Investigations of Genome Instability Utilizing Quantitative Proteomics

Dissertation

zur Erlangung des Grades

„Doktor der Naturwissenschaften“

Am Fachbereich Biologie

der Johannes Gutenberg-Universität in Mainz

Emily Nischwitz

geb. am 17.01.1994 in St. Louis, MO, USA

Mainz, 2023

Table of contents

Abstract

The field of mass spectrometry-based proteomics has had a profound impact on discoveries in almost every field of science. Specifically, bottom-up proteomics enables the exploration of scientific questions at a high-throughput and proteome-wide level. In this thesis, mass spectrometry-based proteomics was employed to understand aspects of genome instability related to: 1) telomere biology in *Caenorhabditis elegans*, 2) phylogenetic diversity of recognition and repair of *in vitro* DNA damage lesions, and 3) DNA damage kinetics in *Tetrahymena thermophila.*

Article I (Dietz et al. 2021) describes the extensive characterization of the first novel double-stranded telomere binders in *C. elegans*, TEBP-1 and TEBP-2. These proteins were discovered using *in vitro* telomere pulldown assays coupled with label-free and dimethyl quantitative mass spectrometry. TEBP-1 and TEBP-2 bind directly and specifically to double-stranded telomeric DNA. Both proteins are critical to the negative and positive regulation of telomere homeostasis. The double knockout strain of *tebp-1;tebp-2* exhibits severe germline arm atrophy and synthetic sterility, suggesting their critical role in fertility. TEBP-1 and TEBP-2 dimerize and directly interact with the known single-stranded binder POT-1, thereby connecting them to the known telomere complex in *C. elegans*.

Article II (Nischwitz and Schoonenberg et al., 2023) explores the conservation of recognition and repair of DNA damage lesions. Due to the imperative need for accurate maintenance of the genome, DNA repair has been highly conserved across all domains of life. To study both the shared and unique elements of the DNA damage response, we conducted a phylointeractomic study to identify enriched binders in 11 different species at the 8-oxoG and abasic lesions, as well as a uracil base incorporated into DNA. While numerous binders were canonical DNA damage factors, we also observed enrichment of proteins not previously associated with DNA repair. Through orthology, network, and domain analysis, we linked 44 proteins that were previously unassociated to DNA repair.

Article III (unpublished, Nischwitz and Schoonenberg et al., xxxx) delves into the kinetics of the DNA damage response (DDR) in the ciliate *Tetrahymena thermophila (Tetrahymena)*. To date,

there have been limited studies that combine the power of proteomics and transcriptomics to investigate DNA damage kinetics across various treatments. Our screen monitored the dynamic DNA damage response over eight hours after exposure to six different mutagens. We observed upregulation of previously associated DNA damage repair pathways, as well as unexpected DDR crosstalk. All treatments elicited a dynamic response at both the transcript and protein level. Through unsupervised machine learning clustering, we examined expression profile trends to gain a more comprehensive understanding of the DDR, as many of these proteins exhibited damage-specific responses. Currently, we are employing a knockdown system to target a subset of these PARP-related proteins to further characterize their specific roles in *Tetrahymena*.

Zusammenfassung

Artikel I (Dietz et al. 2021) beschreibt die umfassende Charakterisierung der ersten bekannten doppelsträngigen Telomerbinder in *C. elegans*, TEBP-1 und TEBP-2. Diese Proteine wurden mit Hilfe von *in vitro* Telomer-Pulldown-Assays in Verbindung mit markierungsfreier und quantitativer Dimethyl-Massenspektrometrie entdeckt. TEBP-1 und TEBP-2 binden direkt und spezifisch an doppelsträngige telomere DNA. Beide Proteine sind entscheidend für die negative und positive Regulierung der Telomer-Homöostase. Der Doppel-Knockout-Stamm von tebp-1;tebp-2 weist eine schwere Keimbahnarmatrophie und synthetische Sterilität auf, was darauf hindeutet, dass beide Proteine eine entscheidende Rolle für die Fruchtbarkeit spielen. TEBP-1 und TEBP-2 dimerisieren und interagieren direkt mit dem bekannten Einzelstrangbinder POT-1, wodurch sie mit dem bekannten Telomerkomplex in *C. elegans* verbunden sind.

Artikel II (Nischwitz und Schoonenberg et al., 2023) befasst sich mit der Erhaltung der Erkennung und Reparatur von DNA-Schäden. Aufgrund der zwingenden Notwendigkeit, das Genom akkurat zu erhalten, ist die DNA-Reparatur in allen Bereichen des Lebens in hohem Maße konserviert. Um sowohl die gemeinsamen als auch die einzigartigen Elemente der DNA-Schadensreaktion zu untersuchen, haben wir eine phylointeraktomische Studie durchgeführt, um in 11 verschiedenen Arten angereicherte Proteine, die an den 8-oxoG- und abasischen Läsionen sowie an einer in die DNA eingebauten Uracilbase binden, zu identifizieren. Bei zahlreichen Bindungsstellen handelte es sich um kanonische DNA-Schadensfaktoren, aber wir beobachteten auch eine Anreicherung von Proteinen, die bisher nicht mit der DNA-Reparatur in Verbindung gebracht wurden. Durch Orthologie-, Netzwerk- und Domänenanalysen konnten wir 44 Proteine identifizieren, die zuvor nicht mit der DNA-Reparatur assoziiert wurden.

Artikel III (unveröffentlicht, Nischwitz und Schoonenberg et al., xxxx) befasst sich mit der Kinetik der DNA-Schadensreaktion (DDR) in dem Ciliaten *Tetrahymena thermophila* (*Tetrahymena*). Bislang gibt es nur wenige Studien, die die Leistungsfähigkeit von Proteomik und Transkriptomik kombinieren, um die Kinetik von DNA-Schäden bei verschiedenen

Behandlungen zu untersuchen. In unserem Screening wurde die dynamische DNA-Schadensreaktion über acht Stunden nach der Exposition gegenüber sechs verschiedenen Mutagenen überwacht. Wir beobachteten die Hochregulierung von zuvor assoziierten DNA-Schadensreparaturwegen sowie unerwartete DDR-Crosstalk. Alle Behandlungen lösten eine dynamische Reaktion sowohl auf der Transkriptions- als auch auf der Proteinebene aus. Mit Hilfe von unüberwachten maschinellen Lernens untersuchten wir die Trends der Expressionsprofile, um ein umfassenderes Verständnis der DDR zu gewinnen, da viele dieser Proteine schadensspezifische Reaktionen zeigten. Gegenwärtig setzen wir ein Knockdown-System ein, um eine Untergruppe dieser PARP-verwandten Proteine anzugreifen und ihre spezifische Rolle in *Tetrahymena* weiter zu charakterisieren.

List of publications

Published articles

Article I: Dietz, S*., Almeida, M.V.*, Nischwitz, E. et al. The double-stranded DNA-binding proteins TEBP-1 and TEBP-2 form a telomeric complex with POT-1. *Nat Commun* 12, 2668 (2021). https://doi.org/10.1038/s41467-021-22861-2
These authors contributed equally.

Article II: Nischwitz, E.*, Schoonenberg, V.A.C.*, Fradera-Sola, A. et al. DNA damage repair proteins across the tree of life. *iScience.* (2023). https://doi.org/10.1016/j.isci.2023.106778
These authors contributed equally.

In preparation

Article III: Nischwitz, E.*, Schoonenberg, V.A.C.*, Mullner, R. et al. A systems view on DNA damage kinetics in *Tetrahymena thermophila*.
These authors contributed equally.

Additional Publications

Fradera-Sola, A., Nischwitz, E., Bayer, M.E. et al. RNA-dependent interactome allows network-based assignment of RNA-binding protein function. *Nucleic Acids Res.* 18:gkad245 (2023). https://doi.org/10.1093/nar/gkad245

Berkyurek, A.C., Furlan, G., Lampersberger, L., Beltran, T., Weick, E-M., Nischwitz, E. et al. The RNA polymerase II subunit RPB-9 recruits the integrator complex to terminate *Caenorhabditis elegans* piRNA transcription. *EMBO J.* 1;40(5), e105565. (2021). https://doi.org/10.15252/embj.2020105565

Introduction

Genome Instability

Replication Stress

Replication stress refers to the challenge or interference of DNA replication and is the most common cause for genomic instability (Aguilera and García-Muse 2013). Defective replication can result in stalled replication forks, reduced reinitiation capacity, reversed replication forks, or most drastically complete collapse. Replication defects can arise at the DNA, transcript, and protein level. DNA lesions frequently cause stalls in the progression of the replication machinery (Saxena and Zou 2022). These defects can often be resolved through DNA damage tolerance pathways. One such pathway is translesion synthesis, which involves replacing the replicative polymerase with a specialized translesion synthesis polymerase (pol ζ or η). These polymerases allow for the bypass of the lesion, enabling replication to continue while providing increased time for repair. However, this process can be highly error-prone (Bi 2015).

When ongoing replication and transcription intersect, collisions can occur between the replisome and transcriptome. Co-directional collisions do not affect replication, but a head-on collision can lead to replication fork pausing or complete collapse if multiple transcriptomes collide with the active replisome. This can have severe consequences , as it often results in the complete collapse of both complexes (Pomerantz and O'Donnell 2010). Additionally, physical protein barriers bound to the DNA can pose a significant threat to continuous replication. To address potential collision, the Fork Pausing/Protection Complex (FPC) slows the replisome (Shyian and Shore 2021). If any form of replication stress is not relieved, sustained DNA damage can occur, which, in extreme cases, can lead to chromosomal fusion, fragmentation, and loss (Aguilera and García-Muse 2013).

Although replication stress is not solely characterized by physical changes in DNA, it often results in the formation of single-stranded DNA (ssDNA). RPA binds to this ssDNA, generating a signal that activates a primer-template junction (Zeman and Cimprich 2014). This junction triggers the recruitment of two critical phosphatidyl- inositol 3-kinase (PI 3-kinase): Ataxia

telangiectasia (ATM) and ATM and Rad3-related (ATR). ATR plays a particularly crucial role. Activation of ATR halts cell cycle progression, regulates origin firing, which slows DNA replication, and allows time for appropriate DNA repair. Specifically at stressed replication forks, ATR stabilizes the replisome, inhibits nucleases, and ultimately controls the restart of replication (Saldivar, Cortez, and Cimprich 2017). In extreme cases when a replication fork cannot be stabilized, it collapses, leading to a double-stranded DNA (dsDNA) break. This activates ATM. Most often, these dsDNA breaks occur in S and G2 phase, but when they occur during M-phase, they can result in anaphase bridges. Anaphase bridges are a hallmark of genomic instability, and they often lead to improper segregation of chromosomes (Aguilera and García-Muse 2013).

Defective DNA repair

DNA Damage Response

The stability of the genome is constantly threatened by both exogenous and endogenous mutagens. These genotoxic stressors can damage the architecture of the DNA, causing single-stranded breaks, double-stranded breaks, or chemical modifications to individual bases. To prevent genomic instability and replication stress, there is a carefully orchestrated DNA damage response (DDR) that functions to identify and subsequently repair damaged DNA (Ciccia and Elledge 2010). This DDR consists of a cascade of: sensors, transducers, mediators, and effectors (Molinaro, Martoriati, and Cailliau 2021).

Sensors are the proteins or complexes of proteins that initially detect the damage. This consists of a large variety of proteins that have specificity towards the recognition of particular damage lesions. Two well characterized sensors are the MRN complex (containing MRE11-RAD50-NBS1), the 9-1-1 complex (containing RAD9-HUS1-RAD-1) (Lee and Park 2020; QIU and HUANG 2021). The MRN complex is especially critical to the recognition of double-stranded breaks, and its activity is a crucial deciding factor between homologous recombination and non-homologous end joining (QIU and HUANG 2021; Ackerson et al. 2021). The 9-1-1 complex, like PCNA (Proliferating Cell Nuclear Antigen), is a sliding clamp which is loaded onto sites of DNA damage. This is critical to the initiation of the DDR. Sensors

specialized towards particular DNA damage lesions are detailed in the following sections.

After recognition by a sensor, transducers are recruited, which are typically kinases that amplify the DNA damage signal. These kinases often directly phosphorylate effectors and aid in the recruitment of additional mediators and effectors. The most critical transducers are in the family of phosphatidylinositol 3` kinase-related kinases, which includes ATM (ataxia-telangiectasia mutated), ATR (ATM- and Rad3-related), and DNA-PKcs (DNA-dependent protein kinase). In addition to the role of ATM and ATR in mediating replication stress, this family of kinases also phosphorylates a wide variety of mediators and/or effectors (Maréchal and Zou 2013). DNA-PKcs is particularly critical to initiating the repair of double-stranded DNA breaks (Yue et al. 2020). These kinases ultimately facilitate arrest of cell cycle, address faulty replication forks, inhibit the origin of replication firing, lower dNTP levels, initiate apoptosis, and induce a transcriptional response (Lanz, Dibitetto, and Smolka 2019). Additionally, all PIKK kinases play a role in the phosphorylation of the variant histone H2AX. When the Ser-139 residue becomes phosphorylated it forms γH2AX. (Podhorecka, Skladanowski, and Bozko 2010). This modified histone variant is critical to the overall orchestration of the DDR, influencing the recognition of DNA damage, recruitment of DNA damage factors, and triggering cell cycle arrest (Mah, El-Osta, and Karagiannis 2010).

Mediators are critical to the facilitation and mediation of DNA damage factors, as well as being largely responsible for the critical arrest of the cell cycle. The PIKK kinase ATR and ATM phosphorylate many of these mediators, such as CHK1, CHK2, MDC1, 53BP1, and BRCA1 (Coster and Goldberg 2010; Giglia-Mari, Zotter, and Vermeulen 2011). These proteins go on to interact with cyclins, cyclin-dependent serine/threonine kinase (CDKs), and p53. Upon damage, DNA cyclins and their associated CDKs ensure a halt in cell cycle progression in order to not pass damaged DNA to daughter cells (Hustedt and Durocher 2017). p53, many times referred to as the guardian of the genome, activation is absolutely critical to DNA repair. p53 acts as a transcriptional regulator to proteins directly involved in cell cycle checkpoint, senescence, DNA repair, and apoptosis (Williams and Schumacher 2016). Finally, effectors are the proteins which perform the final repair steps which return the initial integrity to the previously damaged sequence. Specific effectors are described in detail in the following sections.

The recruitment of the appropriate sensor, transducer, mediator, and effector is conditional on the cell cycle, extent of damage, and type of DNA damage. Together, these elements dictate the needed DNA repair pathway, which can be largely categorized into either excision or double-stranded DNA break repair (DSBR) (Molinaro, Martoriati, and Cailliau 2021).

Excision Repair

The excision repair pathways facilitate the resolution of either a damaged base pair or a mismatched base pair. Each has diverse recognition proteins which are specialized to their associated lesion, nucleases which excise the DNA lesion, and polymerases and helicases that fill in the missing base pairs. Despite fundamentally similar mechanisms, excision pathways vary in recognized substrates and precise members of the pathway. The primary types of excision repair are: nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR), as well as the more recently identified ribonucleotide excision repair (RER) (Figure 1).
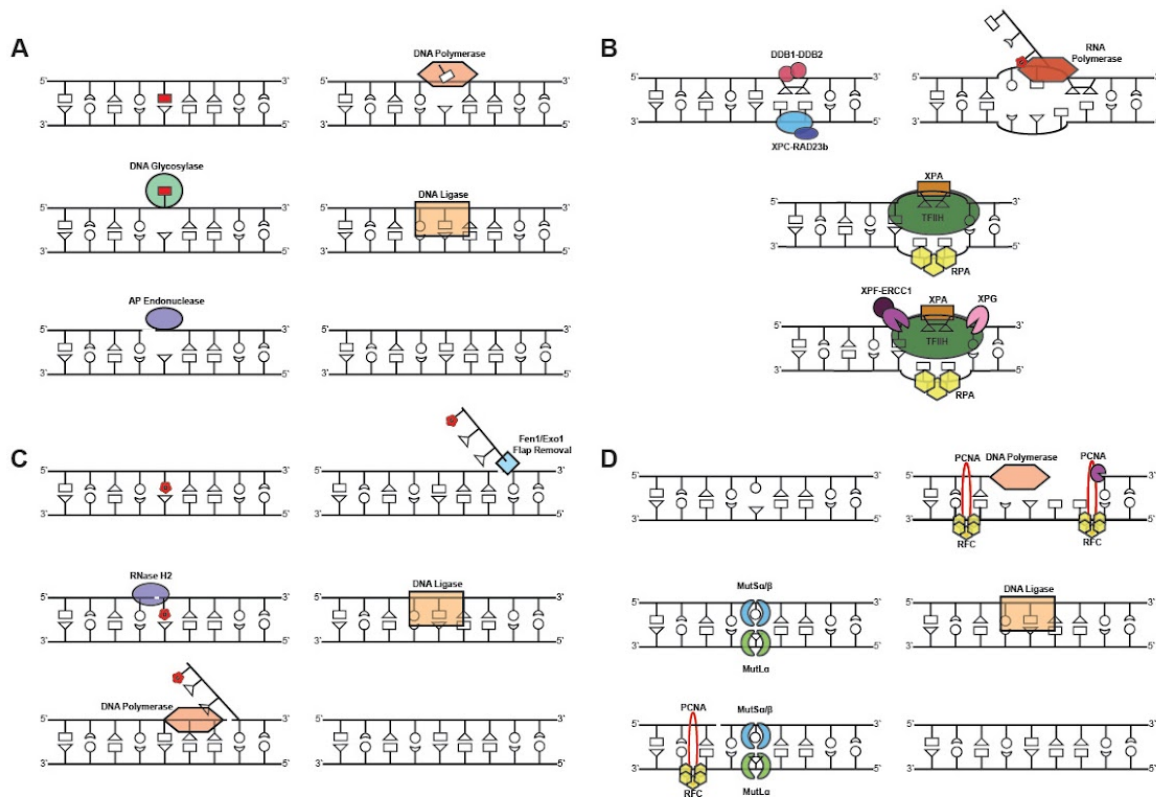

Figure 1. Simplified schematics of excision repair pathways. A) Base excision repair (BER). Damage is recognized by a specialized DNA glycosylase. The damaged base is removed by an AP endonuclease. The missing

bases are filled in by DNA polymerase and sealed by DNA ligase. B) Nucleotide excision repair (NER). Global genome NER recognition depicted (left) and transcription coupled NER recognition (right) are initiated either throughout the entire genome or in tandem to transcribed DNA, respectively. After recognition, the pathways converge, and TFIIH, XPA, and RPA are recruited to the site of damage. XPF-ERCC1 and XPG incise to 5` and 3` sides of the damage, respectively. Missing bases are filled in, and the strands are sealed. C) Ribonucleotide excision repair (RER). When a deoxyribonucleic base is mismatched with a ribonucleic base, RNaseH1 incises the 5` side of the ribonucleoside, allowing for the recruitment of a replicative DNA polymerase. This creates a displaced DNA flap, which is removed by FEN1/EXO1, and the repaired strand is sealed. D) Mismatch repair (MMR). If there is incorrect complementary base pairing, either MutSα or MutSβ are recruited to the site of damage, and MutL is recruited to the opposite strand. PCNA and RFC are recruited and MutL and Exo1 excise the mismatched base. The base is then filled in by the replicative polymerase and sealed.

Nucleotide excision repair (NER) resolves cyclobutane pyrimidine dimers (CPDs) and pyrimidine(6-4) pyrimidone photoproducts (6-4PP), which are often caused by UV light (Marteijn et al. 2014). All lesions repaired by NER cause destabilization of the DNA duplex, which triggers recognition by the NER machinery (Figure 1A). There are two types of nucleotide excision repair processes: global genome NER (ggNER) and transcription-coupled repair (tcNER) (Schärer 2013). TcNER occurs only at points of active transcription, whereas global genome NER proteins scan the entire genome. These pathways primarily differ in their associated DNA damage recognition proteins. TcNER is triggered as a result of stalled transcription, which signals for CSB (Cockayne Syndrome group B), UVSSA (UV-Stimulated Scaffold protein A), and USP7 (Ubiquitin-Specific-Processing protease 7). GgNER is initiated with XPC-RAD23B (Xeroderma Pigmentosum C, RADiation sensitive 23B) genome wide (Tatum et al. 2011). After recognition proteins are recruited and the initial protein complexes are stabilized, in both tcNER and ggNER the transcription and NER factor TFIIH is recruited. Its catalytic subunit XPB unwinds the DNA around the lesion, and XPD translocates along the single-stranded DNA opening it, and stalls at the location of the damage (Sugasawa et al. 2009). This stalling causes XPA, RPA (Replication Protein A), and XPG to come to the site of damage. XPA recruitment shifts lesion-bound protein to allow RPA to coat the ssDNA. Then ERCC1-XPF (Excision Repair Cross Complementation group 1) and XPG to allow for a dual incision to remove the damaged sequence. ERCC1-XPF makes the first incision at the 5` end of the sequence, which triggers the second incision at the 3` end (Fagbemi, Orelli, and Schärer 2011). DNA polymerase κ, ε, or δ fill in the gap created and is sealed by DNA ligase I or DNA ligase IIIα.

Mismatch repair (MMR) is triggered by the incorporation of an inappropriate base pair during

DNA replication (Jiricny 2013). The mismatched base pair is recognized by either the MutSα or MutSβ complexes. MutSα, containing MSH2-MSH6 (MutS Homolog 2-6), detects a singular mismatched base and small insertions, and the MutSβ complex, containing MSH2-MSH3 (MutS Homolog 2-3), detects larger insertions or deletions. When bound to the mismatched base pair, this causes an ATP-dependent conformational change, which allows the binding of the MutL complex (Pećina-Šlaus et al. 2020). In *H. sapiens,* the MutLα, consisting of MLH1 (MutL homolog 1) and PMS2 (postmeiotic segregation increased 2) prompts the recruitment of PCNA and RFC (Replication Factor C) proteins. This causes a conformational change in PSM2 activating the nuclease domain (Kunkel and Erie 2015). Once PSM2 has made an initial incision in the strand with the misincorporated base, EXO1 removes the mismatched base pair. Gap filling is performed by DNA polymerase δ and ligated by DNA ligase I (G.-M. Li 2008).

Base excision repair (BER) is primarily responsible for removing non helix-distorting lesions (Beard et al. 2019). Some of the most prevalent lesions removed via the BER pathway are alkylated or oxidized bases and misincorporated uracil. The repair process generally begins with damage recognition by a DNA glycosylase, which flips out and removes the damaged base, creating an apurinic/apyrimidinic-site (AP site/abasic site). Abasic sites can be formed not only as BER intermediates but also endogenously. When abasic sites are generated, a 5`-cleavage event is triggered by an AP endonuclease, resulting in a 3`-hydroxyl and 5`-deoxyribose phosphate (Krokan and Bjørås 2013). In single nucleotide repair, called short-patch BER, the 5`-deoxyribose is removed primarily by DNA polymerase β or γ, and the resulting gap is then filled. If two or more nucleotides are repaired, called long-patch BER, the 3`-hydroxyl is used for strand displacement synthesis via either DNA polymerase β or δ and ε, typically in conjunction with PCNA (Gredilla, Garm, and Stevnsner 2012). The previously cleaved 5`-deoxyribose strand, often referred to as a 5`-flap, is removed by FEN1. In both instances, the nick is sealed with ligase I or III (Robertson et al. 2009).

Ribonucleotide excision repair (RER) removes rNTPs from DNA. As DNA polymerases replicate DNA, there is a much larger pool of ribonucleotides than nucleotides. For example, in *S. cerevisiae* there are 30 to 200-fold more ribonucleotides than nucleotides (Nick McElhinny et al. 2010). Thus, when a rNTP is incorporated and the proofreading capacity of the replicating

polymerases fails, RER removes it to protect the integrity of the genome. RNaseH2 recognizes the rNMP, and incises at the 5`-side of the ribonucleoside, leaving a 3`-hydroxyl and 5`-phosphate (Sassa, Yasui, and Honma 2019). Similar to BER, the 3`-hydroxyl is used for strand displacement DNA synthesis, either via DNA polymerase δ supported by PCNA or by DNA polymerase ε. The flap that is formed, starting with the 5`-phosphate, is removed by FEN1 or EXO1, after which the repaired strand is ligated (Kellner and Luke 2020).

Double-strand break repair

Double-strand break repair (DSBR) can be broken into two major categories: resection-dependent and resection-independent (Figure 2). Resection-dependent pathways
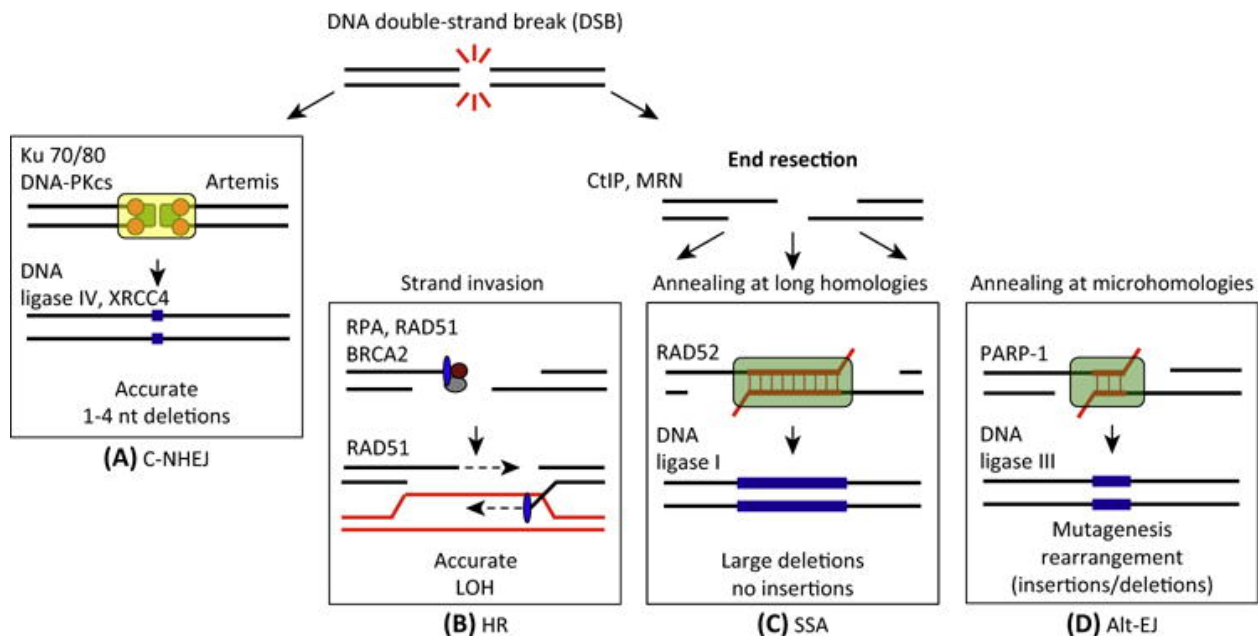

Figure 2. Double-strand break repair in *H. sapiens*. A) Classical-non-homologous end joining (c-NHEJ). Resection-independent pathway that relies on KU70 and KU80 recognize the double-stranded break and recruit DNA-PKcs, which phosphorylate Artemis. The strands are annealed and sealed by XRCC4 and DNA Ligase IV. B) Homologous recombination (HR). Resected ends are bound by BRCA2, RPA, and RAD51 to the break. RAD51 performs the homology search to fill in this missing DNA. C) Single-stranded Annealing (SSA). The long resected single-stranded DNA anneals to homologous sequence. Both RAD52 and EXO1 remove the remaining flats after annealing, and the sequence is filled and ligated. D) Alternative-end joining (alt-EJ). Resected segments containing short stretches of homology anneal with the facilitation of PARP1, and the missing sequence is filled in and ligated. Image originated from Ceccaldi, Rondinelli, and D'Andrea 2016.

include homologous recombination (HR), alternative end joining (alt-EJ), and single strand annealing (SSA). Non-homologous end joining (NHEJ) is the resection-independent pathway. Resection-dependent pathways are most prevalent in the mid-S and mid-G2 phase of the cell cycle (Ceccaldi, Rondinelli, and D'Andrea 2016). Both HR and alt-EJ require templates for recombination, while SSA anneals and resects the complementary strand. The EJ repair system is most prevalent in G0, G1, and G2 of the cell cycle (Hustedt and Durocher 2017). EJ simply anneals resected DSBs without consultation of template DNA. There is a consistent competition between each of these repair pathways (Ackerson et al. 2021).

