bio) into the TDGiKO1 ESCs. RNP assembly was performed according to the manufacturers protocol (IDT). One day after transfection, single cells were sorted with a FACSaria IIIu (BD BioSciences) into 96-well plates coated with inactivated mouse fibroblasts, based on GFP expression from a co-transfected pEGFP-N1 plasmid (Clontech). Screening for the successful disruption of the *Neo^R* gene was done by PCR as described above and assaying the loss of the neomycin resistance with the Cell Counting Kit-8, according to the manufacturer's instructions (CCK-8, Dojindo).

Cell culture and differentiation

ESCs were cultured under a controlled atmosphere (37°C, 5% CO₂, 95% humidity) in serum-free 2i medium with 1000 U/ml LIF, without antibiotics unless indicated otherwise. The 4-OHT (Sigma-Aldrich H7904) was dissolved in DMSO (stock 10 mM) and administered at indicated concentrations for 2 h.

Neuronal differentiation was performed as described before (Bibel *et al.*, 2007; Cortázar *et al.*, 2011; Steinacher *et al.*, 2019). Differentiation towards cardiomyocytes was performed by culturing ESCs in classical ESC medium (ESM) with 15% FCS without LIF in non-adhesive petri dishes (Greiner) for at least 10–14 days.

All pictures were taken with a DM IL LED Fluo microscope and MC170 HD camera (Leica) at 100x magnification.

Molecular analysis of ROSA26 targeting and *miniTdg* excision

Probes for Southern blotting were generated by PCR amplification with Taq DNA Polymerase (Qiagen) from the ROSA26 targeting plasmid using the DIG DNA labeling mix (Roche). Southern blotting was done with 20 µg of genomic DNA, extracted by the “Genomic tip 100G” kit (Qiagen). Probes were hybridized according to the “DIG Application Manual for Filter Hybridization” by Roche. Signals were acquired by exposing the membrane to chemiluminescence detection films (Amersham/GE Healthcare) and digitalized by a CanoScan 8400F scanner with CanoScan Toolbox (Version 4.9.3).

Excision of the *miniTdg* gene by Cre was assessed by qPCR with the “Rotor-Gene SYBR Green PCR Kit” (Qiagen) on a Rotor Gene 3000 (Qiagen) system. After 95°C for 5 min, qPCR reactions followed a two-step cycling (40x) protocol: 95°C for 5 s, 60°C for 30 s, terminated by ramping from 60 to 95°C in 175 s to generate a melting curve. PCRs to exclude background recombination by Cre were done with NEB Phusion polymerase as described above.

Gene expression analyses

Total RNA was extracted with the RNAeasy Mini Kit (Qiagen), including on-column DNase digest, and reverse transcribed with RevertAid First Strand cDNA synthesis kit (ThermoFisher) using oligo-dT primers. qPCR was performed as described above with marker-specific primers (see *Extended data*, Table S1) (Schwarz, 2020).

Protein levels were analyzed by western blotting of 50 µg of NP-40 (or SDS) whole-cell extracts, separated by SDS-PAGE, onto a nitrocellulose membrane (Amersham) and immunodetection with anti-mTDG antibody (rabbit polyclonal L58, Lab P.Schär, dilution 1:10'000, 1 h at 33°C) and anti-GAPDH antibody (Sigma-Aldrich Cat# G9545, RRID:AB_796208 dilution 1:20'000, 1 h at RT). Chemiluminescence signals were detected with a Fusion FX7 system (Software version 16.15.0.0, Vilber).

Detection of oxidized variants of 5-methylcytosine

DNA was extracted and purified with the Genomic tip 100G Kit (Qiagen). High-performance liquid chromatography–tandem mass spectrometry analysis was performed on 10 µg of genomic DNA as described before (Weber *et al.*, 2016).

Statistical analysis

To test for statistical significance ($p \leq 0.05$), we performed two-tailed Student's *t*-tests on the indicated number of replicates in Microsoft Excel (Version 16.0.4954.1000)

Results and discussion

Generation of a chimeric *Tdg* minigene and derivation of murine ESCs carrying a tamoxifen-inducible Cre recombinase

To facilitate genetic manipulation of the *Tdg* gene (~28 kb including the putative promoter), we first constructed a synthetic *Tdg* minigene (*miniTdg*, ~11 kb) (Figure 1A). The minigene consists of the endogenous *Tdg* promoter and terminator sequences flanking the *Tdg* coding sequence (CDS). It also includes the rabbit β-globin intron at the authentic position of the first intron of the *Tdg* transcript variant 2 (Genebank NM_172552.4). This intron allows for the expression of the two naturally occurring *Tdg* splice variants (Um *et al.*, 1998). In addition, the *miniTdg* contains two *loxP* sites in the same orientation for Cre-recombinase-mediated excision of the coding region. A plasmid carrying the *miniTdg* gene (pTCO2-mTDGi.0) was introduced into C57BL/6 mice at a random genomic location by pronuclear injection and offspring carrying the minigene were crossed into heterozygous *Tdg^{tm1Psch}* mutant mice (Kunz *et al.*, 2009). When breeding heterozygous mice for the *miniTdg* transgene and the *Tdg* KO allele, the genotypes of offspring followed the expected Mendelian pattern for two loci (Figure 1B). The *miniTdg* allele segregated with a 3:1 ratio, consistent with a random integration into a single genomic locus. About 25% of born mice with the *miniTdg* gene were homozygous for the *Tdg^{tm1Psch}* KO allele while no homozygous *Tdg* KO mouse was obtained without the transgene. This shows that the *miniTdg* transgene is fully