Homologous recombination is the preferred pathway due to its lower error rate compared to alternative mechanisms (Figure 2B). However, it requires the close proximity of the sister chromatid. The initiation of HR heavily relies on CtIP, which simulates exonuclease activity of the Mre11-Rad50-Nbs1 (MRN) complex (Wright, Shah, and Heyer 2018). This complex plays a role in both cell cycle progression and the resolution of DSBs. The MRN complex possesses both exonuclease and endonuclease activities. It initiates by creating an endonuclease nick and then proceeds with exonuclease activity in the 3` to 5` direction. Further resection is facilitated by EXO1 and BLM-DNA2 in the 5` to 3` direction, generating large overhangs. During this resection, the ssDNA is coated with RPA, which is replaced by the recombinase RAD51 and BRCA2 (Jasin and Rothstein 2013). RAD51 and BRCA2 facilitate homology search preferring sister chromatids or if needed the homologous chromosome. DNA synthesis and ligation occurs, leading to either non-crossover or crossover events. One significant drawback of this pathway is the potential loss of heterozygosity, especially if an excess of crossover events take place (Wright, Shah, and Heyer 2018).

While the loss of heterozygosity is not a concern in SSA or Alt-EJ, there is a potential for undesired mutagenesis. The SSA pathway is preferred when a DSB occurs in a long stretch of homologous sequences (Bhargava, Onyango, and Stark 2016). When the unprotected ends become extensively resected, the resulting 3` overhang allows for annealing of the homologous sequences (Figure 2C). The non-homologous 3` ssDNA flap is removed by RAD52 and EXO1, and the missing sequence is filled in and annealed. This can often lead to large deletions if the

single-strand annealing occurs significantly upstream of its originally complementary sequence (Ceccaldi, Rondinelli, and D'Andrea 2016). Unlike SSA, Alt-EJ (also referred to as microhomology-mediated end joining) requires only a small amount of resected sequence and a short area of complementary sequence (Figure 2D). It has been suggested that both the MRN complex and PARP1 play a critical role in the early processing steps of alt-EJ (Sfeir and Symington 2015). After this short stretch of typically 2-20 base pairs anneals, the resulting 3` flaps are removed and the sequence is ligated (Ackerson et al. 2021). Polymerase θ fills in the sequence and it is annealed by LigIII (Sfeir and Symington 2015).

Classic non-homologous end joining (c-NHEJ) does not rely on the resection activity of CtIP and MRN as in homologous recombination (Figure 2A). Instead, the KU70-KU80 heterodimer is recruited to the site of the DSBs. This complex then recruits DNA-PKcs, a critical kinase in the DDR. DNA-PKcs phosphorylate Artemis which further trims the overhangs of the DSB. Ligase IV and XRCC-4 are recruited and the strands are annealed (Chang et al. 2017). The use of this pathway can result in both insertions and deletions (Ackerson et al. 2021).

DNA damage repair in exceptional organisms

The maintenance of genomic stability is so crucial that there is consistent conservation of DNA repair pathways and proteins across all three domains of life (Kovalchuk 2016). One canonical example of this is the base excision repair pathway. It is assumed that early in evolutionary development, prokarya were subjected to extensive oxidative and alkylative damage (Prorok et al. 2021). The need to repair this type of damage has been highly conserved. However, bacteria possess only six DNA glycosylases, whereas mammals exhibit increased specialization with 11 different types of glycosylases (Krokan and Bjørås 2013; Kurthkoti et al. 2020). These glycosylases have become more specific and sensitive to recognizing particular types of damage in the genome. Similarly, the nucleotide excision repair pathway, which removes small helix-distorting adducts, is conserved across the domains of life. However, in *E. coli* the process involves 6 proteins, whereas in *H. sapiens,* it involves over 20 proteins. Although the overall outcome is equivalent, the actual steps have become more specialized over time (Schärer 2013; Kisker, Kuper, and Van Houten 2013).

Despite the general conservation of DNA factors, there are some specific cases that exhibit notable divergence. For instance, photolyases function as a direct reversal protein, using light to cleave thymine dimers. Photolyases are critical repair proteins in bacteria, archaea, plants, fungi, and animals. However, they lost all DNA repair functionality in placental mammals (Mei and Dvornyk 2015). Reinforcing that even critical and efficient repair proteins and processes can ultimately be replaced. Another repair pathway which has had more significant divergent evolution is intrastrand crosslink repair. This repair pathway addresses crosslinks occurring within the same and between different strands of DNA. *E. coli* remove these lesions through DSBR, while animals employ a novel protein family called Fanconi Anemia (FA). In addition to the FA proteins, various members of the BER, NER, and DSBR pathways are recruited for intrastrand crosslink repair, creating an interesting case of both novel DNA repair mechanisms and adaptation of existing pathways (Moldovan and D'Andrea 2009).

These convergences and divergences highlight the importance of understanding DNA repair using various model organisms. Ciliates, in particular, serve as interesting model organisms for studying DNA repair. One of their remarkable characteristics is their unique nuclear architecture, with a transcriptionally active macronucleus and a transcriptionally silent micronucleus (Orias, Cervantes, and Hamilton 2011). It is well-established that the state of chromatin has a significant impact on the repair process (Stadler and Richly 2017). Therefore, the ability to observe distinct chromatin states within the organism's typical biology provides a powerful tool for DNA repair studies. Additionally, during *Tetrahymena* sexual reproduction (conjugation) undergoes two major rearrangements: the removal of internal eliminated sequences (IESs) and the fragmentation of the micronuclear chromosomes. Approximately 12,000 IES sites are eliminated, resulting in the loss of one-third of the genome (Hamilton et al. 2005). Once these sites are identified through a scnRNA system with an adapted PiggyBac transposase system, the breakage points are rejoined by non-homologous end joining. Alongside the breakage sites created during IES elimination, the micronuclear chromosomes are fractionated at 200 different chromosome breakage sites (CBSs). Altogether, this is a massive undertaking in genome rearrangement which is highly influenced by known DNA repair proteins (Loidl 2021; Loidl and Lorenz 2016). Another unique characteristic of ciliates is their extremely low base-substitution

rate. *Tetrahymena* and *Paramecium aurelia* exhibit the two lowest naturally occurring base-substitution rates in all of eukarya and prokarya (Sung et al. 2012; Long et al. 2016). This exceptionally low mutation rate could potentially be attributed to highly efficient DNA repair. As some of these repair mechanisms might be unique to ciliates, studying the DNA damage response in an unbiased manner can be a large advantage.

Site-specific hotspots of genome instability

There are specific regions of the genome that are particularly prone to replication errors and are more susceptible to chromosome breakage and gross chromosomal rearrangements. Tandem repeats found within LINES, SINES, LTR, and generally within transposons are considered site-specific hotspots (Nesta, Tafur, and Beck 2021). Replication slippage can easily lead to inappropriate extension or shortening of these tandem repeats. Fragile sites are another hotspot for genome instability and are characterized by breakage when exposed to mild replicative stress (Aguilera and García-Muse 2013). While they lack a specific sequence, they often contain stretches of A/T repeats with occasional C/G interruptions. One reason why a subset of these repeats can lead to genome instability is the formation of non-B DNA structures. Examples of non-B DNA structure include Z-DNA, triple-DNA, hairpin, cruciforms, G-quadruplexes, and telomeres. These sequences disrupt regular replication and often require the formation of specific nucleoprotein complexes to facilitate successful replication (Nesta, Tafur, and Beck 2021)

Unlike the other repeats, telomeres have a defined position at the terminus of linear chromosomes. Telomeres protect essential genetic information, while also regulating cellular division. Telomeres solve both the end replication and end protection problems (Shay and Wright 2019). With each round of replication, a small portion of the genome is lost because DNA polymerase cannot fully replicate to the ends of chromosomes. Telomeres typically consist of a G-rich sequence with tandem repeats that create a 3` single stranded DNA overhang. This single strand invades the adjacent double-stranded telomeric DNA, forming a cap (Armanios 2022). Proteins associated with both the single- and double-stranded portions of telomeres act as protectors from inappropriate DNA damage repair. They play a role in repressing ATM, ATR, PARP, and DSBRs (de Lange 2018). Simultaneously, this protein structure must be permissive to telomerase, a reverse transcriptase adding additional telomeric sequence to chromosome ends

(Palm and de Lange 2008; Wu et al. 2017, 201). This protective complex of proteins associated with telomeres is commonly referred to as shelterin (Figure 3A).
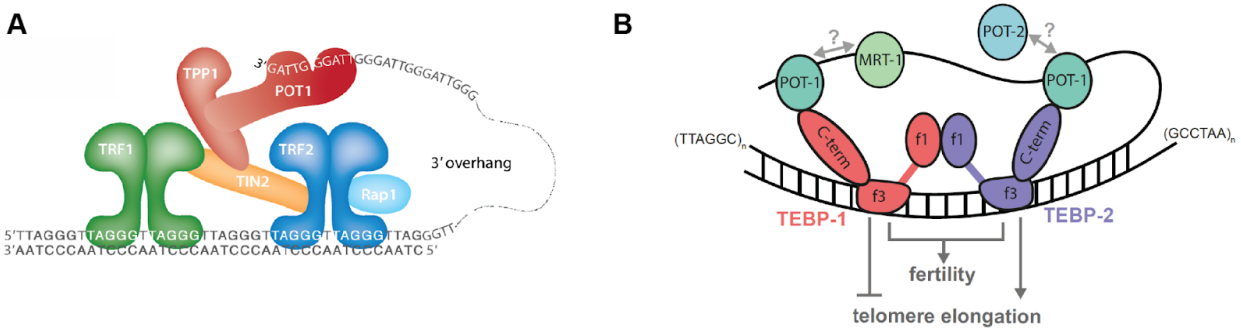

Figure 3. Telomere architecture in *H. sapiens* and *C. elegans*. A) Schematic of *H. sapiens* telomeres. Here, all six proteins of the shelterin complex are depicted: the double-stranded telomere binders TRF1 (green), TRF2 (dark blue) with its binding partner RAP1 (light blue), the bridging proteins TIN2 (yellow) and TPP1 (orange), and the direct 3` single-stranded overhang binding protein POT1 (red). B) Schematic of *C. elegans* telomeres. Here, the five known members of the telomere end binding complex are depicted: the two double-stranded telomere binders, TEBP-1 and TEBP-2 and their direct interactor single-stranded telomere binder, POT-1, as well as the other single-strand binders, MRT-1 and POT-2. Adapted from (de Lange 2018; Dietz et al. 2021).

Although the function of shelterin is highly conserved among organisms, there is variability in their precise composition (Palm and de Lange 2008; Linger and Price 2009). Vertebrates have six shelterin proteins: TRF1, TRF2, POT1, TPP1, TIN2, and RAP1. TRF1 and TRF2 bind double-stranded DNA as homodimers, and POT1 binds the 3` single-stranded overhang (Broccoli et al. 1997; Baumann and Cech 2001). TPP1, TIN2, and RAP1 are critical scaffold proteins (Xin et al. 2007; Kim et al. 2004; B. Li, Oestreich, and de Lange 2000). In vertebrates, *S. cerevisiae,* and *S. pombe* possess proteins that specifically bind either the single-stranded and double-stranded telomeric DNA, performing unique and essential functions. Until recently, only single-stranded binders have been described in *Caenorhabditis elegans*.

C. elegans have an unusual telomere architecture (Figure 3B). At the end of 2 to 9 kb of tandem TTAGGC repeats, their telomeres can terminate in both a 3` G-rich or a 5` C-rich single-stranded overhang, each with unique telomere binders (Raices et al. 2008). The three single-stranded telomere binders in *C. elegans* are Mrt-1, Pot-1 (formerly CeOB2), and Pot-2 (formerly CeOB1). Mrt-1 binds to both the 5` and 3` single-stranded overhang, and its OB fold is necessary for the recruitment of telomerase (Meier et al. 2009). Both Pot-1 and Pot-2 are negative regulators of telomerase via the binding of the 3` and 5` overhang respectively (Shtessel et al. 2013). Pot-2

additionally plays a critical role in the regulation of alternative lengthening of the telomeres (Cheng et al. 2012). While telomerase activity is not required for propagation, all three proteins are involved in the regulation of telomerase (Lackner et al. 2012).

Mass spectrometry-based proteomics

Overview of mass spectrometry-based proteomics

The powerful tool of mass spectrometry utilizes electrical current to identify a variety of analytes, such as metabolites, peptides, lipids, and proteins. In its simplest form, the process of analyzing samples on a mass spectrometer can be broken into three parts: ion generation via an ion source, ion separation via a mass analyzer, and ion detection via a detector. The ion source ensures the analyte enters a gaseous and charged state. The mass analyzer separates ions based on their m/z ratio. The detector measures the changes in image current to ultimately determine their mass/charge value.

Sample preparation

The overall goal of sample preparation is to maintain sample integrity while also creating analytes that can be ionized and measured by the mass spectrometer (Figure 4A). There are various ways to successfully extract proteins from samples. The majority of extraction protocols involve some level of physical disruption of the sample. This disruption can be achieved using a homogenizer, bead-mill, grinder, sonicator, or any other tools that break apart the sample. In addition to physical disruption, many protocols also rely on chemicals or enzymatic digestion (Rogers and Bomgarden 2016).

After successful protein extraction, there are four widely used protocols for processing samples for bottom-up proteomics: in-gel digestion, in-solution digestion, filter-aided sample preparation (FASP), and single-pot, solid-phase-enhanced sample preparation (SP3). In-gel digestion, samples are first separated on an SDS-PAGE gel. Proteins within the sample are then reduced, alkylated, and dehydrated while still in the gel piece. After rehydration, samples are digested with a protease, and peptides are eluted and desalted before measurement (Shevchenko et al. 2006). While in-solution digestion follows similar principles, protein samples are re-solubilized

19

with high concentrations of urea. The same reduction, alkylation, and digestion occur, but the samples do not require dehydration before digestion. Both these methods are often followed by an affinity chromatography column, which removes salts from the sample before measurement (Rappsilber, Mann, and Ishihama 2007). Filter-aided sample preparation (FASP) utilizes centrifugal concentrators to process samples. Similar to in-solution digestion, proteins are initially suspended in 8 M urea, reduced, and alkylated. However, this is done directly on the centrifugal concentrator. After removing excess alkylating agent, a protease is added for digestion. Then, the centrifugal concentrator is ultracentrifuged to release only the digested peptides (Wiśniewski 2018). Single-pot solid-phase-enhanced sample preparation (SP3) utilizes beads with a hydrophobic coating to purify and serve as an interface for processing proteins into peptides (Hughes et al. 2019).
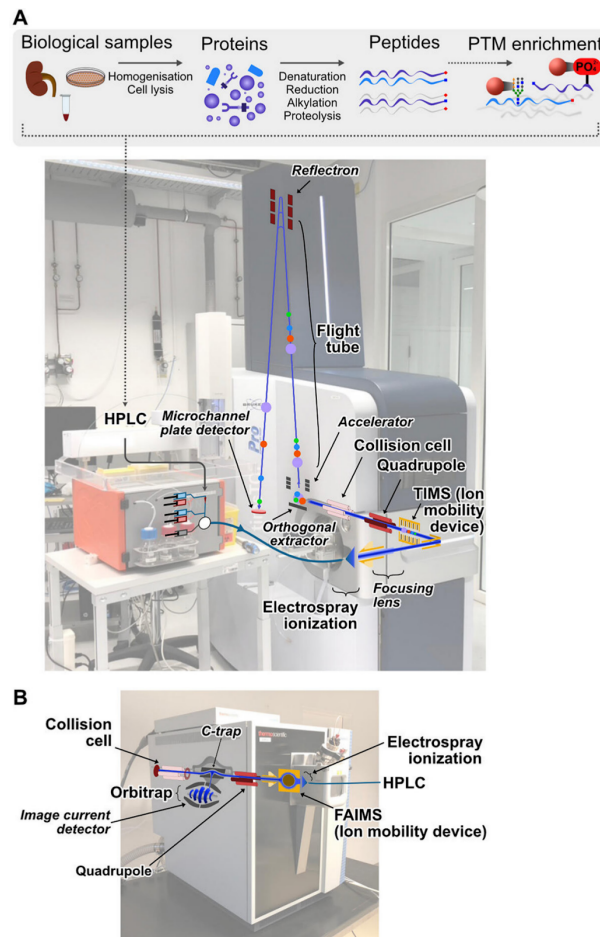

Figure 4. Overview of mass spectrometry sample preparation and equipment. A) Overview of sample preparation for mass-spectrometry based proteomics. Labeled schematic of Matrix-Assisted Laser

Desorption/Ionization Time of Flight (MALDI-TOF). B) Labeled schematic of an Orbitrap. Adapted from (Sinha and Mann 2020).

Liquid Chromatography

Before reaching the mass spectrometer, the prepared samples are separated using liquid chromatography. Chromatography is the process of separating components of a mixture based on their chemical properties. It consists of a mobile and stationary phase. The mobile phase is a mixture of a liquid or gas solvent and the sample of interest, while the stationary phase is the immobile component that interacts with the mobile phase as it passes. The extent of this interaction causes the separation of the sample (Thomas et al. 2022). Mass spectrometers are often paired with high-performance liquid chromatography (HPLC). Unlike traditional LCs that rely on gravity for separation, HPLC applies high pressure to decrease separation time. HPLC can be conducted in normal and reverse phases. Normal phase HPLC has a nonpolar liquid mobile phase that interacts with a polar stationary phase. While uncommon, this system would be beneficial if the sample of interest cannot suspend in polar liquids. The more commonly used reverse phase HPLC has a polar liquid phase and a stationary phase of silica or polymers, such as C18 material (Bhole et al. 2020).

Ion Source

The successful conversion of analytes into a gaseous and ionized state is critical for accurate mass analysis. There is a wide variety of ionization methods, including electron ionization, single photon ionization, chemical ionization-like, plasma-based, electrospray ionization, and laser desorption-based (Bhardwaj and Hanley 2014). Two primary ionization methods used specifically within liquid-chromatography mass spectrometry are electrospray ionization (ESI) and laser desorption ionization (LDI). ESI utilizes slightly acidic conditions within the final preparation solution to protonate all proteins. Droplets of the solution are expelled from the capillary under high voltage (3-5 kV), propelling them towards the grounded MS. As these droplets travel to the MS, the solvent disintegrates and the ions repel each other until each droplet contains a singular ion (Nadler et al. 2017). In laser desorption ionization, a laser exposes the sample to high amounts of energy, causing the sample to become ionized by side products

and undergo desorption, which is the release from an attached surface. Matrix-assisted laser desorption/ionization (MALDI) is a subset of LDI commonly used in bottom-up proteomics (Figure 4B). In MALDI, samples are embedded into a matrix that disperses the intensity of the energy from the laser, ensuring vaporization with the least amount of degradation (Nadler et al. 2017).

Types of mass analyzers and mass detectors

There are four general types of mass analyzers: quadrupole, ion-trap, time-of-flight (TOF), and Fourier-transform ion cyclotron resonance (FTICR).

It is common to pair these mass analyzers together to improve measurement. A common pairing is that between a quadrupole and ion-trap mass analyzer. A quadrupole consists of four parallel rods containing two types of electrical field, fixed direct current (DC) and alternating radio frequency (RF). These electrical fields are manipulated to maintain ions of a desired m/z ratio within the field while removing ions outside of this range (March 2009). Ion traps are quite similar to quadrupoles but, instead of four parallel rods, ions are placed in an RF within a circular electrode (March 2009). A C-trap, which is essentially a flat pole mass analyzer which confines the ions, is paired to an orbitrap. Within the Orbitrap system, a C-trap is utilized to inject small packets of ions into the orbitrap (Eliuk and Makarov 2015). Once inside the Orbitrap, ions rotate around the differentially charged central rod, which has the shape of an elongated and tapered cylinder (Figure 4A). The unique frequencies of the ions are measured in the orbitrap and using Fourier transformation, this signal is converted into a mass spectra, referred to as MS[1]. In a topN approach, the most intense precursor ions are selected in the immediately following population of ions and then transferred to the higher-energy collision dissociation cell (HCD) for fragmentation. These precursor ions are bombarded with Nitrogen gas to break them into fragment ions, which are brought into the C-trap and transferred to the Orbitrap for analysis, resulting in a MS/MS (MS[2]) spectrum (Eliuk and Makarov 2015).

Although not used within these studies, both the TOF-MS and FTICR-MS have unique applications within mass spectrometry. In a TOF system, analytes are accelerated to the same electrical potential within the drift region, allowing only the mass of the ion to determine its velocity. The time of flight, or the length of arrival from the ion source to the detector, is

measured to determine the m/z ratio. Smaller ions accelerate to the detector more quickly, while larger ions do so more slowly, enabling a unique measurement for each ion. This system allows for more efficient measurement time but lower resolution (Boesl 2017). The FTICR method has the highest resolution among the four mass analyzers discussed. Ions are trapped in a magnetic field and exposed to radio frequency via an excitation plate adjacent to a detection plate. This vibrational frequency is measured and transformed using Fourier transformation to create a mass spectra (Nagornov et al. 2022).

Three widely used types of detectors in mass-spectrometry proteomics are the Faraday cup, electron multipliers, and ion cyclotron resonance found in the orbitrap. For all of these detectors, it is critical to have an ion size that isn't too large (range of 10^2-10^5). Within Faraday cups, the ion beam is channeled directly into the cup. When the ions hit the detector at the rear of the cup, a change in electrical current occurs which is measured. Both the degree of electrical change and abundance of ions with that same electrical change are used to calculate the abundance of a particular m/z of an ion. Within an electron multiplier, ions are channeled within a dynode and strike the surface of the dynode. With each strike, secondary electrons are released and these electrons have increased potential energy. The ions are channeled throughout the dynode and with each strike the initial ion signal is increased (Eliuk and Makarov 2015). Uniquely, the cyclotron resonance detector, within the Orbitrap system, acts as both a detector and an analyzer. As previously described, when ions enter the orbitrap with a high energy potential, they rotate around the differentially charged central rod. This unique frequency of the ions is measured and with a Fourier transformation mass to charge are determined (Medhe 2018).

Data analysis

There are various open-source and proprietary software programs that can translate mass spectra into peptide and protein sequences. In this context, the MaxQuant workflow will be briefly explained, as all analysis was performed using this platform (Cox and Mann 2008). The analysis starts by loading raw files into the software, which contain time-resolved mass spectral scans of parental ions MS, and the subset of further fragmented parental ions of interest (MS/MS). Replicates and experimental samples are defined within the experimental design table. Then, a well-annotated FASTA file is uploaded to enable *in silico* digestion. The program proceeds with

feature detection. Through the algorithms included in the search engine Andromeda, two series of peptide searches are conducted on the MS/MS spectra to identify both the parent and precursor ions (Cox et al. 2011). The first search calibrates the mass of the peptides to reduce mass error, while the second search uses the recalibrated mass for identification. After these searches, an Andromeda probability score is assigned to each identified protein using the target-decoy search strategy. The program then calculates the posterior error probability (PEP), which represents the probability that the peptide has been falsely matched to the protein. To calculate this score, the program takes into account the Andromeda probability score, peptide length, charge, and the number of miscleaveages. Peptide misassignment is an unavoidable issue with this method, so to minimize it, a False Discovery Rate (FDR) is calculated by performing the same analysis on a decoy database. The decoy database contains the reverse, shuffled, random, and decoy peptide sequences of the proteins in the initial *in silico* digest (Elias and Gygi 2010). The search is stopped once the FDR has reached 1%. In the final analysis, proteins mapped to the reverse amino acid sequence, proteins with only a single modified peptide for proof of identification, and peptides with seven or less amino acids are filtered out. The pool of proteins that remain after these filtering steps is included in the final analysis (Tyanova, Temu, and Cox 2016).

Quantitative methods in proteomics

The power of top-down proteomics lies not only in the identification of proteins, but also in the measurement of the abundance of their peptides within the sample and in comparison to other conditions. This is achieved through both absolute and relative quantification methods (Bantscheff et al. 2007). Absolute quantification involves the introduction of a known amount of the external standard, which is then compared to the peptide of interest (Gerber et al. 2003). On the other hand, relative quantification relies on different approaches such as metabolic labeling, chemical labeling, or label-free methods. Metabolic labeling entails the introduction of a stable isotope into live tissue. For instance, Stable Isotope Labeling by Amino acids in Cell culture (SILAC) utilizes heavy and light amino acids to distinguish between samples. Chemical labeling encompasses various techniques such as dimethyl labeling (DML) and isobaric mass tag. DML utilizes stable isotopes to induce a mass shift by converting primary amines into dimethyl

amines. These isotopes are classified as heavy, medium, and light, indicating the type of modified isotopes used (Hsu et al. 2003; Boersema et al. 2009). TMT labeling involves a mass reporter and a mass normalizer, in combination known as the isobaric mass tag, as well as a specific amine reactive group that interacts with the peptides (Rauniyar and Yates 2014). With slight but detectable mass differences within the isobaric mass tag, experiments can be multiplexed up to even 18plex. In contrast, label-free quantification does not rely on chemical or metabolic labeling but rather compares intensities across multiple runs. These intensities are normalized against one another and combined with the appropriate statistical testing to determine if there is a significant difference between the intensities (Cox et al. 2014).

Quantitative proteomics in genomic instability

Quantitative mass spectrometry has greatly advanced the field of genomic instability. This tool has allowed for the identification of new factors involved in genomic instability that had been previously uncharacterized (Chen and Chen 2021). Both the use of affinity purification and proximity labeling have been critical to identifying unknown factors, as well as revealing unknown cross talk. Additionally, the use of global proteomic measurements has allowed for a comprehensive view of the DDR. This permits for an unbiased approach studying all aspects of DNA repair, rather than focusing on just the previously associated candidates (Stokes, Zhu, and Farnsworth 2018).

Another significant contribution mass spectrometry has made to the field of DDR is the detection of post-translational modifications. As previously detailed in this thesis, phosphorylation is an integral PTM to all DNA damage responses (Stokes, Zhu, and Farnsworth 2018). Additionally, pol(ADP-ribosyl)ation, ubiquitination, SUMOylation, acetylation, and methylation are critical to the DDR. All excision and double strand break repair pathways previously described are influenced by all or a subset of these modifications (Bai et al. 2020). Phosphorylation, pol(ADP-ribosyl)ation, ubiquitination, and SUMOylation are all detectable by mass spectrometry at a cellular level (Silva et al. 2013; Vivelo and Leung 2015).

Within this thesis, we have utilized quantitative proteomics to address various aspects of DDR through proteomic measurements, affinity purification, DNA affinity purification, phylointeractomic analysis, and temporal proteome profiling. This, in combination with various

molecular biology and biochemical techniques, allowed for novel contributions to the field of genomic instability.

Aims of the thesis

There are three targeted aims in this thesis: 1) characterize novel telomere-end binders in *C. elegans*, 2) characterize novel DNA damage lesion binders across the tree of life, and 3) study the kinetics of the DNA damage response in *Tetrahymena thermophila*.