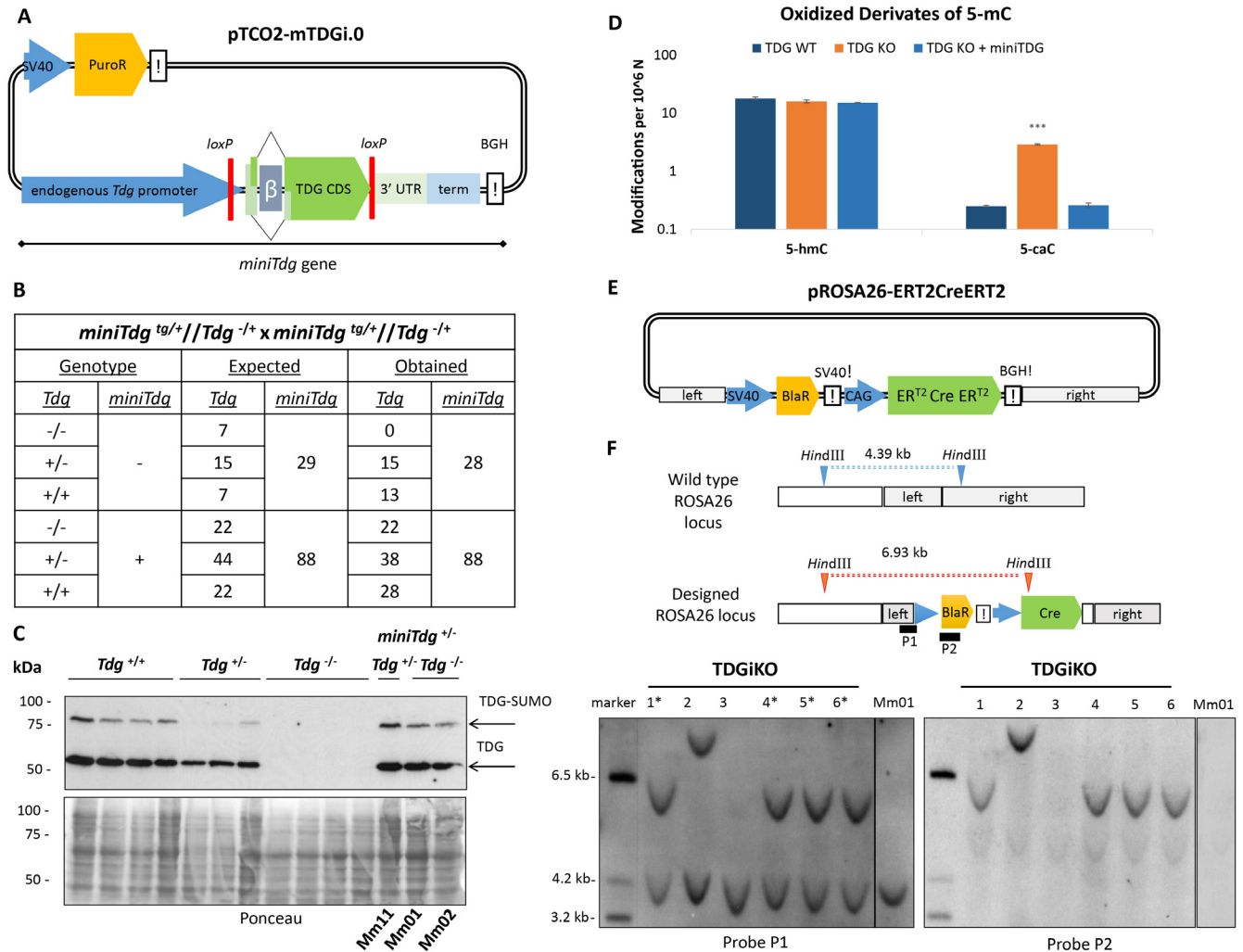


Figure 1. *miniTdg* complementation of *Tdg*^{-/-} mice and mESCs. (A) Synthetic *miniTdg* gene on the plasmid integrated to a random genomic locus by pronuclear injection. *miniTdg* consisting of 6 kb endogenous *Tdg* promoter, the *Tdg* coding sequence (CDS) including a chimeric splice donor (*Tdg* exon 1 of transcript variant 2 NM_172552.4, rabbit β -globin intron), to allow for alternative splicing, 3 kb of the 3' untranslated region (3' UTR) and terminator (term) of *Tdg*, and the bovine growth hormone terminator sequence (BGH!). *loxP* sites are indicated in red. (B) Expected and obtained genotype distribution in offspring (n=116) from crosses of heterozygous mice. (C) Top: Immunoblot for TDG of whole-cell SDS-extracts from derived ESCs with different genotype. Bottom: Loading control by Ponceau staining of the membrane. (D) Accumulation of oxidized 5-mC derivatives measured by HPLC-MS/MS in ESCs with the indicated genotype. For each genotype, means and standard deviations of two independent clones with 1 or 2 repeated measurements are shown. Asterisks indicate *p*-values of Student's *t*-test, compared to TDG WT: **p*-value < 0.001. (E) Scheme of the targeting construct pROSA26-ERT2CreERT2 for the ROSA26 locus with the tamoxifen-inducible Cre recombinase (Matsuda & Cepko, 2007). ERT2-Cre-ERT2 expression is under the control of the synthetic cytomegalo-virus/chicken-actin/ β -globin promoter (CAG). Homology arms for ROSA26 targeting are indicated in light grey. A blasticidin resistance cassette (BlaR) for positive selection is under the control of the SV40 promoter and terminator. (F) Predicted restriction pattern of ROSA26 WT and integration events (top) and fragments detected by Southern blotting (bottom). Left: The hybridization probe (P1) locating to the left ROSA26 homology arm detected a genomic fragment of 6.9 kb. The 4.4 kb fragment represents the wild-type ROSA26 allele. Right: Detection of possible off-target events using a second probe (P2) locating to the blasticidin selection marker within the targeting construct. See *Underlying data* for the raw data behind this figure (Schwarz, 2020).**

functional in complementing the developmental defects and lethality of homozygous *Tdg*^{tm1Psch}KO embryos (Cortázar *et al.*, 2011; Cortellino *et al.*, 2011). *MiniTdg* complemented mice were healthy and fertile with normal lifespan.

Three ESC populations were then derived from male blastocysts from crosses of *miniTdg*^{tg/tg}//*Tdg*^{-/-} with *Tdg*^{+/-} mice (Extended data, Table S2: Mm11, Mm01, Mm02) (Schwarz, 2020). In these

ESCs, TDG protein levels were similar to those in wild-type ESCs (Figure 1C), indicating that the *miniTdg* gene is regulated as the endogenous *Tdg*. Furthermore, no silencing of the transgene was noticed over several generations of animal breeding or culturing of ESCs. To assess functional integrity at the molecular level, we measured the accumulation of oxidized 5-methylcytosine (5-mC) derivatives, 5-hydroxymethylcytosine (5-hmC) and 5-carboxylcytosine (5-caC), the substrate for

excision by TDG (Maiti & Drohat, 2011; Shen *et al.*, 2013) (Figure 1D). Mass spectrometry showed that TDG deficient ESCs accumulated 5-caC as expected, while ESCs expressing TDG solely from the *miniTdg* gene showed 5-caC levels similar to those in wild-type ESCs.