Previously, phylointeractomic screens have been used to study the evolution of proteins binding telomeres across the vertebrate lineage while also identifying novel telomere end binders. Here, we utilized the same workflow to discover novel telomere end binders in *C. elegans*. Through both dimethyl labeled (DML) and label free quantitative (LFQ) mass spectrometry, we identified a number of novel and specific telomere binders. Until now, there was not any *in vivo* characterization of double-stranded telomere end binders in *C. elegans*, and we aimed to identify and characterize these novel candidates in hopes of identifying double-stranded telomere binder(s). Upon thorough characterization, we have conclusive evidence showing TEBP-1 and TEBP-2 are those double-stranded telomere end binders.

Secondarily, we utilized a similar workflow to study the recognition and repair of DNA damage lesions. We focused this study to 8-oxoguanine (8-oxoG) and an abasic site, which are both generally recognized and repaired by base excision repair pathways, as well as an incorporated uracil in DNA. By studying these pathways across the tree of life, we hope to gain insights into the convergence and divergence of these different repair machinery. We anticipated identifying known DNA damage interactors as well as novel candidates. Including organisms across all three domains of life, this study recapitulates previous findings and reveals new candidate proteins with the potential to be involved in DNA damage repair. We provide a large dataset that can be used to propel new discoveries within these specific DNA repair pathways and model organisms.

Finally, we took a systems view of the kinetics of the DNA damage response in *Tetrahymena thermophila*. We treated the cells with six common DNA damaging agents, which are all associated with particular DNA damage repair pathways. Both RNA and protein samples were harvested over an 8 hour time course (H0, H1, H2, H3, H4, H6, H8). We aimed to have an appreciation of both the unique and core dynamic response to these treatments. Additionally, we uncovered dynamic genes and proteins not previously related or understood fully in the context of repair. A large goal of this project was to make the dataset accessible and thus all data and analysis will be available at: https://butterlab.imb-mainz.de/Tt_DDR/.

Publications

The double-stranded DNA-binding proteins TEBP-1 and TEBP-2 form a telomeric complex with POT-1

Summary

In our recent Nature Communications publication, we identified novel telomere binders in *C. elegans* using a telomere DNA pulldown coupled to both label free (LFQ) and dimethyl labeled (DML) quantitative mass spectrometry (MS) (Kappei et al. 2017; Dietz et al. 2021). We then characterized two of these binders, Telomere End Binding Protein 1 and 2 (TEBP-1 and TEBP-2). First, *in vitro* characterization revealed unique and specific binding to double-stranded telomeric sequences with nanomolar affinity. Both proteins were tagged *in vivo* and had nuclear localization as well as co-localization with POT-1, the previously characterized single-stranded telomere end binder. Upon deletion of *tebp-1* and *tebp-2* there was lengthening and shortening of the telomeres, respectively. When these deletion strains were crossed together, the progeny exhibited synthetic sterility and extreme atrophy of the germline arms.

To elucidate the interactors of TEBP-1 and TEBP-2, we performed IP-MS with TEBP-1::3xFLAG and TEBP-2::GFP strains. TEBP-1 and TEBP-2 interact with each other and with known single-stranded telomere end binders, POT-1, MRT-1, and POT-2. A Yeast Two-Hybrid (Y2H) experiment revealed that POT-1 is a direct interactor of TEBP-1 and TEBP-2. POT-1 serves as the critical link between the double-stranded and single-stranded telomere end binders. We mapped the interaction interface between POT-1 with TEBP-1 and TEBP-2. We also determined the TEBP-1 and TEBP-2 interaction interface as well as the DNA binding domain of both TEBP-1 and TEBP-2. Finally, we established the critical nature of both TEBP-1 and TEBP-2 across other nematode species.

Altogether, this study established the first ever double-stranded telomere end binders in *C. elegans*, TEBP-1 and TEBP-2. These proteins are critical to regulating telomere homeostasis, and fertility. TEBP-1 and TEBP-2 directly interact with the previously characterized POT-1 single-stranded binder, which serves as a bridge to the other telomere binders in *C. elegans*.

Zusammenfassung

In unserer kürzlich erschienenen Publikation in Nature Communications identifizierten wir neuartige Telomerbinder in *C. elegans* mithilfe eines Telomer-DNA-Pulldowns, der sowohl mit markierungsfreier (LFQ) als auch mit Dimethyl-markierter (DML) quantitativer Massenspektrometrie (MS) gekoppelt war (Kappei et al. 2017; Dietz et al. 2021). Anschließend charakterisierten wir zwei dieser Binder, Telomere End Binding Protein 1 und 2 (TEBP-1 und TEBP-2). Zunächst zeigte die in vitro-Charakterisierung eine einzigartige und spezifische Bindung an doppelsträngige telomere Sequenzen mit nanomolarer Affinität. Beide Proteine wurden in vivo markiert und wiesen eine Kernlokalisierung sowie eine Ko-Lokalisierung mit POT-1 auf, dem zuvor charakterisierten Binder für einzelsträngige Telomerenden. Nach Deletion von tebp-1 und tebp-2 kam es zu einer Verlängerung bzw. Verkürzung der Telomere. Wenn diese Deletionsstämme miteinander gekreuzt wurden, zeigten die Nachkommen synthetische Sterilität und extreme Atrophie der Keimbahnarme.

Um die Interaktoren von TEBP-1 und TEBP-2 aufzuklären, führten wir IP-MS mit TEBP-1::3xFLAG und TEBP-2::GFP Stämmen durch. TEBP-1 und TEBP-2 interagieren miteinander und mit den bekannten einzelsträngigen Telomerendbindern POT-1, MRT-1 und POT-2. Ein Yeast Two-Hybrid (Y2H) Experiment zeigte, dass POT-1 ein direkter Interaktor von TEBP-1 und TEBP-2 ist. POT-1 dient als kritisches Bindeglied zwischen den doppelsträngigen und einzelsträngigen Telomerendbindern. Wir kartierten die Interaktionsschnittstelle zwischen POT-1 und TEBP-1 und TEBP-2. Außerdem haben wir die Interaktionsschnittstelle von TEBP-1 und TEBP-2 sowie die DNA-Bindungsdomäne von TEBP-1 und TEBP-2 bestimmt. Schließlich haben wir die kritische Natur von TEBP-1 und TEBP-2 bei anderen Nematodenarten festgestellt. Insgesamt wurden in dieser Studie die ersten doppelsträngigen Telomerendbinder in *C. elegans*, TEBP-1 und TEBP-2, nachgewiesen. Diese Proteine sind entscheidend für die Regulierung der Telomer-Homöostase und der Fruchtbarkeit. TEBP-1 und TEBP-2 interagieren direkt mit dem zuvor charakterisierten einzelsträngigen Binder POT-1, der als Brücke zu den anderen Telomerbindern in *C. elegans* dient.

Statement of contribution

This project was initiated and led by Sabrina Dietz and Miguel Vasconcelos Almeida. Within this project, I established protein expression constructs, optimized appropriate expression conditions, and purified full length and fragments of TEBP-1 and TEBP-2. I established appropriate assays and performed data analysis for characterization of binding affinity of TEBP-1 and TEBP-2. I established the DNA-binding fragment in TEBP-1 and TEBP-2. I supported strain establishment, maintenance, and harvest. I gave in depth feedback on data presentation and manuscript composition.

Supervisor Confirmation:_____ *Falk Buttel* _____

The double-stranded DNA-binding proteins TEBP-1 and TEBP-2 form a telomeric complex with POT-1

Sabrina Dietz [1,6], Miguel Vasconcelos Almeida [1,4,5,6], Emily Nischwitz[1], Jan Schreier[1], Nikenza Viceconte [1], Albert Fradera-Sola [1], Christian Renz[1], Alejandro Ceron-Noriega[1], Helle D. Ulrich[1], Dennis Kappei [2,3], René F. Ketting [1] & Falk Butter [1✉]

Telomeres are bound by dedicated proteins, which protect them from DNA damage and regulate telomere length homeostasis. In the nematode *Caenorhabditis elegans*, a comprehensive understanding of the proteins interacting with the telomere sequence is lacking. Here, we harnessed a quantitative proteomics approach to identify TEBP-1 and TEBP-2, two paralogs expressed in the germline and embryogenesis that associate to telomeres in vitro and in vivo. *tebp-1* and *tebp-2* mutants display strikingly distinct phenotypes: *tebp-1* mutants have longer telomeres than wild-type animals, while *tebp-2* mutants display shorter telomeres and a Mortal Germline. Notably, *tebp-1;tebp-2* double mutant animals have synthetic sterility, with germlines showing signs of severe mitotic and meiotic arrest. Furthermore, we show that POT-1 forms a telomeric complex with TEBP-1 and TEBP-2, which bridges TEBP-1/-2 with POT-2/MRT-1. These results provide insights into the composition and organization of a telomeric protein complex in *C. elegans*.

[1] Institute of Molecular Biology (IMB), Mainz, Germany. [2] Cancer Science Institute of Singapore, National University of Singapore, Singapore, Singapore. [3] Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [4] Present address: Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge, UK. [5] Present address: Department of Genetics, University of Cambridge, Cambridge, UK. [6] These authors contributed equally: Sabrina Dietz, Miguel Vasconcelos Almeida. ✉email: f.butter@imb-mainz.de

32

Most telomeres in linear eukaryotic chromosomes end in tandem repeat DNA sequences. Telomeres solve two major challenges of chromosome linearity: the end-protection problem and the end-replication problem[1,2]. The end-protection problem originates from the structural similarity between telomeres and DNA double-strand breaks, which can lead to recognition of the telomere by the DNA damage surveillance machinery[2]. When telomeres are falsely recognized as DNA damage, they are processed by the non-homologous end joining or homologous recombination pathways, leading to genome instability[3,4]. The end-replication problem arises from the difficulties encountered by the DNA replication machinery to extend the extremities of linear chromosomes, which results in telomere shortening with every cell division[5–7]. When a subset of telomeres shorten beyond a critical point, cellular senescence or apoptosis are triggered[8–10].

Specialized proteins have evolved to deal with the complications arising from telomeres, which in vertebrates are composed of double-stranded (ds) $(TTAGGG)_n$ repeats ending in a single-stranded (ss) 3' overhang[11]. In mammals, a telomere-interacting complex of six proteins termed shelterin constitutively binds to telomeres in mitotic cells[12]. This complex consists of the ds telomere binders TRF1 and TRF2, the TRF2-interacting protein RAP1, the ss binding protein POT1 and its direct interactor TPP1, as well as the bridging protein TIN2. Altogether, the proteins of this complex shield telomeres from a DNA damage response by inhibiting aberrant DNA damage signaling[3]. In addition, shelterin components are required for the recruitment of the telomerase enzyme, which adds de novo repeats to the telomeric ends, allowing maintenance of telomere length in dividing cells[6]. Telomerase is a ribonucleoprotein, comprised of a catalytic reverse-transcriptase protein component and an RNA moiety. Besides the core shelterin complex, additional proteins have been described to interact with telomeres and assist in the maintenance of telomere length, e.g., HMBOX1 (also known as HOT1), ZBTB48 (also known as TZAP), NR2C2, and ZNF827[13–17].

In *Schizosaccharomyces pombe*, a shelterin-like complex harboring orthologs of the human shelterin complex was described[18–20]. TAZ1 and POT1 bind to ds and ss telomeric DNA similar to their human counterparts TRF1/TRF2 and POT1, respectively. In turn, *Saccharomyces cerevisiae* has distinct complexes binding to the ds and ss telomere[21–26]. The *S. cerevisiae* ortholog of the TRF2-interacting protein RAP1 binds ds telomeric DNA through two domains structurally related to Myb domains[27]. The ss overhang is not bound by a POT1 homolog but rather by the CST complex[22,23,25]. Overall, this indicates that different telomere-binding complexes have evolved across species to alleviate the challenges of linear chromosome ends, based on variations of recurring DNA-binding modules.

The nematode *Caenorhabditis elegans* has been employed in many seminal discoveries in molecular biology, genetics, and development[28]. Its telomeres have a repeat sequence similar to vertebrate telomeres, consisting of $(TTAGGC)_n$[29]. Moreover, *C. elegans* telomeres have a length of about 2–9 kb[29,30], and it has been proposed that its telomeric structures have both 5' and 3' ss overhangs, each recognized by dedicated ss telomere-binding proteins[31]. Telomere maintenance in this nematode is carried out by the catalytic subunit of telomerase TRT-1[32]. The RNA component of *C. elegans* telomerase has not been identified thus far. Telomeres can be maintained by additional mechanisms, since *C. elegans* can survive without a functioning telomerase pathway by employing alternative lengthening of telomere (ALT)-like mechanisms, creating more heterogeneous telomere lengths[33–37].

In *C. elegans*, four proteins with domains structurally similar to the DNA-binding domain of human POT1 were identified. Three

of those proteins, namely POT-1 (also known as CeOB2), POT-2 (also known as CeOB1), and MRT-1, were confirmed to bind to the ss telomeric overhangs[31,38]. Mutants for these factors show telomere length maintenance defects. Depletion of POT-1 and POT-2 leads to telomere elongation[31,33,35,37], whereas depletion of MRT-1 results in progressive telomere shortening over several generations[38]. Concomitant to telomere shortening, *mrt-1*, *mrt-2*, and *trt-1* mutant animals share a Mortal Germline (Mrt) phenotype, characterized by a gradual decrease in fertility across generations, until animals become sterile[30,32,38]. MRT-1 was proposed to be in a pathway for facilitation of telomere elongation together with the DNA damage checkpoint protein MRT-2, and telomerase TRT-1[38]. Despite the identification of these different telomere-associated proteins, no telomere-binding complex has been described in *C. elegans* yet.

In this work, we performed a quantitative proteomics screen to identify novel telomere-binding proteins in *C. elegans*. We report the identification and characterization of R06A4.2 and T12E12.3, two previously uncharacterized paralog genes, which we named telomere-binding proteins 1 and 2 (*tebp-1* and *tebp-2*), respectively. TEBP-1 and TEBP-2 bind to the ds telomeric sequence in vitro with nanomolar affinity and co-localize with POT-1, a known telomere binder, in vivo. *tebp-1* and *tebp-2* mutants have contrasting effects on telomere length: while *tebp-1* mutants display elongated telomeres, *tebp-2* mutants have shortened telomeres. In addition, TEBP-1 and TEBP-2 have important roles in fertility, as *tebp-1; tebp-2* double mutants are synthetic sterile. Size-exclusion chromatography and interaction studies demonstrate that TEBP-1 and TEBP-2 are part of a complex with POT-1, which bridges the ds telomere binders, TEBP-1 and TEBP-2, with the ss telomere binders POT-2 and MRT-1.

Results

TEBP-1 (R06A4.2) and TEBP-2 (T12E12.3) are double-stranded telomere-binding proteins in *Caenorhabditis elegans*. To identify proteins that bind to the *C. elegans* telomeric sequence, we employed a DNA pulldown assay (Supplementary Fig. 1a, b) previously used to successfully identify telomeric proteins in other species[15,16,39,40]. We incubated concatenated, biotinylated DNA oligonucleotides consisting of either the telomeric sequence of *C. elegans* $(TTAGGC_n)$, or a control sequence $(AGGTCA_n)$, with nuclear-enriched extracts of gravid adult worms. The experiment was performed twice using two different quantitative proteomics approaches: label-free quantitation (LFQ)[41] and reductive dimethyl labeling (DML)[42], which yielded 12 and 8 proteins enriched in telomeric sequence pulldowns, respectively, with an overlap of 8 proteins (Fig. 1a, b and Supplementary Fig. 1a, b). Among these eight proteins, we found the already known ss telomere binders POT-1, POT-2, and MRT-1[31,33,37,38], as well as the CKU-70/CKU-80 heterodimer[43], and three additional proteins: R06A4.2, T12E12.3, and DVE-1.

R06A4.2 and T12E12.3 were of particular interest, as they share 74.3% DNA coding sequence identity and 65.4% amino acid sequence identity (Supplementary Fig. 1c), suggesting that R06A4.2 and T12E12.3 are paralogs. While R06A4.2 and T12E12.3 lack any annotated protein domain, using HHpred v3.2.0[44], we could determine that the N-terminal region of both proteins shows similarity to the homeodomains of human and yeast RAP1 (Supplementary Fig. 1d, e and Supplementary Data file 1). RAP1 is a direct ds telomere binder in budding yeast[21,45], and a member of the mammalian shelterin complex through interaction with TRF2[46].

We validated binding of R06A4.2 and T12E12.3 to telomeric DNA by performing DNA pulldowns with His-tagged recombinant proteins (Fig. 1c). Using CRISPR-Cas9 genome editing, we
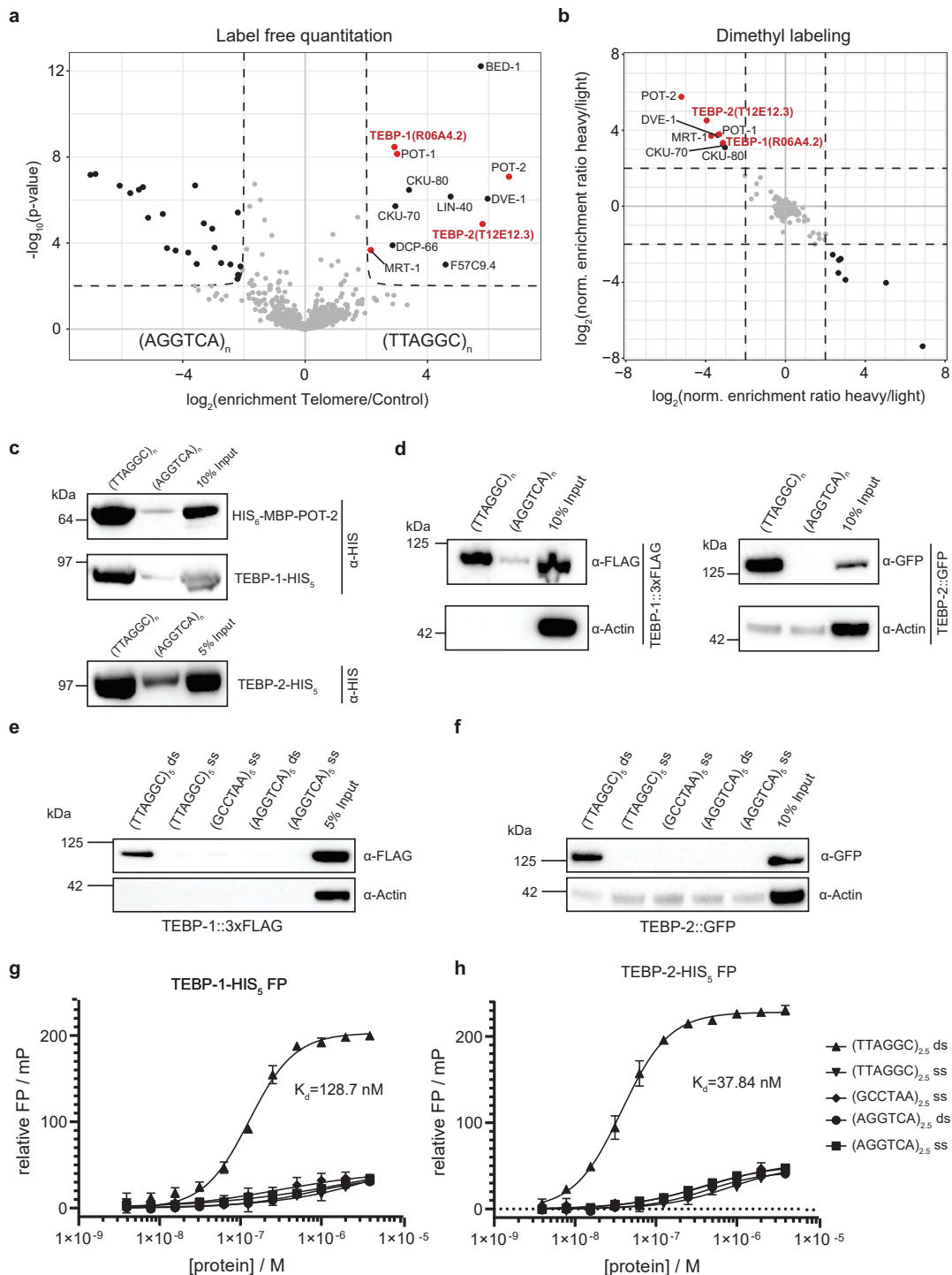
inserted a *gfp* and a *3xflag* sequence directly upstream of the endogenous stop codon of *T12E12.3* and *R06A4.2*, respectively (Supplementary Fig. 1d, e). Using these strains, we could show that the endogenously tagged versions of R06A4.2 and T12E12.3 also bind to the *C. elegans* telomere sequence (Fig. 1d).

Owing to the preparation strategy, our concatenated DNA probes contained both ds and ss DNA, which precludes any conclusions about whether R06A4.2 and T12E12.3 bind ss or ds telomeric DNA. We thus performed additional DNA pulldowns with ss and ds probes specifically designed with five repeats $(TTAGGC)_5$. Both proteins were found to exclusively bind to the ds telomeric repeats, establishing R06A4.2 and T12E12.3 as ds telomere binders (Fig. 1e, f). To confirm and quantify the interaction of R06A4.2 and T12E12.3 with ds telomeric DNA, we

Fig. 1 TEBP-1 (R06A4.2) and TEBP-2 (T12E12.3) are double-stranded telomere binders in *C. elegans*. a Volcano plot representing label-free proteomic quantitation of pulldowns with biotinylated, concatenated oligonucleotide baits of telomeric DNA sequence (TTAGGC)$_n$ or control DNA sequence (AGGTCA)$_n$. Pulldowns were performed with nuclear extracts from synchronized gravid adult animals, in octuplicates per condition (two biological replicates, each with four technical replicates). Log$_2$ fold enrichment of proteins in one condition over the other is presented on the x-axis. The y-axis shows $-\log_{10}$ p-value (Welch t-test) of enrichment across replicates. More than 4-fold enriched proteins with p-value < 0.01 are annotated as black dots, the background proteins as gray dots. Enriched proteins of interest, such as the known ss telomere binders, are annotated as red dots. **b** Scatterplot representing results of reductive dimethyl-labeling-based quantitation of pulldowns with the same extract and DNA baits as in (**a**). Per condition, pulldowns were performed in duplicates and labeled on the peptide level, including an intra-experimental label switch to achieve cross-over sets. The x-axis represents log$_2$ transformed ratios of the reverse experiment, whereas the y-axis represents log$_2$ transformed ratios of the forward experiment (see Supplementary Fig. 1b). Single proteins are depicted by dots in the scatterplot. Enriched proteins (threshold > 4) are annotated as black dots, background proteins as gray dots, and enriched proteins of interest as red dots. **c** Binding of recombinant His-tagged POT-2, TEBP-1 and TEBP-2, from crude *E. coli* lysate, to telomere or control DNA as in (**a**). Chemiluminescence western blot read-out, after probing with α-His antibody. POT-2 is used as a positive control for telomeric repeat binding. MBP: Maltose-binding protein, kDa: kilodalton. Uncropped blots in Source Data. N = 2 biologically independent experiments with similar results, except POT-2 N = 1. **d** DNA pulldowns as in **c** but on embryo extracts of transgenic *C. elegans* lines carrying either TEBP-1::3xFLAG or TEBP-2::GFP. N = 2 independent experiments with similar results, **e**, **f** DNA pulldowns with 5x telomeric (TTAGGC) double-strand (ds) repeats and both respective single-strand (ss) baits, and 5x control (AGGTCA) ds or 5x (AGGTCA) ss repeats. Pulldowns were performed with embryo extracts of TEBP-1::3xFLAG or TEBP-2::GFP animals. Uncropped blots in Source Data. N = 3 biologically independent experiments with similar results, **g**, **h** Fluorescence polarization assays of 4 μM to 4 nM purified TEBP-1-His$_5$ and TEBP-2-His$_5$, respectively. Binding affinities to 2.5x ss and ds telomeric and control repeats of FITC-labeled oligonucleotides. Error bars represent+/- the standard deviation of the mean values. Per data point n = 3 technical replicates. FP, fluorescence polarization; mP, millipolarization, upward triangle: 2.5x TTAGGC double-strand, downward triangle: 2.5x TTAGGC single-strand, diamond: 2.5x GCCTAA single-strand, circle: 2.5x shuffled control double-strand, square: 2.5x shuffled control single-strand.

performed fluorescence polarization with purified, recombinant proteins and FITC-labeled oligonucleotides. Both T12E12.3 and R06A4.2 displayed affinity for the ds telomeric repeat sequence in the nanomolar range (K_d = 128.7 nM for R06A4.2 and K_d = 37.84 nM for T12E12.3, Fig. 1g, h). Both T12E12.3 and R06A4.2 showed highest affinity for the 2.5x telomeric repeat, when incubated with a 2.5x, 2.0x, 1.5x T-rich, and 1.5x G-rich telomeric repeat sequences (Supplementary Fig. S2a–c).

In conclusion, we demonstrate that R06A4.2 and T12E12.3, two proteins with highly similar sequence, bind directly and with high affinity to the *C. elegans* ds telomeric DNA sequence in vitro. Thus, we decided to name R06A4.2 as Telomere-Binding Protein-1 (TEBP-1) and T12E12.3 as Telomere-Binding Protein-2 (TEBP-2).

TEBP-1 and TEBP-2 localize to telomeres in proliferating cells in vivo. To explore the expression pattern of *tebp-1* and *tebp-2* throughout animal development, we used a recently published mRNA-seq dataset[47]. Both genes show the highest expression in embryos, very low abundance during the L1–L3 larval stages, and an increase in expression in L4 larvae and young adults (YAs, Supplementary Fig. 3a–c). The observed increase in *tebp-1* and *tebp-2* mRNA expression from the L4 to YA stages coincides with the increased progression of germline development, which may hint to a higher expression level during gametogenesis. Indeed, using available gonad-specific RNA-seq datasets[48], we confirmed that *tebp-1 and tebp-2* are expressed in spermatogenic and oogenic gonads (Supplementary Fig. 3d). Similar developmental mRNA expression patterns were also found for the known ss telomere binders *pot-1*, *pot-2*, and *mrt-1* (Supplementary Fig. 3a, d). To study the expression at the protein level, we crossed our endogenously tagged strains to generate a *tebp-1::3xflag*; *tebp-2::gfp* strain to monitor protein abundance simultaneously by western blot. The protein expression patterns of TEBP-1 and TEBP-2 are highly similar to the RNA-seq data, with highest detected expression in embryos, a drop during the larval stages L1-L4, ultimately followed by an increase in YA (Fig. 2a).

To study TEBP-1 and TEBP-2 localization in vivo, we focused on embryos and on the germline of adult animals. In these two actively dividing tissues, TEBP-1 and TEBP-2 protein expression is high and condensed chromosomes facilitate visualization of telomeric co-localization. In addition to the *tebp-2::gfp* strain used above, we also generated an endogenously tagged *tebp-1::gfp*

allele, using CRISPR-Cas9 genome editing (Supplementary Fig. 1d). To check for telomeric localization in vivo, we crossed *tebp-1::gfp* and *tebp-2::gfp* each with a germline-specific *pot-1::mCherry* single-copy transgene[37], and imaged the dual-fluorescent animals. TEBP-1::GFP and TEBP-2::GFP co-localize with POT-1::mCherry inside the nuclei of oocytes and embryos (Fig. 2b–e). Confocal microscopy of TEBP-1::GFP in combination with POT-1::mCherry was challenging likely due to bleaching of TEBP-1::GFP. Co-localization of TEBP-2::GFP and POT-1::mCherry was also observed in the mitotic region of the germline and in mature sperm (Fig. 2d). These results clearly establish that TEBP-1 and TEBP-2 co-localize with a known telomeric binder in vivo in proliferating tissues, indicating that their ability to bind ds telomeric DNA in vitro may have functional relevance.