To allow for a controlled excision of the *miniTdg* coding region, we introduced a tamoxifen-inducible Cre recombinase expression cassette (Figure 1E) into the non-essential ROSA26 locus (Soriano, 1999). Screening 60 blasticidin-resistant colonies by PCR for targeted integration at 3' ROSA26 locus yielded six clones (TDGiKO1-6). Southern blotting confirmed correct heterozygous integration of the complete ER^{T2}-Cre-ER^{T2} cassette (Matsuda & Cepko, 2007) for four of the six clones (Figure 1F, left, asterisks). Potential additional off-target integrations were tested and excluded by hybridizing a second probe within the transgene (Figure 1F, right). Original Southern blotting images are available as *Underlying data* (Schwarz, 2020).

Taken together, we established a *miniTdg* gene that complements expression and function of the endogenous *Tdg* gene *in vivo* and *in vitro*. The functional integrity of the transgene is confirmed by rescuing phenotypes of *Tdg* KO defects at the systemic and cellular level, namely the rescue of embryonic lethality and suppression of 5-caC accumulation. The introduction of a Cre-recombinase to the ROSA26 locus, without off-targeted events, will allow conditional inactivation of the *miniTdg* in ESCs.

In addition, we generated ESC clones that ease further genetic manipulations involving selectable marker genes. We eliminated the neomycin resistance gene, which was introduced at the endogenous *Tdg* locus to generate the original *Tdg* KO allele (Kunz *et al.*, 2009), by applying a CRISPR-guided Cas9 nuclease. Using two guide RNAs targeting the 5' and 3' end of the *Neo^R* gene, we generated two ESC clones (TDGiKO1.1 and TDGiKO1.2) that bear a deletion of 776 and 767 bp, respectively, and are no longer resistant to neomycin (*Extended data*, Figure S1A-C) (Schwarz, 2020).

Induction of the Cre recombinase leads to fast and efficient depletion of TDG in ESCs

To make the inducible TDG depletion (Figure 2A) applicable to differentiation experiments that often depend on complex and strict cell culture procedures, we aimed at establishing a short Cre-induction procedure minimizing any interference with standard differentiation conditions. Titrating 4-Hydroxytamoxifen (OHT) and measuring *miniTdg* excision by quantitative PCR, we identified OHT concentrations of 1-5 μ M in serum-free 2i (Figure 2B) and 5-10 μ M OHT in serum-containing medium (Figure 2C) to yield maximal excision efficiencies without impairing cell viability (Figure 2D). We note that lower OHT doses (down to 100 nM) for longer duration (up to 48 h) may be applied as well, depending on the experimental procedure and objective. While the *miniTdg* excision product is readily detectable by PCR in different TDGiKO ESCs after OHT treatment, no unintentional background activity of the Cre recombinase is detectable before OHT induction

(Figure 2E). Following the kinetics of TDG depletion by SDS-PAGE/immuno-detection, we found that TDG was reduced below detection limit (<2% of endogenous TDG) 24 h after Cre induction for 2 h (Figure 2F and *Extended data*, Figure S1D) (Schwarz, 2020).

We also assessed the tissue specific induction of Cre-mediated *miniTdg* excision *in vivo*. We crossed *miniTdg* mice with mice expressing Cre-recombinase under the control of a *FoxN1*-promoter (<http://www.informatics.jax.org/allele/key/62500>) which drives expression in thymic epithelial cells (TECs), but not in thymic cells (TCs). We observed that the mRNA and protein expressed from the *miniTdg* are reliably depleted in TECs expressing *FoxN1*-driven Cre, whereas TDG levels are much less affected in TCs (*Extended data*, Figure S1E/F) (Schwarz, 2020).

These data demonstrate the functionality and tight regulation of the inducible *miniTdg* KO model in murine ESCs as well as in mice. A short-time Cre induction with 1–5 μ M or 5–10 μ M OHT (in 2i+LIF or 15% FCS-containing medium, respectively) is sufficient to mediate excision of the *miniTdg*, resulting in depletion of TDG below detection within 24 hours. This rapid depletion of TDG is presumably facilitated by the cell cycle-regulated proteasomal degradation of TDG via the ubiquitination at its PCNA-interacting peptide (PIP) motif (Hardeland *et al.*, 2007; Shibata *et al.*, 2014).

Molecular and cellular phenotypes of TDG depletion in TDGiKO ESCs

To validate the molecular phenotype of TDG depletion in our newly derived TDGiKO ESCs (TDGiKO1 ESCs and derived *Neo^R*-deleted TDGiKO1.1), we again assessed changes in levels of oxidized 5-mC derivatives, 5-formylcytosine (5-fC) and 5-caC following Cre-mediated *miniTdg* deletion. One week after induction, we measured a 4.4 and 4.2-fold accumulation of the TDG substrates 5-fC and 5-caC in TDG-depleted TDGiKO ESCs, respectively, but no changes of 5-hydroxymethyl cytosine (5-hmC), which is not recognized by TDG (Figure 3A). This accumulation reflects well the previously reported measurements in constitutive *Tdg* KO cells (Shen *et al.*, 2013; Steinacher *et al.*, 2019).