TEBP-1 and TEBP-2 have opposing telomere length phenotypes. As TEBP-1 and TEBP-2 localize to telomeres, we sought to address whether these proteins regulate telomere length, as is the case for the known ss telomere-binding proteins POT-1, POT-2, and MRT-1[31,33,37,38]. Using CRISPR-Cas9 genome editing, we generated *tebp-1* and *tebp-2* deletion mutants encoding truncated transcripts with premature stop codons (Supplementary Fig. 1d–g and Supplementary Fig. 4a, b). *tebp-1* and *tebp-2* mutants are viable and show no immediate, obvious morphological or behavioral defects. We analyzed telomere length in the mutants after propagation for more than 100 generations, sufficient to establish a "steady-state" telomere length phenotype, by carrying out a telomere Southern blot on mixed-stage animals. Interestingly, while *tebp-1(xf133)* shows an elongated telomere phenotype comparable to the *pot-2(tm1400)* mutant, *tebp-2(xf131)* shows a shortened telomere phenotype (Fig. 3a), similar to *mrt-1* mutants[38]. In addition, we performed quantitative fluorescence in situ hybridization (qFISH) in dissected adult germlines, which confirmed our initial observation that *tebp-1* and *tebp-2* mutants display longer or shorter telomeres than wild-type, respectively (Fig. 3b–f). Furthermore, we also measured telomere length in embryos by qFISH. Like in the germline, the telomeres of *tebp-1* mutant embryos are elongated, while the telomeres of *tebp-2* embryos are shortened (Supplementary Fig. 4c–g).

In summary, *tebp-1* and *tebp-2* mutants display opposing regulatory effects on telomere length. These experiments suggest that the TEBP-1 protein counteracts telomere elongation
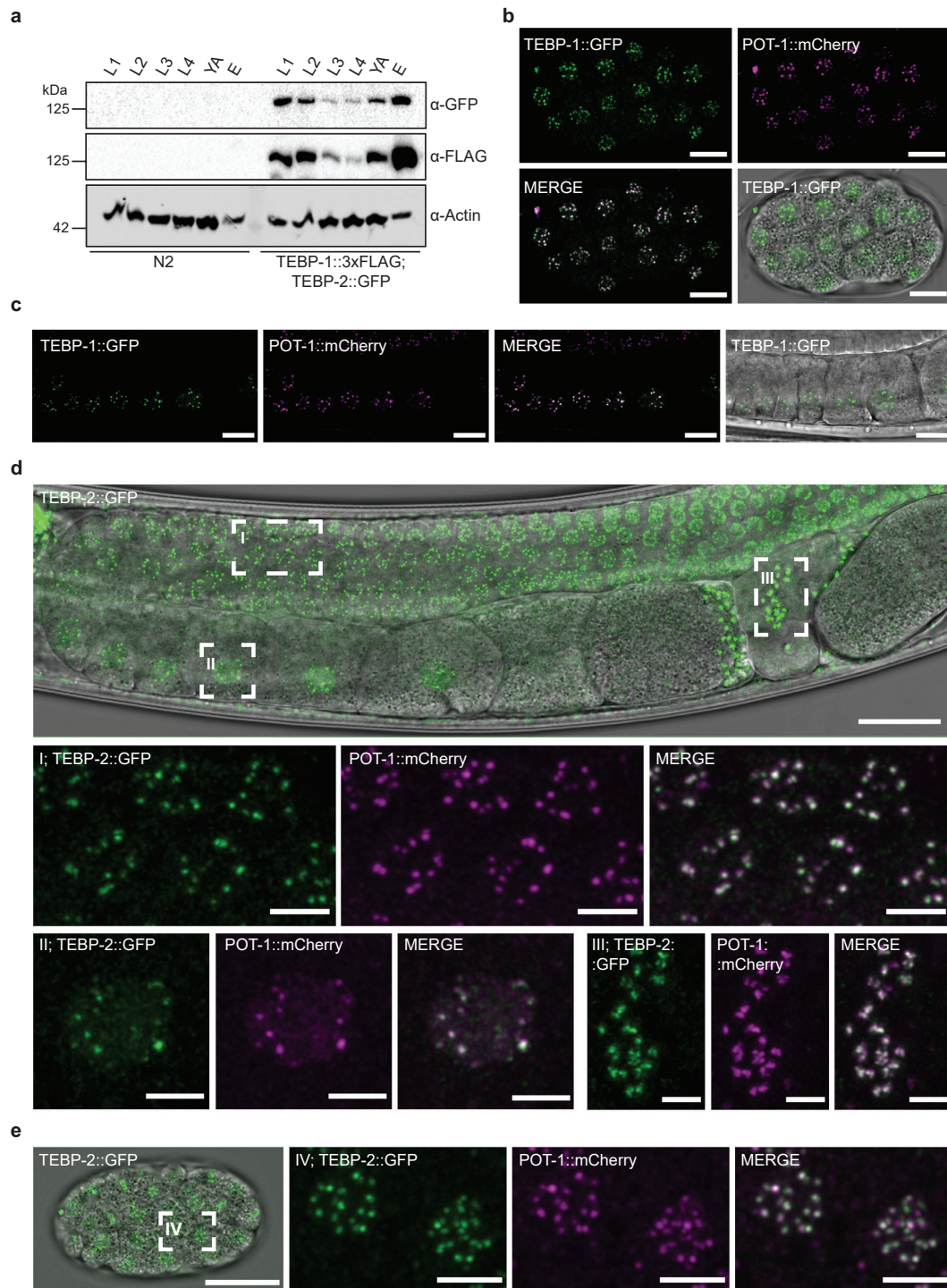
Fig. 2 TEBP-1 and TEBP-2 are expressed throughout *C. elegans* development and localize to telomeres in vivo. a Western blot of TEBP-1::3xFLAG and TEBP-2::GFP expression in different developmental stages of *C. elegans*. Thirty-five micrograms of extract from either N2 or a double transgenic line carrying TEBP-1::3xFLAG and TEBP-2::GFP were used. Actin was used as loading control. kDa: kilodalton. Uncropped blot in Source Data. *N* = 1 **b, c** Maximum intensity projections of representative confocal *z*-stacks of an embryo (**b**), or oocytes (**c**) expressing endogenously tagged TEBP-1::GFP and transgenic POT-1::mCherry. Scale bars, 10 μm. **d, e** Maximum intensity projections of representative confocal *z*-stacks of an adult animal (**d**), or embryo (**e**) expressing both endogenously tagged TEBP-2::GFP and transgenic POT-1::mCherry. Insets show nuclear co-localization in meiotic germ cell nuclei (I), an oocyte (II), spermatozoa (III), and embryonic cells (IV). Scale bars, 20 μm (overview) and 4 μm (insets). All microscopy images were deconvoluted using Huygens remote manager. Representative images from two individual animals per strain, *N* = 2 biologically independent experiments with similar results.
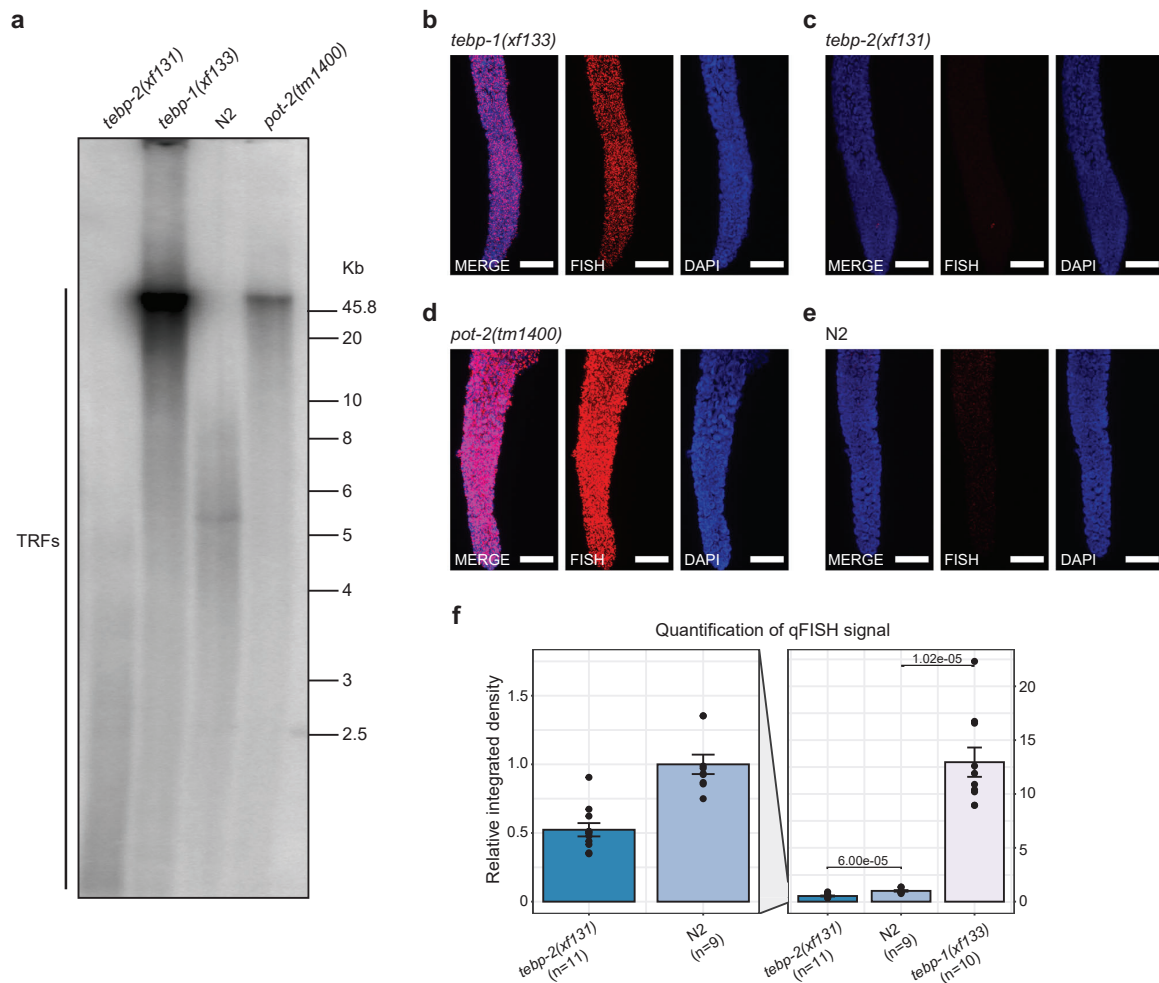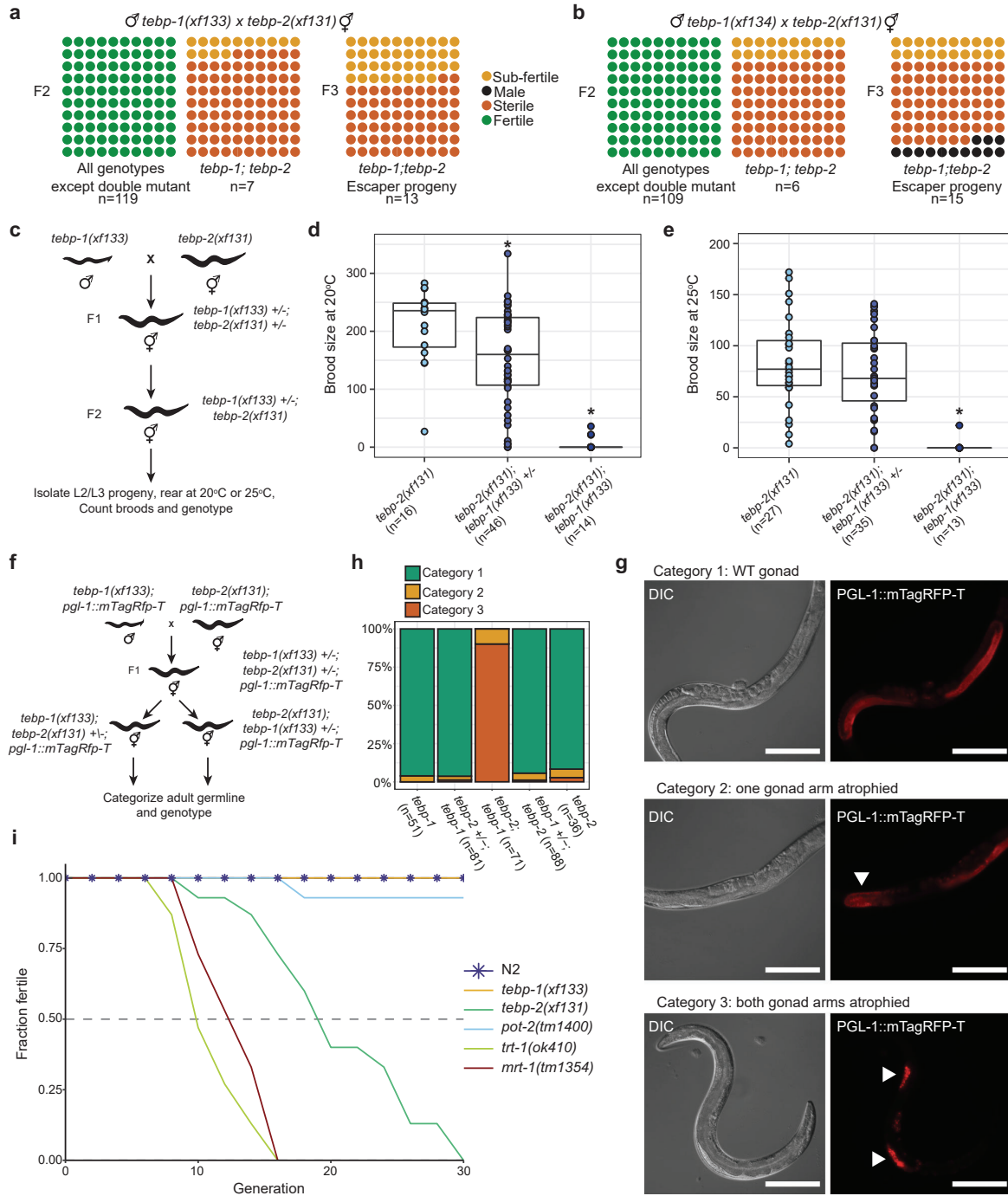
Fig. 3 TEBP-1 and TEBP-2 regulate telomere length. a Southern blot analysis of *C. elegans* telomeres. DNA from four different strains (*tebp-1(xf133)* grown for ~102 generations; *tebp-2(xf131)*, grown for ~124 generations; N2, and *pot-2(tm1400)*) was digested and separated by agarose gel electrophoresis. DNA was transferred to a positively charged nylon membrane and hybridized with a radiolabeled (GCCTAA)$_3$ oligonucleotide. Brightness and contrast of the membrane read-out were adjusted using Fiji. Telomere restriction fragments (TRFs) are indicated in the Fig.. Uncropped blot in Source Data. $N = 3$ independent experiments with similar results. **b–e** Representative maximum projection *z*-stacks of a qFISH assay using dissected adult germlines of the following *C. elegans* mutant strains: *tebp-1(xf133)* (grown for ~98 generations), *tebp-2(xf131)* (grown for ~120 generations), *pot-2(tm1400)*, and wild-type N2. The telomeres of dissected worms of the respective strains were visualized by hybridization with a telomeric PNA-FISH-probe. Nuclei were stained with DAPI. Scale bars, 15 μm. **f** Barplot depicting analysis of qFISH images of the strains in (**b–c**) and (**e**). Average telomere length is indicated by arbitrary units of relative integrated density, with wild-type N2 set to 1. The plot on the left shows the *tebp-2(xf131)* and N2 values zoomed-in. Analyzed n per strain derived from independent animals: *tebp-2(xf131)*: $n = 11$, N2: $n = 9$, *tebp-1(xf133)*: $n = 10$. Error bars represent the standard error of the mean (SEM) and *p*-values were calculated using Welch's *t*-test. $N = 3$ biologically independent experiments with similar results.

independently of telomerase, while TEBP-2 promotes telomere lengthening.

Simultaneous lack of TEBP-1 and TEBP-2 leads to synthetic sterility. To better understand how *tebp-1* and *tebp-2* mutants distinctly affect telomere length, we intended to measure telomere length in *tebp-1; tebp-2* double mutants. Surprisingly, when we crossed our single mutants, we could not establish a double homozygous *tebp-1; tebp-2* mutant strain. In fact, *tebp-1; tebp-2* double mutants displayed highly penetrant synthetic sterility (Fig. 4a). Repeating the cross with another *tebp-1* mutant allele (*xf134*), as well as the reciprocal cross, yielded the same synthetic sterility (Fig. 4b and Supplementary Fig. 5a). Only about 14–38%

of F2 or F3 *tebp-1; tebp-2* animals did not have synthetic sterility (Fig. 4a, b). These "synthetic sterility escapers" were subfertile, siring less than 60 offspring. Importantly, a *tebp-2::gfp* single-copy transgene fully rescued the appearance of sterility, demonstrating that the C-terminal tag does not disrupt TEBP-2 function (Supplementary Fig. 5a). When we combined *tebp-1* mutant animals with *mrt-1*, *trt-1*, or *pot-2* mutations, or *tebp-2* mutant animals with *trt-1* or *pot-2*, the double mutant offspring was fertile (Supplementary Fig. 5a). These results demonstrate that the synthetic sterility is specific to *tebp-1; tebp-2* double mutants, and is not a consequence of crossing shorter telomere mutants with longer telomere mutants. We further quantified the synthetic sterility on brood size by picking L2-L3 progeny of *tebp-2; tebp-1 +/−* mutants, blind to genotype and germline health, rearing those

a ♂ *tebp-1(xf133)* x *tebp-2(xf131)* ♀

All genotypes except double mutant n=119

tebp-1; tebp-2 n=7

tebp-1;tebp-2 Escaper progeny n=13

Sub-fertile
Male
Sterile
Fertile

b ♂ *tebp-1(xf134)* x *tebp-2(xf131)* ♀

All genotypes except double mutant n=109

tebp-1; tebp-2 n=6

tebp-1;tebp-2 Escaper progeny n=15

c

tebp-1(xf133) ♂ x *tebp-2(xf131)* ♀

F1 *tebp-1(xf133) +/-; tebp-2(xf131) +/-* ♂

F2 *tebp-1(xf133) +/-; tebp-2(xf131)* ☿

Isolate L2/L3 progeny, rear at 20°C or 25°C, Count broods and genotype

d Brood size at 20°C

tebp-2(xf131) (n=16)

tebp-2(xf131); tebp-1(xf133) +/- (n=46)

tebp-2(xf131); tebp-1(xf133) (n=14)

e Brood size at 25°C

tebp-2(xf131) (n=27)

tebp-2(xf131); tebp-1(xf133) +/- (n=35)

tebp-2(xf131); tebp-1(xf133) (n=13)

f

tebp-1(xf133); pgl-1::mTagRfp-T ♂ x *tebp-2(xf131); pgl-1::mTagRfp-T* ♀

F1 *tebp-1(xf133) +/-; tebp-2(xf131) +/-; pgl-1::mTagRfp-T* ♂

tebp-2(xf131); tebp-1(xf133) +/-; pgl-1::mTagRfp-T ♂   *tebp-2(xf131); tebp-1(xf133) +/-; pgl-1::mTagRfp-T* ☿

Categorize adult germline and genotype

h

Category 1
Category 2
Category 3

tebp-1 (n=51)
tebp-2 +/-; tebp-1 (n=81)
tebp-2; tebp-1 (n=71)
tebp-1 +/-; tebp-2 (n=88)
tebp-2 (n=36)

g

Category 1: WT gonad
DIC   PGL-1::mTagRFP-T

Category 2: one gonad arm atrophied
DIC   PGL-1::mTagRFP-T

Category 3: both gonad arms atrophied
DIC   PGL-1::mTagRFP-T

i

Fraction fertile

N2
tebp-1(xf133)
tebp-2(xf131)
pot-2(tm1400)
trt-1(ok410)
mrt-1(tm1354)

Generation

animals at 20 °C or 25 °C, later counting their brood sizes, and genotyping each animal (Fig. 4c–e). This revealed that the immediate synthetic sterility phenotype is not dependent on temperature, as the reduction of progeny numbers was apparent at both 20 and 25 °C.

Morphologically, *tebp-1; tebp-2* double mutants displayed a degenerated germline. To visualize this phenotype, we created *tebp-1* and *tebp-2* strains in combination with an endogenously tagged *pgl-1::mTagRfp-T* allele[49,50], which we used as a germ

cell reporter. PGL-1 is expressed in P-granules, perinuclear granules most important for germline development and gene regulation[51,52]. As depicted in Fig. 4f, we repeated the *tebp-1* x *tebp-2* cross with *pgl-1::mTagRfp-T* in the background, isolated cross progeny of the indicated genotypes, reared these animals to adulthood, scored them into three categories of germline morphology, and genotyped them afterwards. The categories can be described as follows: category 1 animals displayed a wild-type or near wild-type morphology (Fig. 4g, upper panels),

Fig. 4 *tebp-1; tebp-2* double mutants have synthetic sterility, and *tebp-2* mutants have a Mortal Germline. a, b Schematics depicting the quantification of fertility of the F2 (two panels on the left) and F3 (panel on the right) cross progeny of the indicated crosses. Each dot represents 1% of the indicated n per square, in a 10 × 10 matrix for 100%. Green dots indicate fertile worms, yellow dots subfertile worms (<60 progeny), orange dots sterile worms, and black dots indicate male worms. The F3 animals used for the panels on the right were the progeny of subfertile F2s, which escaped synthetic sterility. Males with two different *tebp-1* mutant alleles, *xf133* and *xf134*, were used in (**a**) and (**b**), respectively. **c** Schematic of cross performed with *tebp-1(xf133)* and *tebp-2(xf131)* to isolate progeny for determination of brood size at 20 and 25 °C. **d, e** Brood sizes of cross progeny animals, isolated as indicated in (**c**), *which* were grown at 20 °C (**d**), or 25 °C (**e**). Central horizontal lines represent the median, the bottom and top of the box represent the 25th and 75th percentile, respectively. Whiskers represent the 5th and 95th percentile, dots represent the data points used to calculate the box plot. n is indicated on the x-axis label. In (**d**), asterisks indicate the p-values of 9.6e-03 and 2.5e-06, as assessed by two-sided, unpaired Mann–Whitney and Wilcoxon tests comparing *tebp-1* worms with the cross siblings of the other genotypes. In (**e**), asterisk indicates *p*-value = 4.1e-07, computed as in (**d**). **f** Schematic of a repetition of the double mutant cross as in (**c**) with *pgl-1::mTagRfp-T* in the background. Worms heterozygous for one of the *tebp* mutations were singled and their germline categorized at day 2–3 of adulthood, according to germline morphology and assessed by PGL-1::mTagRFP-T expression. Worms were genotyped after categorization and imaging. **g** Representative widefield differential interference contrast (DIC) and fluorescence pictures of the three germline morphology categories defined. Scale bars, 200 μm. Atrophied germlines in categories 2 and 3 are marked with a white arrowhead. **h** Barplot representing the quantification of each category, per genotype as indicated on the x-axis. Number of animals analyzed is shown in the x-axis labels. **i** Plot showing the fraction of fertile populations of each indicated genotype across successive generations grown at 25 °C. *n* = 15 populations per strain.

category 2 animals displayed one atrophied gonad arm (Fig. 4g, middle panels), and category 3 animals had both gonad arms atrophied (Fig. 4g, lower panels). Besides Fig. 4g, representative animals for categories 2 and 3 are shown in Supplementary Fig. 5b. More than 85% of *tebp-1; tebp-2; pgl-1::mTagRfp-T* worms had a category 3 germline, while the remainder had only one gonad arm atrophied (Fig. 4h). Atrophied gonads generally showed under-proliferation of the germ cell nuclei of the mitotic zone and rare entry into meiosis, suggesting severe defects in cell division (Fig. 4g and Supplementary Fig. 5b). In addition, almost 15% (17/114 animals) of the progeny of *tebp-1; tebp-2; pgl-1::mTagRfp-T* synthetic sterility escapers were males, indicative of a high incidence of males (Him) phenotype. The synthetic sterility escaper progenies of previous crosses were also Him, at least in some cases (see F3 escaper progeny in Fig. 4b). Lastly, approximately 8% (8/97) of hermaphrodite *tebp-1; tebp-2; pgl-1::mTagRfp-T* escaper progeny had growth defects: while some reached adulthood but remained smaller than wild-type, others arrested prior to adulthood (Supplementary Fig. 5c).

Overall, these data show that the lack of functional TEBP-1 and TEBP-2 leads to severe germline defects that impede germline development.

TEBP-2 is required for transgenerational fertility. Despite the synthetic sterility of the double mutants, *tebp-1* and *tebp-2* single mutants did not have a baseline reduction in fertility when grown at 20 and 25 °C (Supplementary Fig. 5d, e). Nevertheless, mutants of telomere regulators, like *trt-1* and *mrt-1*, exhibit a Mrt phenotype, characterized by progressive loss of fertility across many generations[32,38]. We thus conducted a Mortal Germline assay at 25 °C using late generation mutants, and found that *tebp-1* and *tebp-2* mutants displayed opposing phenotypes in line with their differing effects on telomere length. While *tebp-1(xf133)* remained fertile across generations, like wild-type, *tebp-2(xf131)* showed a Mrt phenotype (Fig. 4i), the onset of which is delayed compared to *mrt-1(tm1354)* and *trt-1(ok410)*, indicating a slower deterioration of germline health over generations. These results show that TEBP-2 is required to maintain germline homeostasis transgenerationally, while TEBP-1 is not.

TEBP-1 and TEBP-2 are part of a telomeric complex in *C. elegans*. Our initial mass spectrometry approach allowed us to identify proteins associated with the telomeres of *C. elegans*. However, it remains unknown if these factors interact and whether they are part of a telomere-binding complex. To address this,

we performed size-exclusion chromatography with embryonic extracts from a strain expressing TEBP-1::3xFLAG; TEBP-2::GFP. Western blot analysis of the eluted fractions shows that TEBP-1 and TEBP-2 have very similar elution patterns with one peak ranging from 450 kDa to 1.5 MDa, with a maximum at 1.1 MDa (Fig. 5a and Supplementary Fig. 6a). Next, we reasoned that the elution peak would shift if telomeric DNA is enzymatically degraded. To test this, embryonic extracts were treated with *Serratia marcescens* nuclease (Sm nuclease), a non-sequence-specific nuclease, prior to size-exclusion chromatography, but we did not observe a strong shift (Fig. 5b). While we cannot fully exclude the possibility that telomeric DNA was inaccessible to Sm nuclease digestion, the results suggest that TEBP-1 and TEBP-2 are part of a telomeric complex.

To identify proteins interacting with TEBP-1 and TEBP-2, we performed immunoprecipitation (IP) followed by quantitative mass spectrometry (qMS) in embryos (Fig. 5c, d) and YAs (Supplementary Fig. 6b, c). Notably, IP-qMS of TEBP-1 and TEBP-2 baits enriched for MRT-1, POT-1, and POT-2, the three known ss telomere-binding proteins in *C. elegans*. In some cases, (Fig. 5d and Supplementary Fig. 6b) it was difficult to unambiguously assign unique peptides to TEBP-1::3xFLAG and TEBP-2::GFP in our qMS analysis, given their high protein sequence identity (65.4%). However, we confirmed by co-IP experiments that TEBP-1 and TEBP-2 reciprocally interact in embryos and YA (Fig. 5e, f and Supplementary Fig. 6d). Moreover, TEBP-1 and TEBP-2 remain associated with MRT-1, POT-1, and POT-2 even after treatment with Sm nuclease (Supplementary Fig. 6e, f).

POT-1 is required to bridge the double-stranded and the single-stranded telomere. To reveal the architecture of the telomeric complex, we sought to identify direct interactions amongst TEBP-1, TEBP-2, POT-1, POT-2, and MRT-1, using a yeast two-hybrid (Y2H) screen. While TEBP-2 fused to the DNA-binding domain of Gal4 unfortunately self-activated the reporter (Supplementary Fig. 6g), we could identify direct interactions of POT-1 with TEBP-1 and TEBP-2 (Fig. 6a and Supplementary Fig. 6g). Furthermore, in accordance with IP-qMS and co-IP experiments (Fig. 5e, f and Supplementary Fig. 6d), we confirmed interaction between TEBP-1 and TEBP-2 in the Y2H experiment (Fig. 6a and Supplementary Fig. 6g). These results are consistent with a scenario where TEBP-1 and TEBP-2 interact directly with each other and with POT-1.