Finally, we re-tested the pluripotency and differentiation capacity of some of the engineered TDGiKO ESC clones (*Extended data*, Table S2) (Schwarz, 2020), applying a protocol for all-trans retinoic acid (RA) induced *in vitro* differentiation towards the neuronal lineage (Bibel *et al.*, 2007) with minor adaptations. TDG depletion was induced just before an additional culturing period of 16 h in serum containing medium, which promotes the transition from the ground (cultured in 2i medium) to the primed ESC state (Figure 3B). Formation of embryoid bodies and neural progenitor cells was triggered by LIF withdrawal and later the addition of RA, before terminal differentiation was induced by replating the cells in media for neural cells. When subjecting TDG proficient TDGiKO1 ESCs to neural differentiation, we observed an upregulation of the neural marker *Nestin* (Figure 3B, C) and, ultimately, the formation of neuron-like cells (Figure 3B). As expected during embryoid body (EB) formation,

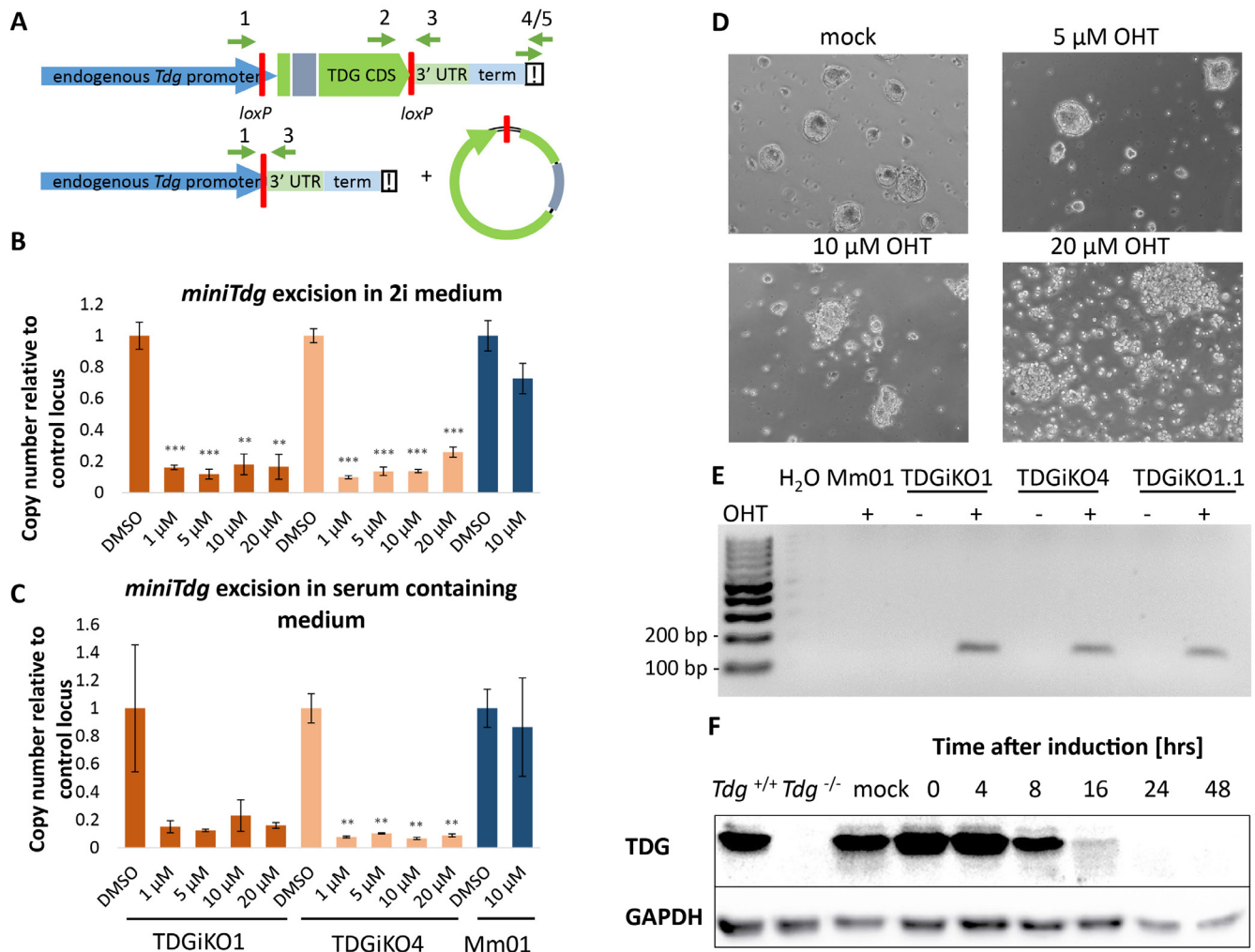


Figure 2. Cre induction for two hours results in the depletion of TDG within 24 hours without affecting cell viability. (A) Scheme of the *miniTdg* cassette before (top) and after (bottom) Cre-mediated recombination. Position of PCR primers are indicated with green arrows. (B) Quantitation of *miniTdg* excision upon OHT administration in TDGiKO1, TDGiKO4 and parental ESCs without the Cre recombinase (Mm01). Genomic DNA of cells was extracted two days after OHT treatment with indicated concentrations for 2 hours. Copy number of the *miniTdg* is measured by qPCR using primers 2 and 3, normalized to the nearby terminator region (primers 4/5) and a control locus on chromosome two. Shown are means and standard error from technical quadruplicates per clone. Asterisks indicate *p*-values of Student's *t*-test, compared to DMSO control: ** *p*-value < 0.01, ****p*-value < 0.001. (C) Quantitation of *miniTdg* excision as in (B), but Cre induction executed in ES medium containing 15% FCS. (D) Phase-contrast images of TDGiKO1 ESCs in 2i medium, two days after OHT treatment for two hours at indicated concentrations. (E) Detection of Cre recombination events (158 bp) by PCR using primers 1 and 3 (A) before and after addition of OHT. (F) Time-course assessment of TDG protein levels expressed from the *miniTdg* gene. Immunodetection with an anti-TDG antibody was performed with western blotted NP-40 cell extracts of 2i cultivated TDGiKO1 ES cells after Cre induction by 1 μM OHT for 2 h. See *Underlying data* for the raw data behind this figure (Schwarz, 2020).

LIF withdrawal resulted in an induction of lineage-specific genes of all three germ layers and the repression of the pluripotency marker gene *Nanog*. This indicates the ability of the TDG proficient TDGiKO ESCs to commit to cell types of all germ layers. In support of this, continued cultivation of non-induced TDGiKO1 and TDGiKO1.1 in ESM without LIF led to a spontaneous differentiation to cardiomyocytes (mesodermal cell type) forming beating organoids (see *Extended data Videos 1 and 2*) (Schwarz, 2020). While the early initiation of lineage-specific

genes seemed to be unaffected by TDG depletion prior to differentiation, the formation of neuron-like cells (Figure 3B, bottom) was impaired. This deficiency to stably commit to and/or maintain the neuronal lineage was indicated by a low number of neuron-like cells formed and the loss of cells with *Nestin* expression following RA application. Upon loss of TDG, we furthermore observed a significant difference in the down-regulation of the pluripotency marker *Nanog* as well as and dysregulation of the endodermal marker *Gata6* (Figure 3C).

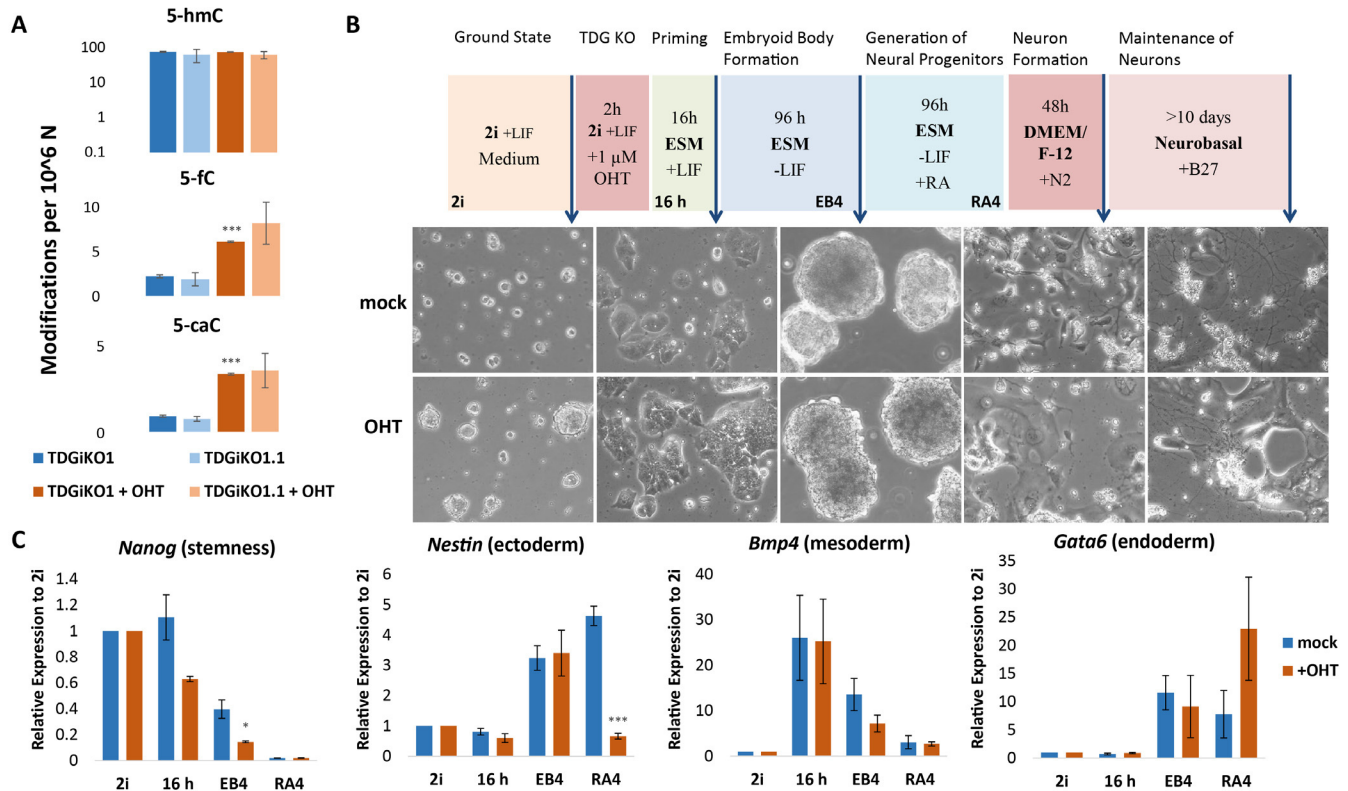


Figure 3. TDGiKO ESCs express a functional TDG and are proficient for cell lineage commitment. (A) Measurement of oxidized derivatives of 5-mC in selected ESCs by HPLC-MS/MS, cultivated in 2i medium + LIF, one week after TDG depletion by 1 μM of OHT for 2 h. Shown are the means and standard deviation of 2-3 biological replica per condition. (B) Scheme of the neural differentiation protocol (Bibel *et al.*, 2007), adapted with a 16 h priming step in ESC medium for ESCs grown in 2i medium. Representative phase-contrast images of TDGiKO1 cells with (+OHT) and without (mock) TDG depletion at the differentiation stages indicated (arrows). (C) Expression of pluripotency and germ layer marker genes were assessed by qRT-PCR in TDGiKO1 cells at the stages of differentiation as marked in (B). Expression was normalized to the housekeeping genes *Rps13* and *Eef1a1*. Shown are means and standard errors of 3 biological replicates. Asterisks indicate *p*-values of Student's *t*-test, compared to the non-induced condition: **p*-value < 0.05, ****p*-value < 0.001. See *Underlying data* for the raw data behind this figure (Schwarz, 2020).

Thus, when subjecting the TDG-depleted TDGiKO ESCs to *in vitro* differentiation, we obtained data consistent with previous observations in constitutive *Tdg* KO ESCs (Steinacher *et al.*, 2019). This highlights the usability of the newly established ESCs in TDG and BER-related research, with the advantage of eliminating clonal divergences between TDG proficient and deficient cells. While here, we focused on the reproduction of published *Tdg* KO phenotypes, the fast and reliable depletion of TDG in the TDGiKO ESC model now provides a tool for in-depth execution point analysis of TDG-BER mediated active DNA demethylation during cell differentiation or in any other context of interest. Furthermore, the *loxP* sites flanking the entire coding region of the *miniTdg* gene enable recombinase-mediated cassette exchange to introduce TDG separation of function variants devoid of catalytic function or harbour mutations that ablate posttranslational modifications (Hardeland *et al.*, 2000). Such an exchange with functional TDG variants or fusion proteins is not easily applicable with other available murine conditional *Tdg* knockout systems (e.g.: Cortellino *et al.*, 2011; Hassan *et al.*, 2017; Hu *et al.*, 2014; Song *et al.*, 2013). Additionally, the availability of mice carrying the *miniTdg*, which can be

combined with Cre-driver mice of interest by crossing, allows the translation of *in vitro* observations into *in vivo* settings. All in all, we provide a versatile, well-controlled and validated toolbox (Extended data, Table S2) (Schwarz, 2020) to facilitate functional and mechanistic research on TDG-mediated active DNA demethylation and BER in cell culture and animal models.