The observed direct interactions suggest that POT-1 may be a critical link between the ds and the ss telomeric region. To test this idea, we performed IP-qMS of TEBP-1 and TEBP-2, in
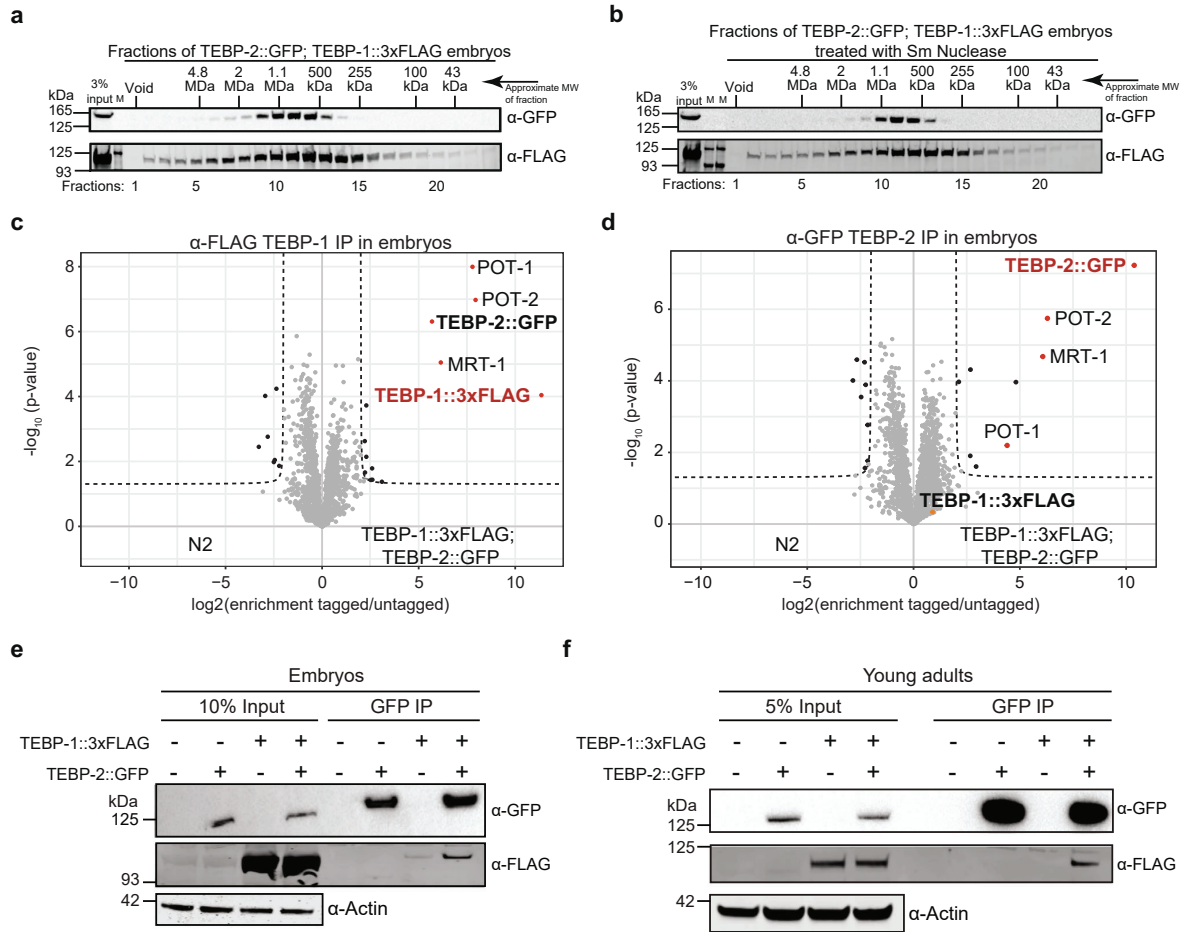
Fig. 5 TEBP-1 and TEBP-2 are part of a telomeric protein complex. a Size-exclusion chromatography of embryo extracts expressing TEBP-1::3xFLAG and TEBP-2::GFP, followed by western blot of the eluted fractions. The approximate molecular weight (MW) of the fractions is indicated on the Fig. panel. $N = 2$ biologically independent experiments with similar results. **b** Identical to (**a**), but with treatment of embryo extracts with Sm nuclease, prior to size-exclusion chromatography. $N = 1$. **c, d** Volcano plots showing quantitative proteomic analysis of either TEBP-1::3xFLAG (**c**) or TEBP-2::GFP (**d**) IPs in embryos. IPs were performed in quadruplicates. Enriched proteins (threshold: 4-fold, p-value < 0.05) are shown as black dots, enriched proteins of interest are highlighted with red or orange dots, and the baits are named in red. Background proteins are depicted as gray dots. **e** Co-IP western blot experiment of TEBP-1::3xFLAG and TEBP-2::GFP. The IP was performed with a GFP-trap, on embryo extracts from strains carrying either one or both of the endogenous tags and wild-type. Actin was used as loading control. **f** Same co-IP experiment as in (**e**) but carried out with extracts from young adult worms. For (**e**) and (**f**) $N = 3$ biologically independent experiments with similar results.

wild-type and mutant *pot-1* backgrounds. These experiments showed that interaction of the ds telomere binders TEBP-1 and TEBP-2 with the ss binders POT-2 and MRT-1, is strongly depleted in *pot-1* mutants (Fig. 6b, c). TEBP-1 and TEBP-2 protein levels are not affected by the *pot-1* mutation, indicating the loss of interaction with POT-2 and MRT-1 is not due to reduced availability of TEBP-1 or TEBP-2 (Supplementary Fig. 6h). In addition, TEBP-1 and TEBP-2 still interact with each other in the absence of POT-1 (Supplementary Fig. 6h).

Next, to map the amino acid sequences responsible for TEBP-1 and TEBP-2 DNA-binding and protein-protein interactions, with each other and with POT-1, we divided their protein sequences into seven fragments (f1–f7), and the protein sequence of POT-1 into three fragments (f1–f3, Fig. 6d). DNA pulldowns with His-MBP-tagged TEBP-1 and TEBP-2 recombinant proteins demonstrated DNA binding by their f3 fragments (Fig. 6d, e), which contain their third predicted homeo-/myb-domain. Furthermore, Y2H experiments using the fragments shown in Fig. 6d, indicate

that the C-terminal tails of TEBP-1 and TEBP-2 (f7) interact with the OB-fold of POT-1 (Fig. 6f, g). Additional Y2H assays demonstrate that TEBP-1 and TEBP-2 interact with each other via their respective f1 fragments, encompassing their first predicted homeo-/myb-domains (Fig. 6h and Supplementary Fig. 6i).

Altogether, our data strongly indicate that TEBP-1 and TEBP-2 are integral parts of a telomeric complex, or complexes, which also include the known ss telomere binders POT-1, POT-2, and MRT-1. We propose a simple working model where TEBP-1 and TEBP-2 bind to the ds telomere via their third predicted homeo-/myb-domains, have opposed effects on telomere dynamics, and are required for fertility (Fig. 6i). POT-1, with the ability of its OB-fold to directly bind the C-terminal tails of TEBP-1 and TEBP-2 (Fig. 6a, f, g), as well as ss telomeric repeats in vitro[31], may link the ds binders to the ss telomere, thereby bringing TEBP-1 and TEBP-2 in close proximity of POT-2 and MRT-1 (Fig. 6i).
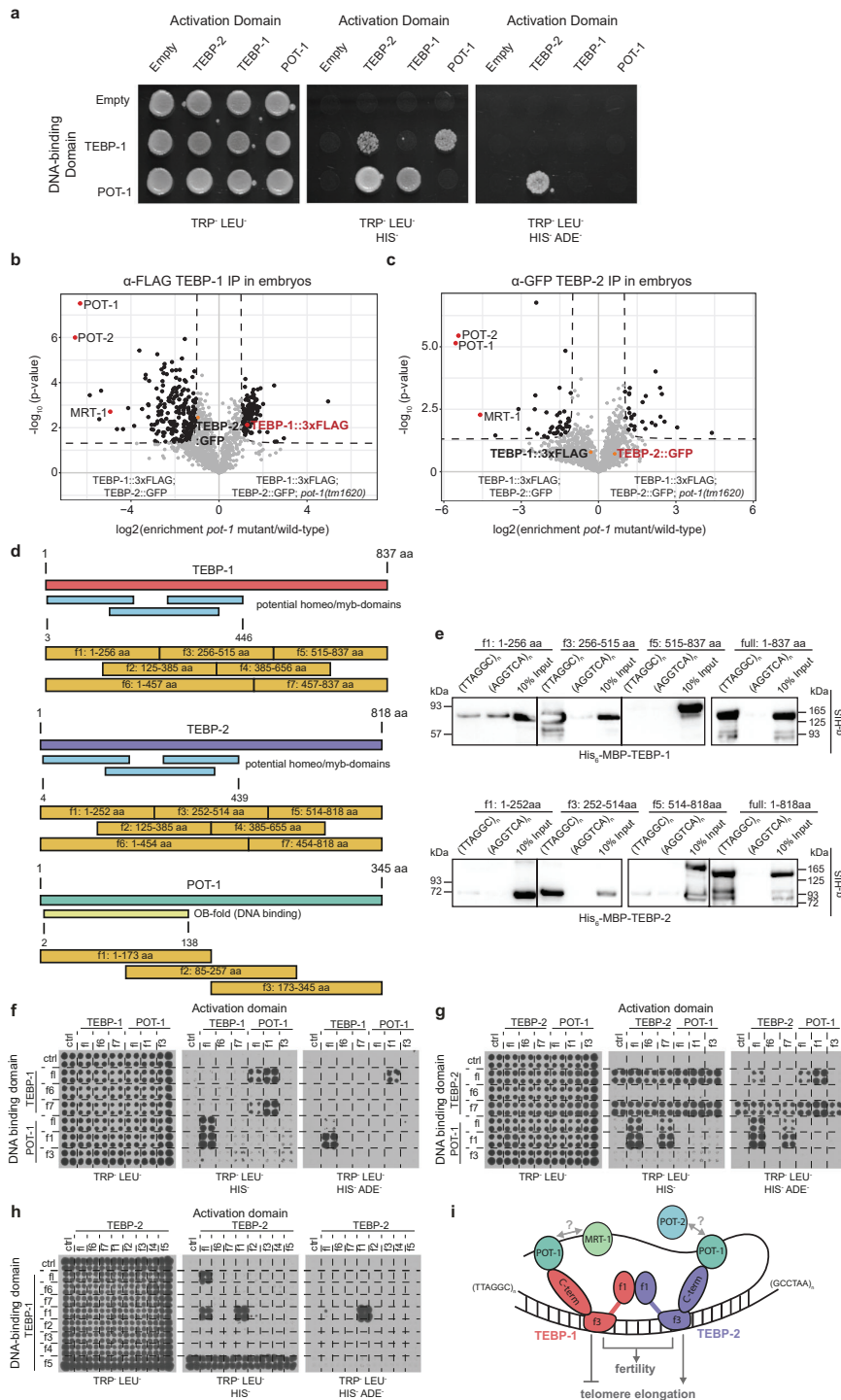
Conservation of *tebp* genes in the *Caenorhabditis* genus. To infer the evolutionary history of *tebp-1* and *tebp-2* genes, we identified protein-coding orthologs by reciprocal BLASTP analysis in the searchable genomes in Wormbase and Wormbase ParaSite databases. Then, we performed a multiple sequence alignment with the ortholog protein sequences, and used it to build a phylogenetic tree (Fig. 7a and Supplementary Data file 2).

Our findings suggest that *tebp* orthologs are present only in the *Caenorhabditis* genus, mostly in the *Elegans* supergroup (which includes the *Elegans* and *Japonica* groups). A distinct number of protein-coding *tebp* genes was identified per species: *C. briggsae*, *C. nigoni*, *C. sinica*, and *C. japonica* have one *tebp* ortholog; *C. elegans*, *C. inopinata*, *C. remanei*, *C. brenneri*, *C. tropicalis*, and *C. angaria* have two *tebp* orthologs; and *C. latens* has three *tebp*

Fig. 6 POT-1 links the ds telomere binders to the ss telomere. a Y2H assay with full length TEBP-1, TEBP-2, and POT-1 fusions to the activation or DNA-binding domains of Gal4. Growth on TRP⁻ LEU⁻ HIS⁻ plates demonstrates interaction. Growth on high stringency TRP⁻ LEU⁻ HIS⁻ ADE⁻ medium suggests strong interaction. TRP:⁻ lacking tryptophan, LEU:⁻ lacking leucine, HIS:⁻ lacking histidine, ADE:⁻ lacking adenine. **b, c** Volcano plots showing quantitative proteomic analysis of either TEBP-1::3xFLAG (**b**) or TEBP-2::GFP (**c**) IPs in embryos. IPs were performed in quadruplicates. Enriched proteins (threshold: 2-fold, *p*-value < 0.05) are shown as black dots, enriched proteins of interest are highlighted with red or orange dots, and annotated. Background proteins are depicted as gray dots and the respective bait protein annotated in red. **d** Scheme for the cloning of different fragments of TEBP-1, TEBP-2 and POT-1 for IP experiments and Y2H. TEBP-1 and TEBP-2 were divided into five fragments (f1–f5) of approx. 30 kDa, as well as two additional fragments covering the N-terminus including the predicted DNA-binding domains (f6) and the C-terminus (f7). POT-1 was divided into three fragments of around 15 kDa (f1–f3). **e** DNA pulldowns as in Fig. 1c with recombinantly expressed and N-terminally His-MBP-tagged fragments f1, f3, and f5 of TEBP-1 and TEBP-2, as well as the full length proteins with the same tags. The western blot was probed with α-His antibody and the signals detected by chemiluminescence. f1–f5: fragments of respective protein, full: full length respective protein, kDa: kilodalton, MBP: maltose-binding protein. $N = 2$ independent experiments with similar results. **f** Y2H assay like in (**a**) but with TEBP-1 and POT-1 full length proteins (fl), as well as N- and C-terminal fragments (f6 and f7 for TEBP-1, or f1 and f3 for POT-1, respectively) fused to the activation or DNA-binding domains of Gal4. Growth determined on the same medium as in **a**. **g** Y2H assay as in (**f**) but with TEBP-2 and POT-1 constructs. **h** Y2H assay as in (**f**) but with all fragments of TEBP-1 including the full length protein fused to the Gal4 DNA-binding domains, as well as all fragments of TEBP-2 including the full length protein fused to the Gal4 activation domain. f1–f7: fragments of respective protein, crtl: control/empty plasmid, fl: full length protein. **i** Proposed working model for the interactions between telomere-binding proteins and telomere repeats in *C. elegans*. TEBP-1 and TEBP-2 fragments 3 (f3), containing a predicted DNA-binding domain, bind to ds telomere repeats and have opposing effects on telomere elongation. Both proteins interact with each other via their N-terminal fragments (f1). TEBP-1, TEBP-2 and POT-1 interact directly via the C-terminal fragment (f7) of TEBP-1/TEBP-2 and the N-terminal fragment (f1) of POT-1. As a result of this interaction, the ss telomere comes in closer contact to the ds telomere. Our current data does not support direct interactions between POT-1, POT-2, and MRT-1, but these factors may interact in the presence of telomeric DNA.

orthologs. The multiple sequence alignment showed the N-terminal region of *tebp* genes, the region with similarity to the homeodomains of human and yeast RAP1 (Supplementary Fig. 1d, e and Supplementary Data file 1), is more similar between orthologs than the C-terminal region (Supplementary Data File 2). However, phylogenetic analysis with only the N-terminal region did not produce major differences on tree topology (Supplementary Fig. 7). In order to derive evolutionary relationships between different *tebp* genes, we evaluated local synteny information. We found a high degree of regional synteny conservation between *C. elegans tebp-1* and one of the *tebp* copies in *C. inopinata, C. remanei, C. briggsae, C. nigoni, C. sinica, C. tropicalis*, and *C. japonica* (Table 1 and Supplementary Data file 2). Conversely, *tebp-2* did not show any signs of regional synteny across *Caenorhabditis* species (Supplementary Data file 2), suggesting that the gene duplication event creating *tebp-2* occurred after divergence from the *C. inopinata* lineage, less than 10.5 million years ago[53]. Neither of the two *tebp* orthologs of *C. brenneri, C. latens, and C. angaria* are in synteny with *C. elegans tebp-1* (Supplementary Data file 2).

To determine whether TEBP proteins are generally telomere-binders in the *Elegans* supergroup, we performed DNA pulldowns, using nuclear extracts prepared from synchronized *C. briggsae* gravid adults. CBG11106, the only *C. briggsae* ortholog of *tebp-1* and *tebp-2*, was significantly enriched in the telomere pulldown (Fig. 7b), demonstrating that it can bind to the TTAGGC telomeric repeat. Of note, CBG22248, one of the two *C. briggsae* orthologs of MRT-1, was also enriched in the telomere pulldown, and CBG16601, the ortholog of POT-1, was just below our significance threshold, suggesting functional similarities to their *C. elegans* orthologs.

Discussion

Telomeres and their associated proteins are important to ensure proper cell division. In the popular model nematode *C. elegans*, only ss telomere-binding proteins were known thus far[31,38]. Here, we describe a telomeric complex with the paralogs TEBP-1 and TEBP-2 as direct ds telomere-binding proteins. POT-1 seems to bridge the ds telomere-binding module of the complex, comprised of TEBP-1 and TEBP-2, with the ss telomere region. Strikingly, despite the high level of sequence similarity between TEBP-1 and TEBP-2, their mutant phenotypes are divergent.

Robust identification of telomere-associated proteins in *C. elegans*. Three lines of evidence demonstrate the validity and robustness of our screen. First, attesting for its technical reproducibility, the two qMS detection strategies employed shared an overlapping set of proteins enriched in telomeric sequence pull-downs (8 overlapping factors out of 12 and 8 hits). Second, within our overlapping set of enriched factors, we detected the previously identified ss telomere-binding proteins POT-1, POT-2, and MRT-1[31,33,37,38]. Lastly, the *C. elegans* KU heterodimer homologs CKU-70 and CKU-80 were enriched in the screens. In other organisms, such as *Saccharomyces cerevisiae, Trypanosoma brucei, Drosophila melanogaster*, and *Homo sapiens*, KU proteins have been shown to associate with telomeres, regulating their length and protecting them from degradation and recombination[54,55]. The *C. elegans* homologs were shown to interact with telomeres, but do not seem to have telomere regulatory functions[43]. However, CKU-70 and CKU-80 were not enriched in the TEBP-1 and TEBP-2 interactome experiments, suggesting that their binding to telomeric DNA occurs independently of the TEBP-1/TEBP-2 complex (Fig. 5 and Supplementary Fig. 6). Alternatively, these factors may be part of the telomeric complex, with no direct interaction with TEBP-1 or TEBP-2.

We identified POT-3 in the background of our LFQ screen (Supplementary Data File 3), supporting the lack of telomeric phenotypes of *pot-3* mutants[31]. Furthermore, a number of factors previously reported to have telomere DNA-binding capability or to regulate telomere length, were not detected or lacked significant enrichment in our quantitative proteomics screen. MRT-2 is a homolog of *S. cerevisiae* checkpoint gene RAD17 and human RAD1, previously reported to regulate telomere length[30]. Much like *tebp-2* and *mrt-1, mrt-2* mutants have shorter telomeres than wild-type and a Mrt phenotype. It is plausible that MRT-2 regulates telomere length beyond the context of direct telomeric binding. PLP-1[56], HMG-5[57], and CEH-37[58], were previously shown to bind to the *C. elegans* telomeric sequence in vitro. PLP-1 was enriched in the (AGGTCA)ₙ scrambled control in our qMS screen (Supplementary Data file 3), suggesting that PLP-1 is a general ds DNA binder, and not a specific telomere binder. Furthermore, HMG-5 was detected in the background, and CEH-37 was not detected altogether in our screen (Supplementary Data file 3). Further studies should clarify if and how these factors interact with the telomere complex described in this work.
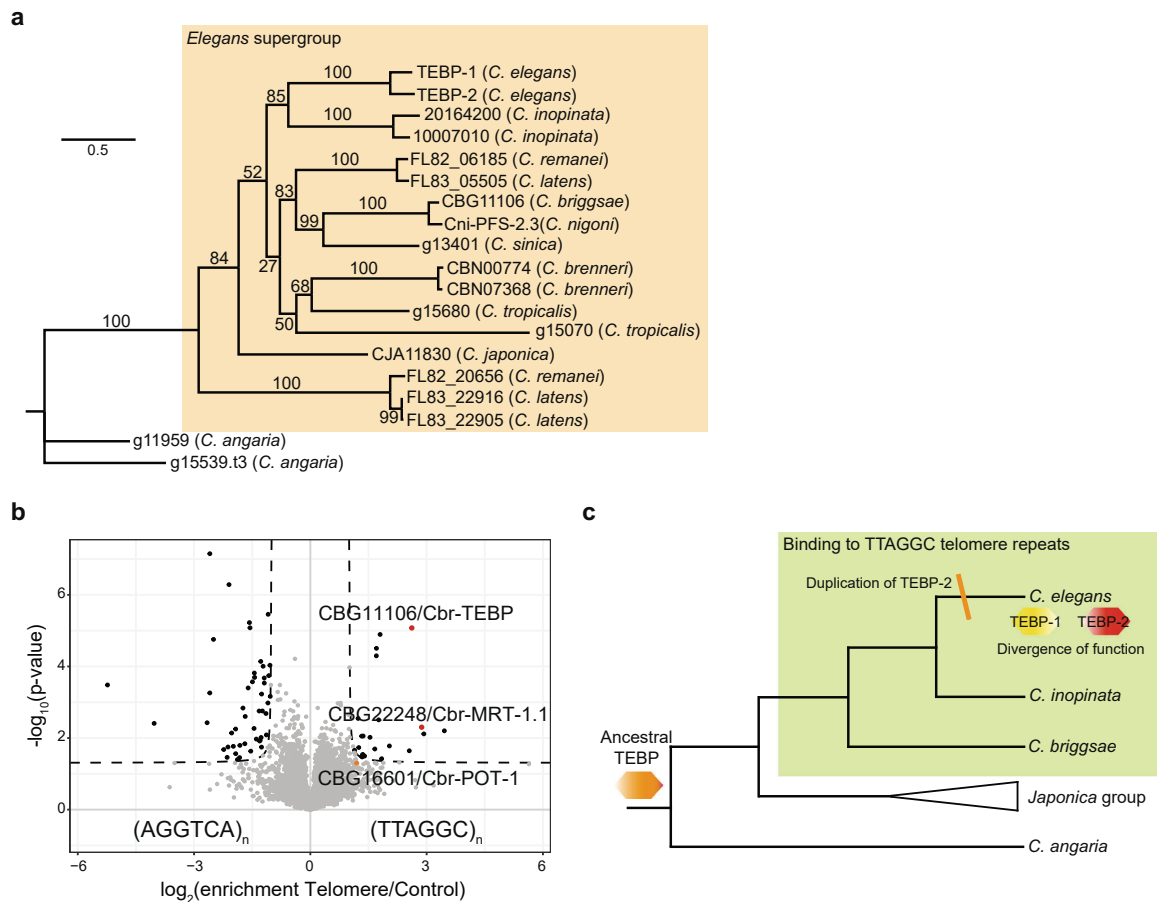
Fig. 7 Conservation of *tebp* genes in the *Caenorhabditis* genus. a Phylogenetic tree constructed with IQ-TREE (v1.6.12), using a MAFFT (v7.452) multiple sequence alignment of the protein sequences of TEBP orthologs (see Supplementary Data file 2, sheet 2). Values on the nodes represent bootstrapping values for 10,000 replicates, set to 100. The TEBP orthologs outside the orange background represent the outgroup of the analysis. **b** Volcano plot of telomere DNA pulldown, as in Fig. 1a, of gravid adult nuclear extracts from *C. briggsae*. Here, pulldowns were performed in quadruplicates, per condition. Enriched proteins (enrichment threshold > 2-fold, *p*-value < 0.05) are labeled as black dots, whereas enriched proteins of interest are labeled with red or orange dots. Proteins below the threshold are depicted as gray dots. Homologs of telomere binders are named. **c** Depiction of the evolution of *tebp* genes in *Caenorhabditis*. We speculate that this family originated from an ancestor TEBP (orange hexagon), presumably required for fertility and capable of binding to telomeres. As we have confirmed telomere binding in *C. elegans* and *C. briggsae* (species in bold indicate confirmed binding of TEBP proteins to telomeric DNA), it is plausible that their common ancestor was able to bind to telomeres. The gene duplication that generated *tebp-2* occurred after the divergence of *C. elegans* and *C. inopinata* (marked as orange stripe), followed by division, or diversification, of functions of these two paralogs (TEBP-1: yellow hexagon, TEBP-2: red hexagon).

The fast-evolving paralogs TEBP-1 and TEBP-2 are required for fertility. TEBP-1 and TEBP-2 share 65.4% of their amino acid sequence, which most likely reflects a common origin by gene duplication. Interestingly, the two paralogs TEBP-1 and TEBP-2 interact with each other, and with the same set of factors, i.e., POT-1, POT-2, and MRT-1 (Fig. 5 and Supplementary Fig. 6). This is striking, considering the divergent phenotypes of *tebp-1* and *tebp-2* mutants: *tebp-1* mutants have longer telomeres than wild-type, while *tebp-2* animals have shorter telomeres than wild-type and a Mortal Germline. Moreover, while the fertility of *tebp-1* and *tebp-2* animals is not compromised, *tebp-1*; *tebp-2* double mutants show highly penetrant synthetic sterility irrespective of the temperature the animals are grown at, indicating that TEBP-1 and TEBP-2 contribute to normal fertility (Fig. 4 and Supplementary Fig. 5). The observed synthetic sterility is likely justified by failure to enter and progress through normal mitosis and meiosis, as judged by the under-proliferation of germ cells.

The synthetic sterility of *tebp-1*; *tebp-2* animals is specific to these two paralogs, as other genetic crosses of shorter versus longer telomere mutants did not result in sterile double mutants. The synergistic role of TEBP-1 and TEBP-2 in fertility provide a puzzling contrast with their opposed telomere length mutant phenotypes. We speculate that the requirement of TEBP-1 and TEBP-2 to fertility may be independent of their functions at telomeres. Future studies on the influence of TEBP-1 and TEBP-2 on germline and embryonic gene expression may shed light on this aspect.

CBG11106, the single homolog of TEBP-1 and TEBP-2 in *C. briggsae*, interacts with telomeric DNA (Fig. 7b), suggesting that TEBP nematode homologs bind to telomeric DNA at least since the divergence of *C. elegans* and *C. briggsae*, from a common ancestor that presumably lived 80–100 million years ago[59]. To verify this, the capability of additional TEBP orthologs to bind to telomeric DNA needs to be experimentally addressed. We

Table 1 Synteny analysis of *tebp* orthologs in other *Caenorhabditis* species.		
***tebp* ortholog**	**Synteny with *tebp-1***	**Synteny with *tebp-2***
10007010 (*C. inopinata*)	−	−
20164200 (*C. inopinata*)	+	−
FL82_06185 (*C. remanei*)	+	−
FL83_05505 (*C. latens*)	−	−
CBG11106 (*C. briggsae*)	+	−
Cni-PFS-2.3 (*C. nigoni*)	+	−
g13401 (*C. sinica*)	+	−
CBN00774 (*C. brenneri*)	−	−
CBN07368 (*C. brenneri*)	−	−
g15680 (*C. tropicalis*)	+	−
g15070 (*C. tropicalis*)	−	−
CJA11830 (*C. japonica*)	+	−
FL83_22916 (*C. latens*)	−	−
FL83_22905 (*C. latens*)	−	−
FL82_20656 (*C. remanei*)	−	−
g15539.t3 (*C. angaria*)	−	−
g11959 (*C. angaria*)	−	−

Overview of synteny of the tebp orthologs of other *Caenorhabditis* species with tebp-1 or tebp-2 of *C. elegans*. A "+" indicates regional synteny, while a "−" is lack of synteny.

speculate that *tebp-1* and *tebp-2* originated from an ancestor *Caenorhabditis tebp* gene required for fertility and with the ability to bind ds telomeric repeats (Fig. 7c). The *tebp-1* ancestor was duplicated after the divergence of *C. inopinata* and *C. elegans*, 10.5 million years ago[53], likely initiating a process of functional diversification of *tebp-1* and *tebp-2*.