Data availability

Underlying data

Zenodo: Inducible TDG knockout models to study epigenetic regulation. <http://doi.org/10.5281/zenodo.4075154> (Schwarz, 2020).

This project contains the following underlying data:

- Raw_Data_Figure1.zip. Database excerpt for genotyping of mouse crossings, original pictures of Southern and western blots and MS-quantification of oxidized mC derivatives (also for figure 3).
- Raw_Data_Figure2.zip. qPCR data of *miniTdg* excision, original pictures of OHT-treated ESCs and western blots.

- Raw_Data_Figure3.zip. Original pictures of differentiating cells and qPCR data of pluripotency and germ line markers.
- Raw_Data_Extended_Figure.zip. Sequencing data for *Neo^R* excision, and predicted off-targets, absorption values of G418-treated ESCs, qPCR data for *miniTdg* excision and original pictures of western blots.

Extended data

Zenodo: Inducible TDG knockout models to study epigenetic regulation. <http://doi.org/10.5281/zenodo.4075154> (Schwarz, 2020).

This project contains the following extended data:

- [Extended Figure and Tables.pdf](#). Supplementary information as referenced in the text.
- [Extended_Video_1_TDGiKO1.mp4](#). Video of “beating bodies” from an ESC clone.

- [Extended_Video_2_TDGiKO1.2.mp4](#). Video of “beating bodies” from a second ESC clone.

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

Acknowledgments

We want to express our thanks to the laboratory of Wolf Reik and the transgenics facility at the Babraham Institute (Cambridge, UK) for performing the pronuclear injection. We also thank the animal facility at the DBM Mattenstrasse and the transgenic core facility of the University of Basel for their help in breeding the mouse strains and establishing ESCs. Additional thanks goes to Carlos Mayer from George Holländer’s Lab (DBM) for the sorting of the thymic cells.

References

- Bibel M, Richter J, Lacroix E, *et al.*: **Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells.** *Nat Protoc.* 2007; **2**(5): 1034–1043.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cortázar D, Kunz C, Selfridge J, *et al.*: **Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability.** *Nature.* 2011; **470**(7334): 419–423.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Cortellino S, Xu J, Sannai M, *et al.*: **Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair.** *Cell.* 2011; **146**(1): 67–79.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hardeland U, Bentele M, Jiricny J, *et al.*: **Separating substrate recognition from base hydrolysis in human thymine DNA glycosylase by mutational analysis.** *J Biol Chem.* 2000; **275**(43): 33449–33456.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hardeland U, Kunz C, Focke F, *et al.*: **Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2.** *Nucleic Acids Res.* 2007; **35**(11): 3859–3867.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hassan HM, Kolendowski B, Isovic M, *et al.*: **Regulation of Active DNA Demethylation through RAR-Mediated Recruitment of a TET/TDG Complex.** *Cell Rep.* 2017; **19**(8): 1685–1697.
[PubMed Abstract](#) | [Publisher Full Text](#)
- He YF, Li BZ, Li Z, *et al.*: **Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA.** *Science.* 2011; **333**(6047): 1303–1307.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Henry RA, Mancuso P, Kuo YM, *et al.*: **Interaction with the DNA Repair Protein Thymine DNA Glycosylase Regulates Histone Acetylation by p300.** *Biochemistry.* 2016; **55**(49): 6766–6775.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hu X, Zhang L, Mao SQ, *et al.*: **Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming.** *Cell Stem Cell.* 2014; **14**(4): 512–22.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ito S, Shen L, Dai Q, *et al.*: **Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.** *Science.* 2011; **333**(6047): 1300–1303.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jacobs AL, Schär P: **DNA glycosylases: In DNA repair and beyond.** *Chromosoma.* 2012; **121**(1): 1–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Kunz C, Focke F, Saito Y, *et al.*: **Base excision by thymine DNA glycosylase mediates DNA-directed cytotoxicity of 5-fluorouracil.** *PLoS Biol.* 2009; **7**(4): e91.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Léger H, Smet-Nocca C, Attmane-Elakeb A, *et al.*: **A TDG/CBP/RARα ternary complex mediates the retinoic acid-dependent expression of DNA methylation-sensitive genes.** *Genom Proteom Bioinf.* 2014; **12**(1): 8–18.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Maiti A, Drohat AC: **Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites.** *J Biol Chem.* 2011; **286**(41): 35334–35338.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Matsuda T, Cepko CL: **Controlled expression of transgenes introduced by in vivo electroporation.** *Proc Natl Acad Sci U S A.* 2007; **104**(3): 1027–1032.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schuermann D, Weber AR, Schär P: **Active DNA demethylation by DNA repair: Facts and uncertainties.** *DNA Repair (Amst).* 2016; **44**: 92–102.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Schwarz SD: **Inducible TDG knockout models to study epigenetic regulation [Data set].** *F1000 Research. Zenodo.* 2020.
<http://www.doi.org/10.5281/zenodo.4075154>
- Shen L, Wu H, Diep D, *et al.*: **Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics.** *Cell.* 2013; **153**(3): 692–706.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shibata E, Dar A, Dutta A: **CRL4Cdt2 E3 ubiquitin ligase and Proliferating Cell Nuclear Antigen (PCNA) cooperate to degrade thymine DNA glycosylase in S phase.** *J Biol Chem.* 2014; **289**(33): 23056–23064.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Song CX, Szulwach KE, Dai Q, *et al.*: **Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming.** *Cell.* 2013; **153**(3): 678–91.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Soriano P: **Generalized lacZ expression with the ROSA26 Cre reporter strain.** *Nat Genet.* 1999; **21**(1): 70–71.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Steinacher R, Barekati Z, Botev P, *et al.*: **SUMOylation coordinates BERosome assembly in active DNA demethylation during cell differentiation.** *EMBO J.* 2019; **38**(1): e99242.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Um S, Harbers M, Benecke A, *et al.*: **Retinoic Acid Receptors Interact Physically and Functionally with the T:G Mismatch-specific Thymine-DNA Glycosylase.** *J Biol Chem.* 1998; **273**(33): 20728–36.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Weber AR, Krawczyk C, Robertson AB, *et al.*: **Biochemical reconstitution of TET1-TDG-BER-dependent active DNA demethylation reveals a highly coordinated mechanism.** *Nat Commun.* 2016; **7**: 10806.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ying Q, Wray J, Nichols J, *et al.*: **The ground state of embryonic stem cell self-renewal.** *Nature.* 2008; **453**(7194): 519–524.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:   

Version 2

Reviewer Report 02 December 2020

<https://doi.org/10.5256/f1000research.30222.r73317>

Appendix IV

© 2020 Bultmann S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Sebastian Bultmann 

Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Munich, Germany

The authors have addressed all concerns.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epigenetics, embryonic stem cells, gene editing

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 07 October 2020

<https://doi.org/10.5256/f1000research.28295.r72019>

© 2020 Bultmann S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Sebastian Bultmann 

Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Munich, Germany

In the manuscript entitled "Inducible TDG knockout models to study epigenetic regulation" the authors present new model system for the conditional depletion of TDG. The authors employ a

strategy in which a loxP flanked mini gene is introduced into a TDG KO background. In combination with an inducible Cre introduced into the ROSA26 locus the authors convincingly show the rapid and efficient depletion of TDG. Finally, the authors validate their system by investigating the effect of acute TDG depletion on differentiation. In summary, the study presents a clever strategy for induced depletion, is well thought through and scientifically sound.

I have only two minor points the authors should address:

1. Is the TDG mini gene randomly integrated into the genome? The authors should make this clear and if possible comment on any observation regarding silencing of the transgene during long-term culture.
2. The section "molecular and cellular phenotypes" is a little confusing. The authors should make clear in the text when they are talking about induced (TDG depletion) or uninduced (control) cells. The authors should clearly explain the experimental setup shown in 3B in the text. In particular in the second paragraph of this section the authors start describing the control cells and then suddenly about TDG depleted cells.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epigenetics, embryonic stem cells, gene editing

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 12 Oct 2020

David Schürmann, University of Basel, Basel, Switzerland

We thank Dr Bultmann for the evaluation of our manuscript. We appreciate his comments helping to improve it. Accordingly, the following adaptations were made :

1. The first point was also raised by Dr Torchia (reviewer 2) and we describe better the random integration of the *miniTdg* gene. We also added a note about the transcriptional stability of the transgene in chapter 1 of the Result section.
2. We apologize for the confusing description of our results. We now clearly indicate whether we refer to TDG proficient or deficient cells. We also added an outline of the experimental setup for the neural differentiation.

Competing Interests: No competing interests were disclosed.

Reviewer Report 30 September 2020

<https://doi.org/10.5256/f1000research.28295.r71089>

© 2020 Torchia J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Joseph Torchia

Department of Biochemistry, Western University, London, ON, Canada

Thymine DNA Glycosylase (TDG) is a BER protein that is essential for the removal of 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-cC) from genomic DNA. TDG knockout is embryonic lethal at E11.5 due to genome wide epigenetic instability. To circumvent embryonic lethality from a deletion of *Tdg*, the authors synthesized a mini-*Tdg* allele consisting of an optimized TDG codon arrangement for expression of two naturally occurring TDG splice variants. The authors introduced the mini-*Tdg* allele in a *Tdg*^{-/-} background to generate *miniTdg*^{tg/tg}/*Tdg*^{-/-} mice, which phenocopy and express TDG similarly to wild-type mice. Furthermore, the TDG codons are flanked by *loxP* sites to allow for efficient spatial and temporal excision of *Tdg* in response to tamoxifen. The deletion of TDG in ESC's results in a global upregulation in 5-caC, and *Tdg* null ESC's fail to differentiate in response to retinoic acid both of which are consistent with previous literature. Overall, the paper describes a novel mouse model to study the role of TDG dependent active DNA demethylation *in vitro* and *in vivo*.

The paper is well written, and the data seems to be convincing. In addition, the experiments appear to be performed to a high standard. However, I feel addressing the following concerns would significantly strengthen the manuscript and increase the novelty of the mouse model:

Major Comments:

1. The authors need to site appropriate literature and/or provide data to address the claim that epigenetic instability resulting from a loss of TDG leads to complications in long-term ESCs cultures. Multiple studies, such as Hu *et al.* 2014¹ (PMID 24529596) and Shen *et al.* 2013² (PMID 23602152), have rigorously used *Tdg*^{-/-} mouse ESC's without any mention of such complications. I feel the argument that the use of such a complicated system to study TDG

deletion in MEFs and ESCs needs further clarification.

2. There are several published conditional knockout mouse models of TDG that incorporate *loxP* sites in the endogenous *Tdg* allele. These mouse models, and their potential drawbacks, need to be discussed to provide a stronger argument for the use of mini-*Tdg* allele approach employed in this manuscript.
3. Its unclear why it was necessary to remove the neo cassette to generate the TDGiKO1.1 and TDGiKO1.2 ESC clones? The use of CRISPR can introduce complications such as off-target mutations and clonal selection that can introduce variability in the model.

Minor comments and questions:

1. Figure 1D, 2B, 2C, 3A are all lacking statistical analysis. Please use the appropriate nomenclature to indicate statistical significance.
2. Please describe the sample size for LC-MS/MS experiment outlined in figure 1D and Figure 3A.
3. Figure 2F: Why is TDG blot not showing a doublet. The SUMOylated TDG band is missing.
4. Why was 4-OHT employed at such a high concentration? In my opinion 100 nM to 500 nM should be more than sufficient to induce excision.
5. Was the mini-*Tdg* allele integrated in the mouse genome? If so, where did the integration occur?

References

1. Hu X, Zhang L, Mao SQ, Li Z, et al.: Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell*. 2014; **14** (4): 512-22 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Shen L, Wu H, Diep D, Yamaguchi S, et al.: Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*. 2013; **153** (3): 692-706 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the rationale for developing the new method (or application) clearly explained?

Partly

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epigenetics, Biochemist

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 12 Oct 2020

David Schürmann, University of Basel, Basel, Switzerland

We thank Dr Torchia for the critical reading of our manuscript and his valuable suggestions for improvement, which helped a lot to strengthen our conclusions. We revised the manuscript according to his major and minor comments.