Given their possible recent divergence, in evolutionary terms the 65.4 % protein sequence similarity observed between the protein sequences of TEBP-1 and TEBP-2 is actually fairly low. This likely reflects fast evolution of TEBP-1 and TEBP-2, in line with the known fast evolution as suggested for other telomere-binding proteins[60]. While it is tempting to establish evolutionary relationships with vertebrate TRF1 and TRF2 proteins, TEBP-1/TEBP-2 and TRF1/TRF2 are not homologs. In addition, TRF1 and TRF2 are binding to telomeric DNA via C-terminal myb-domains[61], while DNA binding in TEBP-1 and TEBP-2 occurs N-terminally. However, on the functional level, similarity between *C. elegans* TEBP-1/TEBP-2 and vertebrate TRF1/TRF2, potentially reflecting convergent evolution between two phylogenetically independent sets of telomere-binding paralogs is possible, but needs further investigation.

A telomere complex in actively dividing tissues in homeostasis. Our size-exclusion chromatography, quantitative proteomics, and Y2H data support the existence of a telomere complex comprising TEBP-1, TEBP-2, POT-1, POT-2, and MRT-1 (Fig. 5 and Supplementary Fig. 6). According to our size-exclusion chromatography data, this complex elutes in a range between 600 kDa and 1.1 MDa. It should be noted that our model does not make any assumptions regarding complex stoichiometry. At the moment, we cannot exclude the existence of remaining DNA fragments in the complex, despite nuclease treatment, which could add to the total molecular weight. Thus, we propose a working model, whereby TEBP-1 and TEBP-2 bind to ds telomere repeats via their third predicted homeo-/myb-domains, and directly interact with the OB-fold of POT-1 with their C-terminal tails. Binding to POT-1 may, in turn, bring the ss telomeric repeats, and thus POT-2 and MRT-1, into closer contact (Fig. 6i). In the absence of POT-1, TEBP-1 and TEBP-2 are not able to interact with POT-2 and MRT-1 (Fig. 6b, c). We speculate that reciprocal regulation

by TEBP-2 and POT-1/TEBP-1 define normal telomere length. In this scenario, TEBP-2 might counteract telomere shortening by POT-1 and TEBP-1 (Fig. 6i). The precise interplay between these telomeric factors, namely the interactions between POT-1, POT-2, and MRT-1, and the mechanism of telomere elongation have to be further elucidated.

The mammalian shelterin complex counteracts recognition of telomeres as DNA double-strand breaks by inhibiting the DNA damage machinery. When shelterin factors are abrogated, catastrophic end-to-end chromosome fusions are observed[62,63]. Previous studies did not identify end-to-end chromosome fusions in *pot-1* and *pot-2* mutants[31,33,37]. It remains to be determined if *tebp-1* and/or *tebp-2* mutations lead to telomere fusions and whether the *C. elegans* telomeric complex is required to protect telomeres from DNA damage. It is possible that the synthetic sterility and high frequency of males observed in *tebp-1; tebp-2* double mutants, as well as the Mortal Germline phenotype of *tebp-2* and *mrt-1*, may be downstream of germline genome instability.

A germline-specific MAJIN/TERB1/TERB2 telomere-binding complex has been described in mouse testes[64–66]. Knock-outs of these factors lead to meiotic arrest and male sterility[64–66], similar to the observed phenotype in *tebp-1; tebp-2* double mutants. This mammalian protein complex tethers telomeres to the nuclear envelope, a process essential for meiotic progression. A previous study has shown that POT-1 is required in *C. elegans* to tether telomeres to the nuclear envelope during embryogenesis[67]. Given the interaction of TEBP-1 and TEBP-2 with POT-1 in vitro and in vivo, the telomeric complex may be dynamically involved in this process.

The distinct compartmentalization of post-mitotic soma versus actively dividing germline, together with a plethora of genetic tools, make *C. elegans* an enticing model organism for telomere biology in vivo, in homeostatic conditions. The identification of a telomeric complex in *C. elegans* allows further investigation of telomere regulation in this popular model organism.

Methods

***C. elegans* nuclear-enriched protein extract preparation**. Nuclear extract preparation of gravid adult worms was done as described[68]. The worms were synchronized by bleaching and harvested at the gravid adult stage by washing them off the plate with M9 buffer. After washing the worms in M9 buffer for 4 times, they were pelleted by centrifugation at 600 x *g* for 4 min, M9 buffer was removed and extraction buffer (40 mM NaCl, 20 mM MOPS pH 7.5, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 10% Glycerol, 2 mM DTT, and 1x complete protease inhibitors Roche) was added. Worms resuspended in extraction buffer were frozen in liquid nitrogen. The resulting pellets were ground to a fine powder in a pre-cooled mortar and transferred to a pre-cooled glass douncer. When thawed, the samples were sheared with 30 strokes, piston B. The worm suspension was pipetted to pre-cooled 1.5 ml reaction tubes (1 ml per tube) and cell debris, as well as unsheared worms were pelleted by centrifugation at 200 x *g* for 5 min at 4 °C for two times. To separate the cytoplasmatic and nuclear fractions, the supernatant was spun at 2000 x *g* for 5 min at 4 °C. The resulting pellet containing the nuclear fraction was washed twice by resuspension in extraction buffer and subsequent centrifugation at 2000 x *g* for 5 min at 4 °C. After the washing steps, the nuclear pellet was resuspended in 200 μl buffer C + (420 mM NaCl, 20 mM Hepes/KOH pH 7.9, 2 mM MgCl₂, 0.2 mM EDTA, 20% Glycerol, 0.1% Igepal CA 630, 0.5 mM DTT, 1x complete protease inhibitors). Nuclear extract of gravid adult worms of *C. briggsae* was prepared as described above.

Oligonucleotides. All oligonucleotides used throughout this manuscript (cloning, sequencing, DNA pulldowns, fluorescence polarization etc.) are listed in Supplementary Data file 4 with their name and sequence.

DNA pulldowns

Preparation of biotinylated DNA for pulldown experiments. Biotinylated telomeric and control DNA for the DNA pulldown for detection of telomeric interactors was prepared as previously published[16,39,40]. In short, 25 μl of 10-mer repeat oligonucleotides of either telomeric or control sequence were mixed 1:1 with 25 μl of their respective reverse complement oligonucleotide and 10 μl annealing buffer (200 mM Tris-HCl, pH 8.0, 100 mM MgCl2, 1 M KCl). The mixture was brought to

100 µl final volume with H$_2$O, heated at 80 °C for 5 min, and left to cool. Once at room temperature (RT), the samples were supplemented with 55 µl H$_2$O, 20 µl 10x T4 DNA ligase buffer (Thermo Scientific), 10 µl PEG 6000, 10 µl 100 mM ATP, 2 µl 1 M DTT and 5 µl T4 Polynucleotide Kinase (NEB, 10 U/µl, #M0201) and left at 37 °C for 2 h to concatenate. Finally, 4 µl of T4 DNA Ligase (Thermo Scientific, 5 WU/µl, #EL0011) were added and the samples incubated at RT overnight for ligation and polymerization. The ligation process was monitored by running 1 µl of the reaction on a 1% agarose gel. The samples were cleaned by phenol-chloroform extraction. For this, 1 vol. of H$_2$O and 200 µl of Phenol/Chloroform/Isoamyl Alcohol (25:24:1; pH 8; Invitrogen, # 15593049) was added to the mixture, vortexed and centrifuged at 16,000 x g for 2 min. After centrifugation the aqueous phase was transferred to a fresh tube and the DNA precipitated by addition of 1 ml 100% Ethanol and incubation at −20 °C for 30 min. Afterwards the suspension was centrifuged at 16,000 x g for 45 min at 4 °C. The resulting DNA pellet was resuspended in 74 µl H$_2$O and 10 µl 10x Klenow-fragment reaction buffer (Thermo Scientific), 10 µl 0.4 mM Biotin-7-dATP (Jena Bioscience, #NU-835-BIO) and 6 µl Klenow-Fragment exo- polymerase (Thermo Scientific, 5 U/µl, # EP0422) added. Biotinylation was carried out by incubation at 37 °C over night. The reaction was cleaned up by size-exclusion chromatography using MicroSpin Sephadex G-50 columns (GE Healthcare, #GE27-5330-01).

Pulldown experiments. Biotinylated DNA and Dynabeads™ MyOne™ Streptavidin C1 (Thermo Scientific, #65001) were mixed with PBB buffer (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5% NP 40, 5 mM MgCl$_2$, 1 mM DTT) and incubated at room temperature for 15 min on a rotating wheel to immobilize the DNA on the beads. After three washes with PBB buffer, the DNA coupled beads were resuspended in PBB buffer and Salmon sperm (10 mg/ml, Ambion, #AM9680) was added 1:1000 as competitor for unspecific DNA binding. The pulldowns were performed with different amounts of protein extract (see below) and incubated at 4 °C on a rotating wheel for 90 min. Following incubation the beads were washed three times with PBB buffer and resuspended in 1x Loading buffer (4x NuPAGE LDS sample buffer, Thermo Scientific, #NP0008) supplemented with 100 mM DTT. For elution, the samples were boiled at 70 °C for 10 min and afterwards loaded on a gel and processed as indicated above for MS, or below for western blot. In pulldown-MS experiments, the pulldowns were prepared in either technical quadruplicates (LFQ), or technical duplicates (DML) per condition, whereas for western blot all conditions were prepared with one replicate and an input. In all, 200–400 µg of nuclear worm extract and of *Escherichia coli* extract were used for the mass spectrometry screen and pulldowns of Fig. 1c, respectively. In all, 0.4–0.7 mg of total protein extract were used for the pulldowns shown in Fig. 1d–f. Four-hundred micrograms of *E. coli* extract was used in DNA-binding domain pulldowns in Fig. 6e.

Mass spectrometry: sample preparation, data acquisition, and analysis

In-gel digest. In-gel digestion was performed as previously described[16,69] with the exception of the DML samples (see below). Samples were run on a 10% Bis-Tris gel (NuPAGE; Thermo Scientific, #NP0301) for 10 min (IP samples) or on a 4–12% Bis-Tris gel (NuPAGE, Thermo Scientific, #NP0321) for 20 min (LFQ-measured telomeric DNA pulldowns) at 180 V in 1x MOPS buffer (NuPAGE, Thermo Scientific, #NP0001). Individual lanes were excised and cut to approximately 1 mm × 1 mm pieces with a clean scalpel, and transferred to a 1.5 ml tube. For the LFQ telomeric DNA pulldowns, the lanes were split into four fractions. The gel pieces were destained in destaining buffer (50% 50 mM NH$_4$HCO$_3$ (ABC), 50% ethanol p.a.) at 37 °C under rigorous agitation. Next, gel pieces were dehydrated by incubation in 100% acetonitrile for 10 min at 25 °C shaking and ultimately dehydrated using a Concentrator Plus (Eppendorf, #5305000304, settings V-AQ). The gel pieces were incubated in reduction buffer (50 mM ABC, 10 mM DTT) at 56 °C for 60 min and subsequently incubated in alkylation buffer (50 mM ABC, 50 mM iodoacetamide) for 45 min at room temperature in the dark. Gel pieces were washed in digestion buffer (50 mM ABC) for 20 min at 25 °C. Next, gel pieces were dehydrated again by incubation in 100% acetonitrile and drying in the concentrator. The dried gel pieces were rehydrated in trypsin solution (50 mM ABC, 1 µg trypsin per sample, Sigma-Aldrich, #T6567) and incubated overnight at 37 °C. The supernatant was recovered and combined with additional fractions from treatment with extraction buffer (30% acetonitrile) twice and an additional step with pure acetonitrile for 15 min at 25 °C, shaking at 1400 rpm. The sample solution containing the tryptic peptides was reduced to 10% of the original volume in a Concentrator Plus, to remove the acetonitrile and purified using the stage tip protocol.

Dimethyl labeling. Dimethyl labeling (DML) was done as previously described[70]. For DML, in-gel digest was performed as indicated in the last section, with the exception of exchanging ABC buffer for 50 mM TEAB (Fluka, #17902) after alkylation. The volume of the extracted peptides was reduced in a Concentrator Plus. For labeling, either 4% formaldehyde solution (Sigma-Aldrich, #F8775) for light labeling or 4% formaldehyde-D2 (Sigma-Aldrich, #596388) solution for medium labeling, as well as 0.6 M NaBH$_3$CN (Sigma-Aldrich, #156159) were added to the samples and mixed briefly. The mixture was incubated for 1 h at 20 °C, shaking at 1000 rpm and afterwards quenched by addition of a 1% ammonia solution (Sigma-Aldrich, #30501) and acidified with 10% formic acid solution (Merck, #1.00264.1000). After the labeling reaction, the respective light and

medium samples were mixed 1:1 (light telomere: medium control; medium telomere: light control) and purified by stage tip purification.

Stage tip purification. Stage tip purification was performed as previously described[71]. Desalting tips were prepared by using two layers of Empore C18 material (3 M, #15334911) stacked in a 200 µl pipet tip. The tips were activated with pure methanol. After two consecutive washes with Buffer B (80% acetonitrile, 0.1% formic acid) and Buffer A (0.1% formic acid) for 5 min the tryptic peptide samples were applied and washed once more with Buffer A. Upon usage, peptides were eluted with Buffer B. The samples were centrifuged in a Concentrator Plus for 10 min to evaporate the acetonitrile and adjusted to 14 µl with Buffer A.

MS measurement and data analysis. For MS measurement 5 µl of sample were injected. The desalted and eluted peptides were loaded on an in-house packed C18 column (New Objective, 25 cm long, 75 µm inner diameter) for reverse-phase chromatography. The EASY-nLC 1000 system (Thermo Scientific) was mounted to a Q Exactive Plus mass spectrometer (Thermo Scientific) and peptides were eluted from the column in an optimized 2 h (pulldown) gradient from 2 to 40% of 80% MS grade acetonitrile/0.1% formic acid solution at a flow rate of 225 nL/min. The mass spectrometer was used in a data-dependent acquisition mode with one MS full scan and up to ten MS/MS scans using HCD fragmentation. All raw files were processed with MaxQuant (version 1.5.2.8) and searched against the *C. elegans* Wormbase protein database (Version WS269), as well as the Ensembl Bacteria *E. coli* REL606 database (version from September 2018) for proteins from the feeding strain OP50. Carbamidomethylation (Cys) was set as fixed modification, while oxidation (Met) and protein N-acetylation were considered as variable modifications. For enzyme specificity, trypsin was selected with a maximum of two miscleavages. LFQ quantification (without fast LFQ) using at least 2 LFQ ratio counts and the match between run option were activated in the MaxQuant software. Fractions and conditions were indicated according to each experiment. Data analysis was performed in R using existing libraries (ggplot2-v 3.2.1, ggrepel-v 0.8.1, stats-v 3.5.2) and in-house scripts. Protein groups reported by MaxQuant were filtered removing known contaminants, protein groups only identified by site and those marked as reverse hits. Missing values were imputed at the lower end of LFQ values using random values from a beta distribution fitted at 0.2–2.5%. For statistical analysis, p-values were calculated using Welch's t-test. Enrichment values in the volcano plots represent the mean difference of log$_2$ transformed and imputed LFQ intensities between the telomere and the control enriched proteins. Peptide labels created by the dimethyl-labeling reaction were selected in the MaxQuant software as "N-terminal Dimethyl 0" and "Dimethyl 0" for the light samples, as well as "N-terminal Dimethyl 4" and "Dimethyl 4" for the heavy labeled samples. The re-quant option was activated. An incorporation check was run additionally to confirm incorporation of the dimethyl labels of at least 95% in each sample. Protein groups resulting from MaxQuant analysis were filtered identically to LFQ. The normalized ratios for each protein were log$_2$ transformed and plotted in the scatterplot. Filtering and analysis were done in R using existing libraries and an in-house script.

In vitro single- or double-strand binding of proteins from *C. elegans* extract.

For this assay, biotinylated oligonucleotides (Metabion) were used, containing a five times repeat of telomeric G-rich, C-rich, or control sequences. To allow for proper annealing, all oligonucleotides contained unique sequences flanking both sides of the repeats. Double-stranded oligonucleotides were prepared by mixing the biotinylated forward oligonucleotide 1:1 with the respective non-biotinylated reverse complement oligonucleotide and addition of annealing buffer (200 mM Tris-HCl, pH 8.0, 100 mM MgCl$_2$, 1 M KCl). The mix was heated at 80 °C for 5 min and cooled to room temperature. The single-stranded oligonucleotides were treated similarly, only replacing the reverse compliment oligonucleotide with H$_2$O. The pulldown itself was performed as described above with 0.5 mg (TEBP-2::GFP) or 0.4 mg (TEBP-1::3xFLAG) *C. elegans* embryo total protein extract of the respective strains. After elution, the samples were run on a 4–12% Bis-Tris gel (NuPAGE, Thermo Scientific, #NP0321) at 150 V for 120 min and transferred to a membrane. Western blot detection of the tagged proteins was carried out as described below.

Expression and purification of recombinant protein from *E. coli*. Auto-

induction[72] was used for expression of His$_6$-MBP-POT-2. An overnight culture of the expression strain BL21(DE3) was cultured at 37 °C in YG medium (2% Yeast extract, 0.5% NaCl, 3.5% Glycerol) supplemented with the respective antibiotic. A growing culture in YG medium was prepared by inoculating it with 1:50 volume of the overnight culture. At an OD$_{600}$ of 0.7, a culture of auto-induction medium (2% Peptone, 3% Yeast extract, 25 mM Na$_2$HPO$_4$/KH$_2$PO$_4$, 0.05% Glucose, 2.2% Lactose, 0.5% Glycerin, 50 mM NH4Cl, 5 mM Na$_2$SO$_4$, 2 mM MgSO$_4$, 1x Trace Metal Solution) was inoculated with the growing culture to a density of OD$_{600}$ 0.004. 1000x Trace Metal Solution used for the auto-induction medium, has the following constitution: of 50 mM FeCl$_3$/HCl, 20 mM CaCl$_2$, 10 mM Mn(II)Cl$_2$, 10 mM ZnCl$_2$, 2 mM CoCl$_2$, 2 mM Cu(II)Cl$_2$, 2 mM NiCl$_2$, 2 mM NaMoO$_4$, 2 mM Na$_2$SeO$_3$. The auto-induction culture was incubated at 25 °C for 24 h and then harvested by centrifugation at 4000 x g.

TEBP-1-His$_5$ and TEBP-2-His$_5$ were expressed in Rosetta 2 (DE3) pLysS competent cells (Novagen,#71401). An overnight culture was grown in LB containing the respective antibiotic. A growing culture was inoculated and after reaching mid-log growth at 37 °C, the cultures were induced with 1 mM IPTG. Cells were grown at 18 °C and harvested after 24 h. IPTG-induced or auto-induction cultures were pelleted in 50 ml reaction tubes by centrifugation at 4000 x g after growth and lysed according to the protocol for the respective downstream use.

POT-2 expression pellets were resuspended in Tris buffer (50 mM Tris/HCL pH 7.5, 100 mM NaCl, 10 mM MgCl$_2$, 1x EDTA-free protease inhibitor (Roche, #4693132001)) and divided into 2 ml flat lid micro tubes containing 0.1 mm zirconia beads (Carl Roth, #N033.1). Lysis of the cells was achieved with a FastPrep -24™ Classic (MP Biomedicals, #116004500) using the setting 6 m/s for 30 s for two times. In between the disruption cycles the samples were centrifuged at 21,000 x g for 2 min to pellet debris, followed by an incubation on ice for 5 min before the second cycle. After lysis the suspension was centrifuged at 21,000 x g for 10 min at 4 °C.

TEBP-1 and TEBP-2 expression pellets were lysed via sonication with a Branson Sonifier 450 (duty cycle: 50%, output control: 3, 3.5 min with 5 mm tip) in lysis buffer (25 mM Tris-HCl pH 7.5, 300 mM NaCl, 20 mM imidazole) with 1 mM DTT, and protease inhibitor cocktail tablets (Roche, #4693132001). Lysates were centrifuged at 4613 x g for 10 min at 4 °C. For both preparation methods the supernatant was afterwards transferred to fresh reaction tubes.

His-MBP tagged TEBP-1 and TEBP-2 fragments were expressed in E.coli ArcticExpress DE3 cells (Agilent, #230192). Cultures were grown overnight in 5 ml LB supplemented with the respective antibiotic for the expression vector. Next day the expression culture was inoculated from the overnight culture and grown to mid-log phase at 30 °C, and then induced with 1 mM IPTG. Cultures were incubated at 12 °C and harvested after 24 h. The pellet was resuspended in binding buffer (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 50 mM imidazole) with 1 mM DTT, complete protease inhibitor cocktail tablets (Roche, #4693132001), and 100 μg DNase I (NEB, M0303S). Cells were lysed using a Branson Sonifier (duty cycle: 50%, output control: 4, 6 min (3 min sonication, 3 min ice, 3 min sonication) with 9 mm tip). Lysates were cleared at 4613 x g for 10 min at 4 °C, and used for subsequent assays.

Protein expression, purification, and fluorescence polarization assay. *E.coli* ArcticExpress DE3 cells (Agilent, #230192) were grown overnight in 5 ml LB supplemented with the respective antibiotic for the expression vector. Next day the expression culture was inoculated from the overnight culture and grown to mid-log phase at 30 °C, and then induced with 1 mM IPTG. Cultures were incubated at 12 °C and harvested after 24 h. The pellet was resuspended in binding buffer (20 mM Tris-HCl pH 7.5, 500 mM NaCl, 50 mM imidazole) with 1 mM DTT, complete protease inhibitor cocktail tablets (Roche, #4693132001), and 100 μg DNase I (NEB, M0303S). Cells were lysed using a Branson Sonifier (duty cycle: 50%, output control: 4, 6 min (3 min sonication, 3 min ice, 3 min sonication) with 9 mm tip). Lysates were ultracentrifuged (Beckman Optima XE-100) at 75,000 x g for 30 min at 4 °C. After loading the lysate, the HisTrap HP column (GE Healthcare, #GE17-5247-01) was washed with binding buffer, and proteins were eluted in binding buffer containing 500 mM imidazole in 250 μl fractions. Proteins were dialyzed with the PD-10 Desalting Column (GE Healthcare, #GE17-0851-01) in a buffer consisting of 20 mM Tris-HCl pH = 7.5, 1 mM MgCl2, 150 mM NaCl, 10% (v/v) glycerol, and 1 mM DTT, and were concentrated. These fractions were then utilized for the fluorescence polarization assays.

The purified protein stocks were used from a maximum concentration of 4 μM, to a minimum concentration of 2 nM in twofold serial dilutions in ice-cold buffer containing 20 mM HEPES pH 7.0, 100 mM NaCl, and 5% (v/v) glycerol. FITC-labeled oligonucleotides (Metabion) carrying 2.5x, 2.0x, and 1.5x repeats of either telomeric (G- or C-rich), or control sequence were used for this assay. Double-stranded oligonucleotides were prepared by mixing 1:1 with the respective reverse complement oligonucleotide. For annealing, oligonucleotides were heated to 95 °C and then cooled at 0.1 °C/s until 4 °C. Diluted proteins were incubated with a final concentration of 20 nM FITC-labeled probe for 10 min at room temperature. Samples were measured with a Tecan Spark 20 M (Tecan). Experiments were conducted using three replicates for each condition. Analysis was performed with Graph Pad Prism 9.0 and specific binding was measured with Hill slope.

***C. elegans* complete protein extract preparation**. Animals were washed off the plates with M9 buffer, synchronized by bleaching and grown to the desired stage, at which point worms were collected and washed 3–4 times with M9 buffer. Worms were washed 3–4 times in M9, washed one last time with H$_2$O and frozen in 100–200 μl aliquots. Upon extract preparation, the aliquots were thawed, mixed 1:1 with 2x Lysis Buffer (50 mM Tris/HCl pH 7.5, 300 mM NaCl, 3 mM MgCl$_2$, 2 mM DTT, 0.2 % Triton X-100, Protease inhibitor tablets), and sonicated in a Bioruptor 300 (Diagenode) for 10 cycles with 30 s on/off, on high level. After sonication, the samples were centrifuged at 21,000 x g for 10 min to pellet cell debris. The supernatant was transferred to a fresh tube. With the exception of embryos (see below), extract of all developmental stages of *C. elegans* was prepared as described above. Samples of each developmental stage (for Fig. 2a) were collected in the following time points after plating of synchronized L1s: L1s were collected ~7 h after plating to recover

from starvation; L2s, ~12 h; L3s, ~28 h; L4, ~49 g; and YAs were collected ~ 56 h after plating.

For mixed-stage embryo extract preparations, synchronized gravid adults were harvested by washing them off the plate with M9 buffer. The worm suspension was washed with M9 until the supernatant was clear. Then, animals were bleached until all gravid adults were dissolved and only mixed-staged embryos remained. The embryos were subsequently washed in M9 buffer for three times then transferred to a new tube and washed one more time. In the last wash step the embryos were resuspended in 1x lysis buffer (25 mM Tris/HCl pH 7.5, 150 mM NaCl, 1.5 mM MgCl$_2$, 1 mM DTT, 0.1 % Triton X-100, protease inhibitors) and frozen in liquid nitrogen. After freezing, the pellets were ground to a fine powder in a pre-cooled mortar, then transferred to a cold glass douncer and sheared for 40 strokes with piston B. The suspension was pipetted to 1.5 ml tubes and spun down at 21,000 x g for 15 min at 4 °C. Finally, the supernatant was transferred to a fresh tube.

Immunoprecipitation (IP)
GFP IP. IPs with GFP-tagged proteins were performed with GFP-binding magnetic agarose beads (GFPtrap MA, Chromotek, #gtma-20). Per IP sample, 10 μl of bead slurry was used and washed two times with 500 μl Wash Buffer (10 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1:1000 Pepstatin A/Leupeptin, 1:100 PMSF). Afterwards, the beads were resuspended in 450 μl Wash Buffer and up to 1 mg of complete extract of the respective *C. elegans* strain (of mixed-stage embryos or young adults) was added to a final volume between 500 and 750 μl. The IP samples were incubated at 4 °C rotating for 2 h. Following three washing steps with 500 μl Wash Buffer the beads were resuspended in 1x LDS (4x NuPAGE LDS sample buffer, Thermo Scientific, #NP0008) supplemented with 100 mM DTT and boiled at 70 °C for 10 min. When used for mass spectrometry, the samples were prepared in quadruplicates per strain/condition. In the IP-MS related to Supplementary Fig. 5e, f, the Wash Buffer was supplemented with 2 mM MgCl$_2$ and 0.05% of recombinant endonuclease from *Serratia marcescens*, or Sm nuclease[73], produced by the IMB's Protein-Production Core Facility.