Major comments:

1. We understand the reviewers comment that this statement is rather speculative and currently poorly supported by data. We therefore removed it in the revised version. On the other hand, we observed in our experiments with constitutive TDG KO ESCs subtle and rather global differences (e.g. cell division time), which we do not yet fully understand. They could originate from clonal adaptation but also from TDG-driven epigenetic mechanisms. However, to be able to discriminate between these possibilities was the reason for the development of the here presented inducible TDG depletion system. Apart from the not yet ruled out role of TDG as repair enzymes for damaged DNA, the increase of fC and caC might affect DNA-protein interactions and consequently chromatin landscaping. Irrespective of these hypothetical scenarios, where long-term culturing of cells could result in both a genetic as well as an epigenetic drift, we feel that an isogenic system is always the preferable experimental system, even when it seems to be more complicated and laborious.
2. We are aware that several other conditional *Tdg* allele were generated and available as well, which are genetically easier to handle and could perfectly serve their experimental purposes. In fact, before having started to engineer our complementing *miniTdg* system, we intended but failed to introduce loxP sites at the extremities of the endogenous *Tdg* coding region to be able to delete the entire gene, omitting putative dominant truncated transcripts. We suspect that besides a poor accessibility of some regions of the *Tdg* gene, the multitude of pseudogenes consisting basically of *Tdg* cDNA sequences interferes with successful genome engineering in certain circumstances as they may act as off-targets, not only for classical gene targeting but also CRISPR/Cas9-based approaches. Having a functional *miniTdg* gene, in which the entire coding region is flanked by loxP sites, not only allows for an inducible TDG depletion but also enables further engineering steps in a flexible manner by Cre-mediated cassette exchange (although not designed specifically for this purpose). We refer to other available conditional *Tdg* allele in the revised version of our manuscript. In the last paragraph, furthermore, we point to

this advantage and possibility. This was, by the way, one of the reasons why we deleted the neomycin resistance gene (see response to comment 3 below), allowing for the use of this selectable marker for such a replacement strategy.

3. We utterly agree with the reviewer about the potential drawbacks of the CRISPR/Cas9 genome editing approach. Therefore, we added our off-target analysis for the two ESC clones generated in the "Extended data".

In fact, most of the here presented data from cultured cells and *in vivo* are obtained from the systems without the additional deletion of the neomycin resistance gene (except LC-MS/MS data in Fig. 3A). For the sake of completeness of our system, we included these two ESC clones as they might be of interest when additional selection marker-based transgene integration is intended. These notions may have been presented ambiguously in the first version and are now improved by slightly rephrasing and restructuring of the respective paragraphs.

In any case, the possibility to work with the inducible TDG depletion systems ensures an isogenic background, at least for the role of TDG, which could be validated in different in ESC clones. Although one can never completely rule out some issues of clonal selection, we still consider it an advantageous approach.

Minor comments and questions:

1. We appreciate reviewer's pointing to the missing statistical indicators, which we now added.
2. More detailed information about the analysis by LC-MS/MS was added into the figure legends and the respective paragraph in the Methods section.
3. This is a very attentive comment by the reviewer. In contrast to Figure 1, indeed, the SUMO-modified TDG is not apparent in Figure 2F because we used NP-40 soluble protein extraction methods without the addition of specific inhibitors of SUMO proteases (f.e. N-Ethylmaleimide). As de-SUMOylation happens quite fast and efficiently during mild lysis condition even at 4°C, SUMOylated TDG is in our hands only detectable when immediate lysis in Laemmli buffer and/or detergent-based lysis buffers supplement by the inhibitor. We now added the technical information about the extraction method also in the figure legend.
4. We agree with the reviewer that lower 4-OH tamoxifen concentration might work as well and added a comment about suitable Cre induction conditions in the 2nd chapter of the results.

Based on an initial literature search, we found that commonly used concentrations of OHT are in the range of 100 nM – 10 µM. To get a maximal flexibility and timely controlled TDG depletion in experimental protocols of differentiation, we aimed at rather short durations of induction. Such single induction 4-OHT treatments of as short as 2 h are rarely found in literature and associated with limited recombination efficiency (e.g. Buelow et al. 2008), while lower concentrations are more frequently used for Cre induction for up to 48 h, sometimes also with repeated administration. For our purpose, we therefore opted for 4-OHT concentrations in the intermediate to high range but avoiding cytotoxic effects.

5. This is indeed a relevant point for a potential user of our system. The *miniTdg* gene was introduced by pronuclear injection of a plasmid that contained no homology arms for a targeted integration into the mouse genome. Thus, it is a random integration event as stated in the legend of Figure 1. We added now this information also the first result section.

Unfortunately, our attempts to identify the genomic locus of integration by conventional molecular methods were so far not successful. Yet, we might get some hints from ongoing and future global NGS analyses, which would be amended to this paper as well as to the MGI entry.

Competing Interests: No competing interests were disclosed.

Reviewer Report 21 September 2020

<https://doi.org/10.5256/f1000research.28295.r71086>

© 2020 Sczepanski J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Jonathan Sczepanski

Department of Chemistry, Texas A&M University, College Station, TX, USA

In the method article "Inducible TDG knockout models to study epigenetic regulation" Schwarz *et al* establish mice and mESC cell line that allows for inducible KO of TDG. Building on their prior work, the key features of this system is a TDG mini-gene that resides between two *loxP*-Cre recombinase sequences, along with a tamoxifen-inducible Cre recombinase expression cassette. The authors demonstrate this design achieves near complete (~98%) depletion of TDG upon tamoxifen-induction within 24 h. Importantly, the TDG mini-gene recapitulates the epigenetic and repair functions of endogenous TDG prior to depletion, and cell morphology and pluripotency are not altered until after tamoxifen-induced depletion. These molecular and cellular phenotype studies are particularly well done, and clearly demonstrate agreement between inducible TDG KO ESCs and previously reported constitutive TDG KO ESCs. Overall, this work is very rigorous, and the controls are carefully designed and extensive. I anticipate that inducible TDG KO will be a valuable tool for mechanistic studies aimed at uncovering the role of TDG in DNA epigenetics and repair. I recommend this paper for indexing without further revisions.

Minor comment: ESC is written "ECS" in Figure 3 title.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: DNA repair, base excision repair, DNA demethylation, thymine DNA glycosylase

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 12 Oct 2020

David Schürmann, University of Basel, Basel, Switzerland

We are very grateful to Dr Sczepanski for taking the time to evaluate our manuscript and were pleased to learn about his overall approval and appreciation of our experimental work and data.

The spelling mistake in the title of figure 3 was corrected.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research