FLAG IP. IPs with FLAG-tagged protein were performed with Protein G magnetic beads (Invitrogen™ Dynabeads™ Protein G; #10004D) and α-FLAG antibody (Monoclonal ANTI-FLAG® M2 antibody produced in mouse, Sigma Aldrich, #F3165). Per IP, 30 μl of beads were used and washed three times with 1 ml Wash Buffer (25 mM Tris/HCl pH 7.5, 300 mM NaCl, 1.5 mM MgCl$_2$, 1 mM DTT, 1 complete Mini protease inhibitor tablet per 50 ml). The beads were resuspended in 450 μl Wash Buffer and up to 1 mg of complete protein extract from the respective *C. elegans* strains was added. Finally, 2 μg of FLAG antibody were added and the samples were incubated for 3 h, rotating at 4 °C. After the incubation, the samples were washed three to five times with 1 ml Wash Buffer (see washing steps before), the beads were resuspended in 1x LDS/DTT, and the samples were boiled at 95 °C for 10 min. For mass spectrometry, IPs were prepared in quadruplicates per strain/condition. When doing the IP with Sm nuclease, the wash buffer was supplemented with 0.05% Sm nuclease (as indicated above).

Western blot. Protein samples were boiled at 70 °C for 10 min and loaded on a 4–12% Bis-Tris gel (NuPAGE, Thermo Scientific, #NP0321), running at 150-180 V for 60–120 min in 1x MOPS. After the run, the gel was shortly washed in VE H$_2$O and equilibrated in transfer buffer (25 mM Tris, 192 mM Glycine, 20% Methanol). A nitrocellulose membrane (Amersham Protran, VWR, #10600002) was equilibrated in transfer buffer as well. Membrane and gel were stacked with pre-wet Whatman paper (GE Healthcare-Whatman, #WHA10426892) and immersed in a blotting tank (Bio-Rad) filled with ice-cold transfer buffer and additionally cooled with a cooling element. The proteins were blotted at 300 mA for 60–120 min depending on the size. If blotted for 90–120 min for larger proteins, the transfer was carried out with a blotting tank on ice to keep the temperature. After blotting, the membranes were further prepared according to the respective antibody protocol.

Anti-His antibody. Membranes were blocked in Blocking Solution (PentaHis Kit, Qiagen, #34460) for 1 h at room temperature. After three 5 min washes in TBS-T (1x TBS, 0.1% Tween-20, 0.5% Triton X-100) the membranes were incubated with the Anti-His-HRP conjugated antibody in a dilution of 1:1000 in Blocking Solution for 1 h at room temperature. The membranes were then washed again three times in TBS-T and incubated with ECL Western Blot reagent (Thermo Scientific™ SuperSignal™ West Pico PLUS Chemiluminescent Substrate, #15626144; mixed 1:1) for detection. Western blot ECL detection was performed with the ChemiDoc XRS+system (BioRad, Software: Image Lab 5.2.1).

Anti-GFP, Anti-FLAG, and Anti-Actin antibodies. Western blot analysis was performed using the following primary antibodies: an anti-GFP antibody (Roche, Anti-GFPfrom mouse IgG1κ (clones 7.1 and 13.1), #11814460001; 1:1000 in Skim Milk solution), an anti-FLAG antibody (Sigma-Aldrich, mouse Monoclonal ANTI-FLAG® M2 antibody, # F3165; 1:5000 in Skim Milk solution), and an anti-Actin antibody (Sigma-Aldrich, rabbit anti-actin, #A2066; 1:500 in Skim Milk solution). After blotting, membranes were blocked in Skim Milk solution (1x PBS, 0.1% Tween-20, 5% (w/v) Skim Milk Powder) for 1 h at room temperature. The

incubation with the primary antibody was carried out at 4 °C, rotating overnight. Membranes were washed in PBS-T (1x PBS, 0.1% Tween-20) three times for 10 min, they were incubated with an HRP-linked secondary antibody (for anti-flag and anti-GFP with Cell Signaling Technology, anti-mouse IgG, #7076; 1:10,000 dilution in Skim Milk Solution; for anti-actin the secondary used was GE Healthcare, anti-Rabbit IgG, #NA934; 1:3000 in Skim Milk solution) for 1 h rotating at room temperature. Following three washes in PBS-T the membranes were incubated with ECL solution (Thermo Scientific™ SuperSignal™ West Pico PLUS Chemiluminescent Substrate, #15626144; mixed 1:1) for detection. Western blot ECL detection was performed with the ChemiDoc XRS+system (BioRad, Software: Image Lab 5.2.1). Incubation with Anti-Actin antibody was typically performed after detection of GFP/FLAG and subsequent washes.

Antibody protocol for co-IPs (LI-COR antibodies). For co-IP experiments, we first probed the IP bait with HRP-linked secondary antibodies, as described above. Then, we probed for the co-IP using LI-COR secondary antibodies. After incubation with primary antibody, as described above, membranes were washed and incubated with secondary antibodies compatible with the LI-COR System (FLAG/GFP: Licor IRDye® 680RD Donkey anti-Mouse IgG (H+L), #926-68072; Actin: Licor IRDye® 800CW Donkey anti-Rabbit IgG (H+L), #926-32213; both 1:15,000 in Skim Milk solution) for 1 h at room temperature. After three additional washes with PBS-T, the membranes were imaged using an Odyssey CLx scanner and processed using Image Studio software (LI-COR, Version 3.1).

C. elegans culture and strains. *C. elegans* was cultured under standard conditions on Nematode Growth Medium (NGM) plates seeded with *E. coli* OP50 bacteria[74]. For proteomics experiments, animals were grown on OP50 high-density plates (adapted from ref. [75]). In specific, the yolks of commercially available chicken eggs were isolated, added to LB medium (50 ml per egg yolk) and thoroughly mixed. Subsequently, the mix was incubated at 65 °C for 2–3 h. Pre-grown OP50 liquid culture is added to the mix (10 ml per egg), after the yolk-LB mixture cooled down. This preparation was poured on 9 cm plates (10 ml per plate) and plates are decanted the next day. Plates remained for 2–3 days at room temperature, for further bacterial growth and drying.

Animals were grown at 20 °C, except when noted. The standard wild-type strain used in this study was N2 Bristol. Strains used and created in this study are listed in Supplementary Table 1.

Fertility assays. For brood size counts of the homozygous single mutants, L3 worms were isolated, per strain and were grown either at 20 or 25 °C. After reaching adulthood, worms were transferred to a new plate every day, until no eggs were laid in 2 consecutive days. Viable progeny was counted approximately 24 h after removing the parent. For the experiment shown in Fig. 4d, e, a cross between *tebp-1(xf133)* males and *tebp-2(xf131)* hermaphrodites was performed, the genotypes of the F1 and F2 were confirmed by PCR genotyping. L2/L3 progeny of F2 *tebp-1(xf133)/ + ; tebp-2(xf131)* mothers were isolated and grown at 20 °C, or 25 °C. During adulthood, the viable brood size was counted as mentioned above. The assayed F3s were genotyped 2 days after egg laying stopped. For all brood size experiments, worms that died before egg laying terminated, e.g., by dehydration on the side of plate, were excluded from the analysis.

Mortal germline assay. All strains used in the Mortal Germline assay were outcrossed with wild-type N2 two times before the experiment. Six L3 larvae of the chosen strains were picked per plate (*n* = 15 plates per strain) and grown at 25 °C. Six L3 larvae were transferred to a fresh plate every 5 days (equivalent to two generations). This procedure was followed until plates were scored as sterile, when the six worms transferred failed to produce six offspring to further isolate, on 2 consecutive transfer days.

***pgl-1::mTagRfp-T; tebp-1* x *pgl-1::mTagRfp-T; tebp-2* cross and definition of categories of germline defects**. We crossed *pgl-1::mTagRfp-T; tebp-1* males with *pgl-1::mTagRfp-T; tebp-2* hermaphrodites. F1 cross progeny was confirmed by genotyping. 300 F2 progeny were singled and left to self-propagate. After genotyping F2 worms, we isolated 60 F3 worms from three different *tebp-1(xf133);tebp-2(xf131)/ +*, 60 F3 worms from three different *tebp-1(xf133)/ + ;tebp-2(xf131)* mothers, as well as 10 F3 worms from two different single mutant mothers as controls. Additionally, all synthetic sterility escaper progeny from *tebp-1; tebp-2* double-homozygous worms were singled to check their fertility. Germline health, as well as growth and other phenotypes for all singled worms were determined at day 2 of adulthood. Germlines were categorized by microscopy with a Leica M80 stereomicroscope with a fluorescence lamp (Leica EL 6000), according to the morphology of the germline, as assessed by PGL-1::mTagRFP-T expression: category 1, near wild-type morphology; category 2, one gonad arm is atrophied; category 3, both gonad arms are atrophied. After germline categorization, worms were genotyped. We repeated this procedure until the F5, always using the progeny of *tebp-1(xf133);tebp-2(xf131)/ +* or *tebp-1(xf133)/ + ,tebp-2(xf131)* mothers, as well as sibling controls. The barplots depicting the final distribution of germline categories across all scored generations was created using R and publicly available packages (ggplot2-v 3.2.1, reshape–v 0.8.8, viridis–v 0.5.1, scales–v 1.0.0).

Scoring crosses of *tebp-1* x *tebp-2* mutant animals. Owing to the onset of synthetic sterility in F2 *tebp-1; tebp-2* double mutant animals, > 100 of F2 progeny was singled from the F1 heterozygous parent. F2 worms were genotyped at the adult stage after 3–4 days of egg laying and genotypes were determined and correlated with fertility. Progeny descending from *tebp-1; tebp-2* double mutant synthetic sterility escaper F2s were singled and allowed to grow and lay eggs for 3–4 days. Subsequently, these double mutant F3s were genotyped and their fertility was determined. Boxplots depicting the results were created using R and publicly available packages (ggplot2-v 3.2.1, reshape–v 0.8.8, viridis–v 0.5.1, scales–v 1.0.0).

Creation of mutants using CRISPR-Cas9 technology. Mutants were created as described[76], with the following specifications. To create *tebp-2(xf131)*, N2 animals were injected with a mix of three constructs: 25 ng/μl of co-injection marker pCFJ104 (*Pmyo-3:mCherry:unc-54 3'UTR*, a gift from Erik Jorgensen, Addgene plasmid #19328; http://n2t.net/addgene:19328; RRID:Addgene_19328); 100 ng/μl of a construct expressing Cas9 and a sgRNA targeting the sequence ACAT-GAGTCTGTGTTTACGG (derived from pDD162, which was a gift from Bob Goldstein, Addgene plasmid # 47549; http://n2t.net/addgene:47549; RRID: Addgene_47549); and 75 ng/μl of a construct expressing a sgRNA targeting ACGGCTCATAAGAGACTTGG (derived from p46169, which was a gift from John Calarco, Addgene plasmid # 46169; http://n2t.net/addgene:46169; RRID: Addgene_46169).

To produce *tebp-1(xf133)* and *tebp-1(xf134)*, the following mix was injected into N2 animals: 25 ng/μl of pCFJ104; 150 ng/μl of a construct expressing Cas9 and a sgRNA targeting the sequence GCATGTCGAGATTCTACTGG (derived from pDD162); and 80 ng/μl of a construct expressing a sgRNA targeting GCTTCAAAATTTCTCCAGGG (derived from p46169). After isolation, PCR genotyping and confirmation by Sanger sequencing, mutants were outcrossed four times against the wild type.

Creation of endogenous tags and a *tebp-1; pot-2* double mutant via CRISPR-Cas9-mediated genome editing. Protospacer sequences were chosen using CRISPOR (http://crispor.tefor.net)[77], cloned in pRFK2411 (plasmid expressing Cas9 + sgRNA(F+E);[78] derived from pDD162) or pRFK2412 (plasmid expressing sgRNA(F+E)[78] with Cas9 deleted; derived from pRFK2411) via site-directed, ligase-independent mutagenesis (SLIM)[79,80]. pDD162 (Peft-3::Cas9 + Empty sgRNA) was a gift from Bob Goldstein (Addgene plasmid # 47549; http://n2t.net/addgene:47549; RRID:Addgene_47549)[81]. All plamids were purified using NucleoSpin® Plasmid from Macherey-Nagel, eluted in sterile water and confirmed by enzymatic digestion and sequencing. All Cas9 nuclease induced double-strand breaks (DSBs) were within 20 bp distance to the desired editing site. All CRISPR-Cas9 genome editing was performed using either *dpy-10(cn64)* or *unc-58(e665)* co-conversion strategies[82]. Single-stranded oligodeoxynucleotides (ssODN, 4 nmole standard desalted Ultramer™ DNA oligo from IDT) and PCR products (purified using QIAquick® PCR Purification Kit from QIAGEN) served as donor templates for small (3xFLAG epitope tag, protospacer sequences) and big (GFP tag) insertions, respectively. The *gfp* coding sequence including three introns and flanking homology regions was amplified from pDD282, which was a gift from Bob Goldstein (Addgene plasmid # 66823; http://n2t.net/addgene:66823; RRID: Addgene_66823)[83]. All donor templates contained ~35 bp homology regions[84,85]. Plasmid vectors, ssODN and PCR products were diluted in sterile water and injected at a final concentration of 30–50 ng/μl, 500–1000 nM and 300 ng/μl, respectively. For GFP insertions, the protospacer sequence used for the *dpy-10* co-conversion was transplanted to the editing site to generate d10-entry strains[86], which was used as reference strains for further injections. DNA mixes were injected in both gonad arms of 10–25 1-day-old adult hermaphrodites maintained at 20 °C. Co-converted F1 progeny were screened for insertions by PCR. Successful editing events were confirmed by Sanger sequencing. All generated mutant strains were outcrossed at least two times prior to any further cross or analysis. CRISPR-Cas9 genome editing reagents and DNA injection mixes are listed in Supplementary Data file 5. The *pgl-1::mTagRfp-T* is described elsewhere[49,50].

Creation of transgenic worms using MosSCI. A TEBP-2::GFP fusion transgene was produced as previously described[87], and as indicated in www.wormbuilder.org. Animals of the strain EG6699 were injected, in order to get insertions in locus ttTi5605 on LGII. The injection mix contained all the injection constructs listed in www.worbuilder.org, using the recommended concentrations, including 50 ng/μl of a repair template containing the *tebp-2::gfp* sequence. Selection was performed as recommended in www.wormbuilder.org[76].

Extraction of genomic DNA from *C. elegans*. Mixed-staged animals were washed off plates with M9 and washed two to three more times in M9. Next, worms were resuspended in Worm Lysis buffer (WLB: 0.2 M NaCl, 0.1 M Tris/HCl pH 8.5, 50 mM EDTA, 0.5% SDS) and aliquoted in 250 μl samples. For genomic DNA extraction the aliquots were brought to a final volume of 500 μl with WLB and Proteinase K (30 μg/ml). To lyse the worms, the samples were incubated at 65 °C at 1400 rpm for > 2 h until all carcasses were dissolved. The samples were then centrifuged at 21,000 x *g* for 5 min to pellet debris and the supernatant was transferred to a fresh tube. Afterwards, 500 μl of Phenol:Chloroform:

47

Isoamylalcohol were added, the samples shaken vigorously for 30 s and spun down at 16,000 x g for 5 min. Additionally, 500 µl of chloroform were added to the samples and again shaken vigorously for 30 sec and spun at 16,000 x g for 5 min. The aqueous phase of the samples was transferred to fresh 2 ml reaction tubes and 50 µg RNase A were added to digest the RNA. The tubes were inverted once and incubated at 37 °C for > 1 h. After RNA digestion the samples were again purified by phenol:chloroform:isoamylalcohol and chloroform addition (as before). The aqueous phase was transferred to fresh tubes and the DNA was precipitated with 350 µl isopropanol for > 15 min at −80 °C. To pellet the DNA, the samples were centrifuged at 21,000 x g for 20 min at 4 °C. The supernatant was carefully removed and the DNA pellet washed once with 1 ml of ice-cold 70% ethanol and spun at 21,000 x g for 5 min at 4 °C. Washing was repeated if the samples still smelled of phenol. After washing the supernatant was completely removed, the pellet air dried for ca. 10 min, and resuspended in 20 µl H$_2$O. To fully resuspend the DNA, the samples were kept at 4 °C overnight and mixed again the next day.

Telomere Southern blot. For denatured telomere Southern blot 15 µg of *C. elegans* genomic DNA were digested in 80 µl total volume with 40 U HinfI (New England BioLabs, #R0155) and RsaI (New England BioLabs, #R0167), respectively. The digestion was incubated at 37 °C overnight and the next day additional 10 U of each enzyme were added and the samples incubated 1–2 h further. Afterwards the samples were evaporated in a Concentrator Plus at 45 °C to end up with a volume of 20–30 µl and supplemented with 2x DNA loading dye. A 0.6% agarose gel was prepared (with 1x TBE and 16 µl SYBR Safe DNA stain, Thermo Fischer Scientific, #S33102) and the samples loaded after boiling at 95 °C for 10 min. The GeneRuler 1 kb (Thermo Scientific, #SM0312), as well as the 1 kb extended markers (New England BioLabs, #N3239) were used. The samples were secured in the gel by running it at 100 V for 20–30 min then the voltage was set to 60 V for a run overnight (16–19 h). With a crosslinker set to 1 min crosslinking time, the DNA was broken and the gel afterwards equilibrated in transfer buffer (0.6 M NaCl, 0.4 M NaOH) for at least 20 min. After equilibration, an upward alkaline transfer was set up with whatman paper and a positively charged nylon membrane (Byodine B membrane; Pall, #60207), all equilibrated in transfer buffer. The transfer was set up overnight. Following blotting, the membrane was fixed by incubation in 0.4 M NaOH for 15 min with slight agitation and neutralized with two washes in 2x SSC for 5 min each. To keep hydrated the membrane was sealed in cling film with 2x SSC until hybridization.

The membrane was pre-hybridized in a glass hybridization tube with 20 ml hybridization buffer (3.3x SSC, 0.1% SDS, 1 mg/ml Skim Milk powder) for at least 1 h at 42 °C rotating in a hybridization oven. The oligonucleotide used for detection was a TTAGGC reverse complement triple repeat (GCCTAA)$_3$. The probe was radioactively labeled with 3 µl 32P-[γ]-ATP by a polynucleotide Kinase reaction and cleaned up using a MicroSpin Sephadex G-50 column (GE Healthcare, #GE27-5330-01). The labeled oligonucleotide was denatured at 95 °C for 10 min and mixed with 20 ml fresh hybridization buffer. This mix was added to the membrane after removing the previous buffer and incubated for 3.5 days rotating at 42 °C.

After hybridization the membrane was washed by first rinsing it twice with Wash Buffer 1 (2x SSC, 0.1% SDS), then incubating it twice for 5 min in 20 ml Wash Buffer 1. For the last wash, the membrane was incubated for 2 min in Wash Buffer 2 (0.2x SSC, 0.1% SDS), then rinsed in 2x SSC to re-equilibrate the salt concentration. The membrane was dried on a Whatman paper for 3 h, sealed in cling film and exposed to a phosphoimager screen for 3 days. The screen was read out with the Typhoon Scanner with the settings 1000 V PMT and 200 µm pixel size. Contrast and brightness of the resulting tif-file were optimized using Fiji.

Microscopy

Co-localization microscopy. Strains carrying TEBP-1::GFP or TEBP-2::GFP were crossed with strain YA1197 expressing POT-1::mCherry. Adult animals were washed in M9 buffer, immobilized in M9 buffer supplemented with 40 mM sodium azide and mounted on freshly made 2% agarose pads. For imaging embryos, adult hermaphrodites were washed and dissected in M9 buffer before mounting. Animals were immediately imaged using a TCS SP5 Leica confocal microscope equipped with a HCX PL APO 63x water objective (NA 1.2), Leica hybrid detectors (HyD), and the acquisition software Leica LAS AF. Deconvolution was performed using Huygens Remote Manager and images were further processed using Fiji.

PGL-1 fluorescence microscopy. For imaging PGL-1::mTagRFP-T in animals of each category of germline morphology, adult worms were picked to a droplet of M9 to remove OP50 bacteria, then transferred to a drop of M9 buffer supplemented with 40 mM sodium azide in M9 for immobilization on a 2% agarose pad. Animals were immediately imaged with a Leica AF7000 widefield microscope using a 20x objective (NA 0.4) and red fluorescence filters (N3), as well as TL-DIC (acquisition software: Leica LAS X, camera: Hamamatsu, Orca Flash 4.0 V2). Images were processed using Fiji (brightness changes applied only in DIC channel for better visualization).

Quantitative FISH (qFISH). For telomere length determination, fluorescence in situ hybridization (FISH) was utilized in a quantitative manner[88]. The staining protocol was optimized after the work of Seo and Lee[89]. Per strain, 100 gravid adults were

picked to an unseeded small NGM plate to remove the majority of OP50 bacteria. From there, worms were picked to a 5 µl drop of Egg buffer (25 mM HEPES/KOH pH 7.4, 118 mM NaCl, 48 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 1% Tween-20) on a cover slip and dissected using 20 gauge needles (Sterican, Roth #C718.1) to release embryos and gonads. The samples were fixed by adding 5 µl of 2% Formaldehyde solution and incubating for 5 min. To remove the Formaldehyde solution, samples were washed on the cover slip by adding and removing Egg buffer carefully by pipetting. For permeabilization of the cuticle, the worms were afterwards treated by freeze cracking[90]. The cover slips were put on a Poly-lysine coated slide (Sigma Aldrich, #P0425) and the slides transferred to an aluminum block on dry ice for freezing. After 15 min freezing on the aluminum block, the cover slips were removed and the slides immersed first in ice-cold methanol, then in ice-cold acetone for 5 min, respectively. To remove the solutions the slides were washed in 1x PBS (10 mM Na2HPO4, 2 mM KH2PO4, 137 mM NaCl, 2.7 mM KCl) for 15 min. For additional permeabilization the samples were incubated in permeabilization buffer (20 mM Tris/HCl pH 7.5, 50 mM NaCl, 3 mM MgCl$_2$, 300 mM Sucrose, 0.5% Triton X-100) at 37 °C for 30 min followed by a wash in 1x PBS for 5 min at room temperature. To prevent unspecific binding of the FISH probe, the samples were treated with 20 µl RNase A solution (1x PBS, 0.1% Tween-20, 10 µg/ml RNase A) at 37 °C for 1 h in a humid chamber. Afterwards the slides were washed in 1x PBS-T (1x PBS, 0.1% Tween-20) for 10 min at room temperature and dehydrated by successive 3 min washes in 70%, 85 and 100% ethanol and air dried. For pre-hybridization 50 µl of hybridization solution (3X SSC, 50% Formamide, 10% (w/v) Dextran-Sulfate, 50 µg/ml Heparin, 100 µg/ml yeast tRNA, 100 µg/ml sheared salmon sperm DNA) were added to the sample and the slides incubated in a humid chamber for 1 h at 37 °C. The FISH probe (PNA-FISH TTAGGC telomeric probe, Panagene, resuspended to 100 µM, fluorophore: Alexa-555) was prepared as a 1:500 dilution in hybridization solution and denatured for 5 min at 70 °C. After pre-hybridization, the solution on the slides was removed as much as possible by pipetting and 20 µl of FISH probe were added, then covered by a cover slip. For hybridization of the probe the slides were denatured on a heat block prepared with wet paper towels for humidity at 80 °C for 3 min and transferred to a humid chamber for incubation overnight at 37 °C. The next day the slides were washed twice in 1x PBS-T for 5 min to remove the probe. To fixate the staining, the samples were incubated in hybridization wash solution (2X SSC, 50% Formamide) for 30 min at 37 °C. As a last step the slides were washed in 1x PBS-T twice for 15 min at room temperature and mounted by adding 10–20 µl Vectashield mounting medium containing DAPI (Vector laboratories, #H-1200-10). The pictures were taken with a Leica TCS SP5 confocal microscope (objective: CX PL APO CS 63sx oil NA: 1.4, pinhole 60.05 µm, 2x zoom, PMT detectors, acquisition software Leica LAS AF). The images stacks were composed of a sequence of pictures acquired every 0.5 µm on the z-axis. The laser and gain settings were adjusted according to the sample with the lowest FISH intensity. For analysis, images were opened in Image J/Fiji and the channels split into the DAPI and red channel. A mask of the image was created to infer the volume of the imaged object. The threshold function of the software was used with activated plugins for identification of round objects (Otsu). After setting the threshold for the image in the histogram settings, the z-stack was converted to a binary mask and using the 3D OC Options menu volume, mean gray values and integrated density of the FISH foci were calculated. Additionally, the 3D Object counter menu was used and the filters set to a minimum of 2. The values obtained by this analysis were averaged over several images of either germlines or embryos of the same strain and used for quantitative comparison of telomere length. For comparison, all values obtained for the mutant strains were scaled relative to the average of the wild type values. The barplots were created using R with standard and publicly available scripts (RColorBrewer-v 1.1-2, ggpubr-v 0.4.0, plyr-v 1.8.6, viridis-v 0.5.1, viridisLite-v 0.3.0, ggforce-v 0.3.2, ggsignif-v 0.6.0, dplyr-v 1.0.2, ggplot2-v 3.3.3, readr-v 1.4.0).

Yeast two-hybrid assay. Yeast two-hybrid assays were conducted in the yeast strain PJ69-4α as described before[91,92]. The respective Gal4 activation and DNA-binding domain plasmid pairs were co-transformed in PJ69-4α. The resulting transformants were resuspended in ddH$_2$O and pinned on SC Trp-Leu-, SC Trp-Leu-His-, and SC Trp-Leu-His-Ade- plates. For Fig. 6a an additional round of plasmid transformation was performed, as a biological duplicate, and the results were identical. Colonies were imaged with a ChemiDoc XRS+system (BioRad, Software: Image Lab 5.2.1) for Fig. 6a and Supplementary Fig. S6g, and scanned with an Epson Scanner (Perfection V700 Photo, Software version 3.81) for Fig. 6f–h and supplementary Fig. 6i.

Size-exclusion chromatography. Size-exclusion chromatography was performed as previously described[76,92]. For the first run (Supplementary Fig. 5a) two embryo samples were prepared and combined. Using a centrifugal filter with a 10 kDa cutoff (Merck, Amicon Ultra 0.5 ml 10 K, #UFC5010) the sample was concentrated to a final volume of 550 µl. Between 3.6 and 3.8 mg of total extract was separated on a Superose 6 10/300 GL column (GE Healthcare, 17517201) operated on a NGC Quest System (Bio-Rad) using lysis buffer without Triton X-100 as running buffer (25 mM Tris/HCl pH 7.5, 150 mM NaCl, 1.5 mM MgCl$_2$, 1 mM DTT, protease inhibitors). Five-hundred microliter fractions were collected according to the scheme in Supplementary table 2. Selected fractions were concentrated to 30 µl using 10 kDa cutoff centrifugal filters (Merck, Amicon Ultra 0.5 ML 10 K,

#UFC5010). The samples were supplemented with 4x LDS (NuPAGE) and 100 mM DTT to a final volume of around 40 µl and boiled at 95 °C for 10 min. After spinning down, a part of each sample was run on a 4-15% Criterion TGX Stain-Free Protein Gel (26 wells, Bio-Rad, #5678085) in 1x SDS running buffer at 200 V for 32 min. Transfer of proteins to a nitrocellulose membrane (Bio-Rad, #1620112) was performed using the Trans-Blot Turbo Transfer System (Bio-Rad). Following the transfer, western blot was performed as described above. For the second run (Fig. 5a, b), four embryo extracts were prepared, combined and concentrated, as above, to 1 ml. Then half of the sample was treated with Sm nuclease for 30 min at 4 °C, prior to size-exclusion chromatography, while the other half was not.

Phylogenetic and synteny analysis. The protein sequences of *C. elegans* TEBP-1 and TEBP-2 were extracted from Wormbase (WS275). These sequences were used separately as queries for Wormbase BLASTP search in the available genomes. Orthologs of TEBP-1 and TEBP-2 were defined based on two criteria: (1) BLASTP hit had an E-value lower than 1.00e-15; and (2) reciprocal BLASTP of the hit, querying the *C. elegans* proteome, resulted in TEBP-1 and TEBP-2 as top hits. Sequences of the identified orthologs were obtained from Wormbase (WS275) and Wormbase ParaSite (WBPS14/WS271). The list of identified orthologs and BLASTP results can be found in Supplementary Data file 2 (sheet 1).

The full-length protein sequences of TEBP orthologs were used for multiple sequence alignment using MAFFT, version 7.452[93]. Alignment was performed using default settings, including an automatic determination of best alignment strategy, which provided the L-INS-I result[94]. Multiple sequence alignment can be found in Supplementary Data file 2 (sheet 2). Then, the multiple sequence alignment in fasta format was used as an input for IQ-TREE version 1.6.12[95], with branch supports obtained with ultrafast bootstrap[96]. IQ-TREE was first ran to determine the best fit substitution model, which was VT+F+R3. Then, analysis was repeated with the following parameters: -redo -m VT+F+R3 -bb 10000 -o Cang_2012_03_13_00535. g11959_Cang, Cang_2012_03_13_01061.g15539.t3_Can, where -m is the best fit model, -bb is the number of ultrafast bootstrap replicates, and –o represents the defined outgroups. Output.tree file was visualized in FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The *C. angaria* TEBP orthologs were used as outgroups, as this species is not part of the *Elegans* and *Japonica* groups, according to recent phylogenetic studies[97]. To create an additional tree with the N-terminal region only, the initial multiple sequence alignment was trimmed to the 600 initial alignment positions. The alignment of this region (with similarity to the homeodomain of RAP1) was substantially more reliable, as assessed by higher GUIDANCE2 scores[98]. Using this edited alignment, another tree was constructed as described above. IQ-TREE best fit model was VT+F+I+G4, parameters used: -m VT+F+I+G4 –bb 10000 -o Cang_2012_03_13_00535.g11959_Cang, Cang_2012_03_13_01061.g15539.t3_Can.

We defined local synteny across species as the maintenance of linkage in at least one of the neighboring genes upstream and downstream of the respective *tebp* gene. We used two different strategies to determine synteny. (1) Synteny was determined by navigating genome browser tracks through regions containing *tebp* orthologs, using Wormbase ParaSite (WBPS14/WS271). Currently annotated genes, adjacent to *tebp* orthologs, were selected, their predicted protein sequences were retrieved and BLASTP was performed in the *C. elegans* genome to find the corresponding ortholog. Results are summarized in Supplementary Data file 2 (sheet 3). (2) The protein sequences obtained previously by reciprocal BLASTP of TEBP-1 and TEBP-2 were used as an entry for WormBase ParaSite BioMart tool (https://parasite.wormbase.org/biomart). We recouped the neighboring 13 genes upstream and 13 genes downstream, and, with the resulting gene ID list, we determined a set of orthologous genes with the following series of 'Output attributes': gene stable ID, chromosome/scaffold, start (bp) and end (bp) coordinates that were to be listed in the result from ten available complete *Caenorhabditis* genomes. Subsequently, we filtered only those genes that share the same chromosome/scaffold with the *tebp* orthologous gene, finally, we evaluate if the enlarged group meets our definition of local synteny. We repeated this process taking each of the *tebp* genes in the ten species as a reference and evaluated the filtered groups for local synteny. In the specific case of *C. remanei*, WormBase ParaSite provides three different assemblies: PRJNA248909, PRJNA248911 and PRJNA53967. The latter was the only assembly where we were able to identify synteny of *tebp-1* with BioMart, although we could verify it manually for PRJNA248911. Results are summarized in Supplementary Data file 2 (sheet 4). This strategy was not applicable to *C. angaria*, as the genome of this species is not implemented in WormBase ParaSite BioMart.

Analysis of previously published RNA-seq datasets. For the expression data of the telomeric proteins during development of *C. elegans* (Supplementary Fig. 2a–c), RNAseq data was taken from a previously published dataset[47]. To probe expression of the telomeric genes in spermatogenic and oogenic gonads (Supplementary Fig. 2d), previously published transcriptome data was used[48]. Gene expression and genome browser tracks were plotted using Gviz[99] and GenomicFeatures[100] on an R framework (R Core Team 2018).

RNA extraction and library preparation. RNA was extracted as described[47]. Synchronized young adult animals were frozen in 50–100 µl of H2O after harvest.

After thawing, 500 µl TRIzol LS reagent (Invitrogen, # 10296010) was added and the worms were lysed with six freeze-thaw cycles (frozen in liquid nitrogen for ca. 30 s, then thawed for 2 min in a 37 °C waterbath and vortexed). Following lysis, the samples were spun down at full speed for 2 min to pellet debris. Supernatant was transferred to a fresh tube, mixed 1:1 with 100% ethanol and the mix was transferred to a column of the Direct-zol RNA MiniPrep Plus Kit (Zymo Research, #R2070). The following purification steps were done according to manufacturer's instructions, including the recommended in-column DNase I treatment for 25–40 min. RNA samples were eluted in 30–32 µl of RNase-free H2O.

Library preparation for mRNA sequencing was performed with Illumina's TruSeq stranded mRNA LT Sample Prep Kit following Illumina's standard protocol (Part # 15031047 Rev. E). Libraries were prepared by using only ¼ of the reagents with a starting amount of 250 ng and they were amplified in ten PCR cycles. Libraries were profiled in a High Sensitivity DNA on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life technologies). Libraries were pooled in an equimolar ratio and sequenced on one NextSeq 500 Highoutput Flowcell, SR for 1 × 75 cycles plus 1 × 7 cycles for index read.

mRNA read processing and mapping. The library quality was assessed with FastQC (version 0.11.8) before alignment against the *C. elegans* genome assembly WBcel235 and a custom.GTF file, which included gene annotations from *C. elegans* (WormBase, c_elegans.PRJNA13758.WS269) and *E. coli* (EnsemblBacteria, Escherichia_coli_b_str_rel606.ASM1798v1). Alignment was performed with STAR aligner[101] version 2.6.1b. Reads mapping to annotated features in the custom.GTF file were counted with featureCounts[102] version 1.6.2 using featureCounts functionality. Counts aligning to *E. coli* were removed at this point from downstream analysis. Coverage tracks were generated with deepTools[103] version 2.27.1 and plotted using Gviz[99] on an R framework (R Core Team 2018).

Data availability
The datasets supporting the conclusions of this article are available in the ProteomeXchange Consortium via Pride repository, PXD019241; and in the SRA, BioProject PRJNA630690.

Code availability
Code is available upon request.

References
1. Jain, D. & Cooper, J. P. Telomeric strategies: means to an end. *Annu. Rev. Genet.* **44**, 243–269 (2010).
2. de Lange, T. How telomeres solve the end-protection problem. *Science* **326**, 948–952 (2009).
3. Doksani, Y. The response to DNA damage at telomeric repeats and its consequences for telomere function. *Genes* **10**, 318 (2019).
4. Doksani, Y. & de Lange, T. The role of double-strand break repair pathways at functional and dysfunctional telomeres. *Cold Spring Harb. Perspect. Biol.* **6**, a016576 (2014).
5. Lingner, J., Cooper, J. P. & Cech, T. Telomerase and DNA end replication: no longer a lagging strand problem? *Science* **269**, 1533–1535 (1995).
6. Palm, W. & de Lange, T. How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* **42**, 301–334 (2008).
7. Soudet, J., Jolivet, P. & Teixeira, M. T. Elucidation of the DNA end-replication problem in *Saccharomyces cerevisiae*. *Mol. Cell* **53**, 954–964 (2014).
8. Allsopp, R. C. et al. Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl Acad Sci. USA* **89**, 10114–10118 (1992).
9. Harley, C. B., Futcher, A. B. & Greider, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458–460 (1990).
10. Hemann, M. T., Strong, M. A., Hao, L.-Y. & Greider, C. W. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell* **107**, 67–77 (2001).
11. Moyzis, R. K. et al. A highly conserved repetitive DNA sequence, (TTAGGG) n, present at the telomeres of human chromosomes. *Proc. Natl Acad Sci. USA* **85**, 6622–6626 (1988).
12. de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).

13. Conomos, D., Reddel, R. R. & Pickett, H. A. NuRD–ZNF827 recruitment to telomeres creates a molecular scaffold for homologous recombination. *Nat. Struct. Mol. Biol.* **21**, 760–770 (2014).

14. Déjardin, J. & Kingston, R. E. Purification of proteins associated with specific genomic loci. *Cell* **136**, 175–186 (2009).

15. Jahn, A. et al. ZBTB48 is both a vertebrate telomere-binding protein and a transcriptional activator. *EMBO Rep.* **18**, 929–946 (2017).

16. Kappei, D. et al. HOT1 is a mammalian direct telomere repeat-binding protein contributing to telomerase recruitment. *EMBO J.* **32**, 1681–1701 (2013).

17. Li, J. S. Z. et al. TZAP: A telomere-associated protein involved in telomere length control. *Science* **355**, 638–641 (2017).

18. Baumann, P. & Cech, T. R. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**, 1171–1175 (2001).

19. Cooper, J. P., Nimmo, E. R., Allshire, R. C. & Cech, T. R. Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature* **385**, 744–747 (1997).

20. Miyoshi, T., Kanoh, J., Saito, M. & Ishikawa, F. Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length. *Science* **320**, 1341–1344 (2008).

21. Conrad, M. N., Wright, J. H., Wolf, A. J. & Zakian, V. A. RAP1 protein interacts with yeast telomeres in vivo: Overproduction alters telomere structure and decreases chromosome stability. *Cell* **63**, 739–750 (1990).

22. Grandin, N., Reed, S. I. & Charbonneau, M. Stn1, a new Saccharomyces cerevisiae protein, is implicated in telomere size regulation in association with Cdc13. *Genes Dev.* **11**, 512–527 (1997).

23. Grandin, N., Damon, C. & Charbonneau, M. Ten1 functions in telomere end protection and length regulation in association with Stn1 and Cdc13. *EMBO J.* **20**, 1173–1183 (2001).

24. Levy, D. L. & Blackburn, E. H. Counting of Rif1p and Rif2p on Saccharomyces cerevisiae telomeres regulates telomere length. *Mol. Cell. Biol.* **24**, 10857–10867 (2004).

25. Lin, J.-J. & Zakian, V. A. The Saccharomyces CDC13 protein is a single-strand TG1–3 telomeric DNA-binding protein in vitro that affects telomere behavior in vivo. *Proc. Natl Acad. Sci. USA* **93**, 13760–13765 (1996).

26. Moretti, P., Freeman, K., Coodly, L. & Shore, D. Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. *Genes Dev.* **8**, 2257–2269 (1994).

27. König, P., Giraldo, R., Chapman, L. & Rhodes, D. The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell* **85**, 125–136 (1996).

28. Corsi, A. K., Wightman, B. & Chalfie, M. A. Transparent window into biology: a primer on *Caenorhabditis elegans*. *WormBook* 1–31 https://doi.org/10.1895/wormbook.1.177.1 (2015).

29. Wicky, C. et al. Telomeric repeats (TTAGGC)n are sufficient for chromosome capping function in Caenorhabditis elegans. *Proc. Natl Acad. Sci. USA* **93**, 8983–8988 (1996).

30. Ahmed, S. & Hodgkin, J. MRT-2 checkpoint protein is required for germline immortality and telomere replication in *C. elegans*. *Nature* **403**, 159–164 (2000).

31. Raices, M. et al. *C. elegans* telomeres contain G-strand and C-strand overhangs that are bound by distinct proteins. *Cell* **132**, 745–757 (2008).

32. Meier, B. et al. trt-1 Is the *Caenorhabditis elegans* catalytic subunit of telomerase. *PLOS Genet.* **2**, e18 (2006).

33. Cheng, C., Shtessel, L., Brady, M. M. & Ahmed, S. *Caenorhabditis elegans* POT-2 telomere protein represses a mode of alternative lengthening of telomeres with normal telomere lengths. *Proc. Natl Acad. Sci. USA* **109**, 7805–7810 (2012).

34. Lackner, D. H. & Karlseder, J. C. elegans survivors without telomerase. *Worm* **2**, e21073 (2013).

35. Lackner, D. H., Raices, M., Maruyama, H., Haggblom, C. & Karlseder, J. Organismal propagation in the absence of a functional telomerase pathway in *Caenorhabditis elegans*. *EMBO J.* **31**, 2024–2033 (2012).

36. Seo, B. et al. Telomere maintenance through recruitment of internal genomic regions. *Nat. Commun.* **6**, 1–10 (2015).

37. Shtessel, L. et al. *Caenorhabditis elegans* POT-1 and POT-2 repress telomere maintenance pathways. *G3: genes, genomes. Genetics* **3**, 305–313 (2013).

38. Meier, B. et al. The MRT-1 nuclease is required for DNA crosslink repair and telomerase activity in vivo in *Caenorhabditis elegans*. *EMBO J.* **28**, 3549–3563 (2009).

39. Casas-Vila, N., Scheibe, M., Freiwald, A., Kappei, D. & Butter, F. Identification of TTAGGG-binding proteins in Neurospora crassa, a fungus with vertebrate-like telomere repeats. *BMC Genomics* **16**, 965 (2015).

40. Kappei, D. et al. Phylointeractomics reconstructs functional evolution of protein binding. *Nat. Commun.* **8**, 1–9 (2017).

41. Cox, J. et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteom.* **13**, 2513–2526 (2014).

42. Hsu, J.-L., Huang, S.-Y., Chow, N.-H. & Chen, S.-H. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* **75**, 6843–6852 (2003).

43. Lowden, M. R., Meier, B., Lee, T. W., Hall, J. & Ahmed, S. End Joining at *Caenorhabditis elegans* telomeres. *Genetics* **180**, 741–754 (2008).

44. Zimmermann, L. et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).

45. Lustig, A. J., Kurtz, S. & Shore, D. Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. *Science* **250**, 549–553 (1990).

46. Li, B., Oestreich, S. & de Lange, T. Identification of human Rap1: implications for telomere evolution. *Cell* **101**, 471–483 (2000).

47. Almeida, M. V., Domingues, A. M., de, J. & Ketting, R. F. Maternal and zygotic gene regulatory effects of endogenous RNAi pathways. *PLOS Genet.* **15**, e1007784 (2019).

48. Ortiz, M. a, Noble, D., Sorokin, E. P. & Kimble, J. A New dataset of spermatogenic vs. oogenic transcriptomes in the nematode *Caenorhabditis elegans*. *G3 (Bethesda, Md.)* https://doi.org/10.1534/g3.114.012351 (2014).

49. Schreier, J. et al. A membrane-associated condensate drives paternal epigenetic inheritance in *C. elegans*. Preprint at *bioRxiv* https://doi.org/10.1101/2020.12.10.417311 (2020).

50. Gudipati, R. K. et al. Protease-mediated processing of Argonaute proteins controls small RNA association. Preprint at *bioRxiv*. https://doi.org/10.1101/2020.12.09.417253 (2020).

51. Kawasaki, I. et al. PGL-1, a predicted RNA-binding component of germ granules, is essential for fertility in *C. elegans*. *Cell* **94**, 635–645 (1998).

52. Strome, S. & Updike, D. Specifying and protecting germ cell fate. *Nat. Rev. Mol. Cell Biol.* **16**, 406–416 (2015).

53. Kanzaki, N. et al. Biology and genome of a newly discovered sibling species of *Caenorhabditis elegans*. *Nat. Communications* **9**, 3216 (2018).

54. Fisher, T. S. & Zakian, V. A. Ku: A multifunctional protein involved in telomere maintenance. *DNA Repair* **4**, 1215–1226 (2005).

55. Riha, K., Heacock, M. L. & Shippen, D. E. The role of the nonhomologous end-joining DNA double-strand break repair pathway in telomere biology. *Annu. Rev. Genet.* **40**, 237–277 (2006).

56. Im, S. H. & Lee, J. PLP-1 binds nematode double-stranded telomeric DNA. *Mol. Cells* **20**, 297–302 (2005).

57. Im, S. H. & Lee, J. Identification of HMG-5 as a double-stranded telomeric DNA-binding protein in the nematode Caenorhabditis elegans. *FEBS Lett.* **554**, 455–461 (2003).

58. Kim, S. H., Hwang, S. B., Chung, I. K. & Lee, J. Sequence-specific binding to telomeric DNA by CEH-37, a homeodomain protein in the nematode *Caenorhabditis elegans*. *J. Biol. Chem.* **278**, 28038–28044 (2003).

59. Coghlan, A. & Wolfe, K. H. Fourfold faster rate of genome rearrangement in nematodes than in drosophila. *Genome Res.* **12**, 857–867 (2002).

60. Saint-Leandre, B. & Levine, M. T. The telomere paradox: stable genome preservation with rapidly evolving proteins. *Trends Genet.* **36**, 232–242 (2020).

61. de Lange, T. Shelterin-mediated telomere protection. *Annu. Rev. Genet.* **52**, 223–247 (2018).

62. He, H. et al. POT1b protects telomeres from end-to-end chromosomal fusions and aberrant homologous recombination. *EMBO J.* **25**, 5180–5190 (2006).

63. van Steensel, B., Smogorzewska, A. & de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413 (1998).

64. Long, J. et al. Telomeric TERB1–TRF1 interaction is crucial for male meiosis. *Nat. Struct. Mol. Biol.* **24**, 1073–1080 (2017).

65. Shibuya, H. et al. MAJIN links telomeric DNA to the nuclear membrane by exchanging telomere cap. *Cell* **163**, 1252–1266 (2015).

66. Wang, Y. et al. The meiotic TERB1-TERB2-MAJIN complex tethers telomeres to the nuclear envelope. *Nat. Commun.* **10**, 1–19 (2019).

67. Ferreira, H. C., Towbin, B. D., Jegou, T. & Gasser, S. M. The shelterin protein POT-1 anchors *Caenorhabditis elegans* telomeres through SUN-1 at the nuclear periphery. *J. Cell Biol.* **203**, 727–735 (2013).

68. de Albuquerque, B. F. M. et al. PID-1 is a novel factor that operates during 21U-RNA biogenesis in *Caenorhabditis elegans*. *Genes Dev.* **28**, 683–688 (2014).

69. Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–2860 (2007).

70. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* **4**, 484–494 (2009).

71. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).

72. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

73. Ball, T. K., Saurugger, P. N. & Benedik, M. J. The extracellular nuclease gene of *Serratia marcescens* and its secretion from *Escherichia coli*. *Gene* **57**, 183–192 (1987).

74. Brenner, S. The genetics of *Caenorhabditis Elegans*. *Genetics* **77**, 71–94 (1974).

75. Schweinsberg, P. J. & Grant, B. D. C. elegans gene transformation by microparticle bombardment. *WormBook* 1–10 https://doi.org/10.1895/wormbook.1.166.1 (2013).

76. Almeida, M. V. et al. GTSF-1 is required for formation of a functional RNA-dependent RNA Polymerase complex in *Caenorhabditis elegans*. *EMBO J.* **37**, e99325 (2018).

77. Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **17**, 148 (2016).

78. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).

79. Chiu, J., March, P. E., Lee, R. & Tillett, D. Site-directed, Ligase-independent mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res.* **32**, e174–e174 (2004).

80. Chiu, J., Tillett, D., Dawes, I. W. & March, P. E. Site-directed, ligase-independent mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. *J. Microbiological Methods* **73**, 195–198 (2008).

81. Dickinson, D. J., Ward, J. D., Reiner, D. J. & Goldstein, B. Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nat. Methods* **10**, 1028–1034 (2013).

82. Arribere, J. A. et al. Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in *Caenorhabditis elegans*. *Genetics* **198**, 837–846 (2014).

83. Dickinson, D. J., Pani, A. M., Heppert, J. K., Higgins, C. D. & Goldstein, B. Streamlined genome engineering with a self-excising drug selection cassette. *Genetics* **200**, 1035–1049 (2015).

84. Paix, A. et al. Scalable and versatile genome editing using linear DNAs with microhomology to Cas9 sites in *Caenorhabditis elegans*. *Genetics* **198**, 1347–1356 (2014).

85. Paix, A., Schmidt, H. & Seydoux, G. Cas9-assisted recombineering in C. elegans: genome editing using in vivo assembly of linear DNAs. *Nucleic Acids Res.* **44**, e128 (2016).

86. Mouridi, S. E. et al. Reliable CRISPR/Cas9 genome engineering in *Caenorhabditis elegans* using a single efficient sgRNA and an easily recognizable phenotype. *G3* **7**, 1429–1437 (2017).

87. Frøkjær-Jensen, C. et al. Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nat. Genet.* **40**, 1375–1383 (2008).

88. Lansdorp, P. M. et al. Heterogeneity in telomere length of human chromosomes. *Hum. Mol. Genet.* **5**, 685–691 (1996).

89. Seo, B. & Lee, J. Observation and quantification of telomere and repetitive sequences using fluorescence in situ hybridization (FISH) with PNA probes in *Caenorhabditis elegans*. *JoVE (Journal of Visualized Experiments)* e54224 https://doi.org/10.3791/54224 (2016).

90. Duerr, J. S. Antibody staining in *C. Elegans* using 'freeze-cracking'. *JoVE (Journal of Visualized Experiments)* e50664 https://doi.org/10.3791/50664 (2013).

91. James, P., Halladay, J. & Craig, E. A. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144**, 1425–1436 (1996).

92. Rodrigues, R. J. C. et al. PETISCO is a novel protein complex required for 21U RNA biogenesis and embryonic viability. *Genes Dev.* **33**, 857–870 (2019).

93. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2019).

94. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518 (2005).

95. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

96. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

97. Stevens, L. et al. Comparative genomics of 10 new Caenorhabditis species. *Evolution Lett.* **3**, 217–236 (2019).

98. Sela, I., Ashkenazy, H., Katoh, K. & Pupko, T. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* **43**, W7–W14 (2015).

99. Hahne, F. & Ivanek, R. in *Statistical Genomics: Methods and Protocols* (eds. Mathé, E. & Davis, S.) 335–351 (Springer, 2016).

100. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLOS Computational Biol.* **9**, e1003118 (2013).

101. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

102. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

103. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

Author contributions

Conceptualization, S.D., M.V.A., D.K., R.F.K., and F.B.; investigation, S.D., M.V.A., E.N., J.S., N.V., C.R.; formal analysis, S.D., M.V.A., E.N., N.V., A.F.-S., A.C.-N., and F.B.; visualization, S.D., M.V.A., E.N., J.S., A.F.-S., and F.B.; writing–original draft, S.D., M.V.A., and F.B.; writing–review & editing, all authors contributed; supervision: H.D.U., R.F.K., and F.B.; project administration: F.B.; funding acquisition, H.D.U., R.F.K., F.B.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-22861-2.

Correspondence and requests for materials should be addressed to F.B.

Peer review information *Nature Communications* thanks Jerome Dejardin, Jan Karlseder and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

SUPPLEMENTARY INFORMATION

The double-stranded DNA-binding proteins TEBP-1 and TEBP-2 form a telomeric complex with POT-1

Sabrina Dietz[1,5], Miguel Vasconcelos Almeida[1,2,5], Emily Nischwitz[1], Jan Schreier[1], Nikenza Viceconte[1], Albert Fradera-Sola[1], Christian Renz[1], Alejandro Ceron-Noriega[1], Helle D. Ulrich[1], Dennis Kappei[3,4], René F. Ketting[1], Falk Butter[1,*]

[1]Institute of Molecular Biology (IMB), Ackermannweg 4, 55128 Mainz, Germany

[2]Present address: Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QN, UK; and Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK.

[3]Cancer Science Institute of Singapore, National University of Singapore, 117599, Singapore

[4]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 117596 Singapore

[5]These two authors contributed equally to this work

*Correspondence: f.butter@imb-mainz.de, +49 6131-39-21570

a Label free quantitation

b Dimethyl labeling

c

d *tebp-1*

e *tebp-2*

f

g
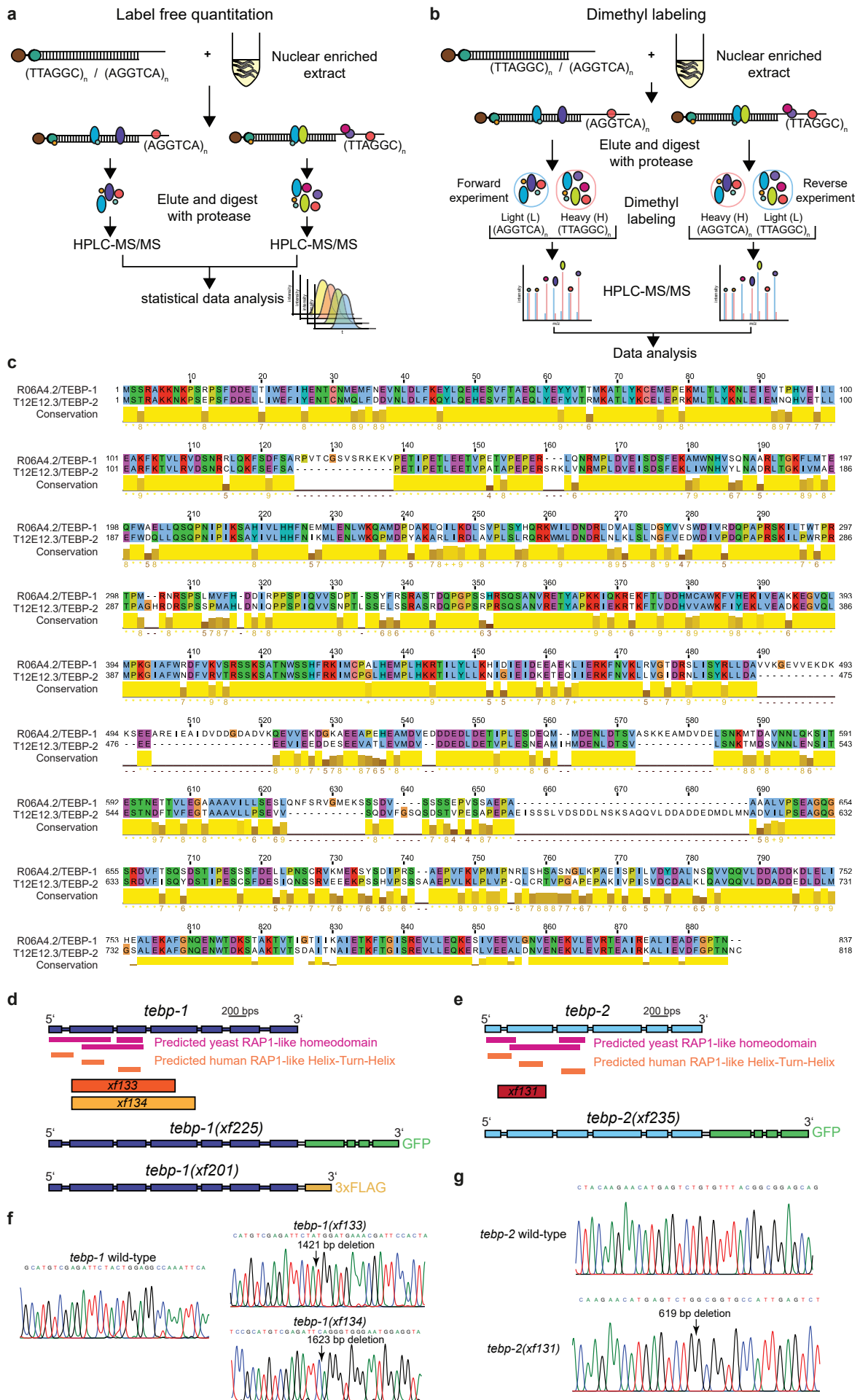
Supplementary Fig. 1. A quantitative proteomics screen for telomere binders identifies the paralogs TEBP-1 and TEBP-2.

(a) Scheme representing the label free quantitation workflow. Telomere $(TTAGGC)_n$, or control DNA $(AGGTCA)_n$ baits are incubated with nuclear extract. Samples are processed and measured independently, and later compared by statistical data analysis.
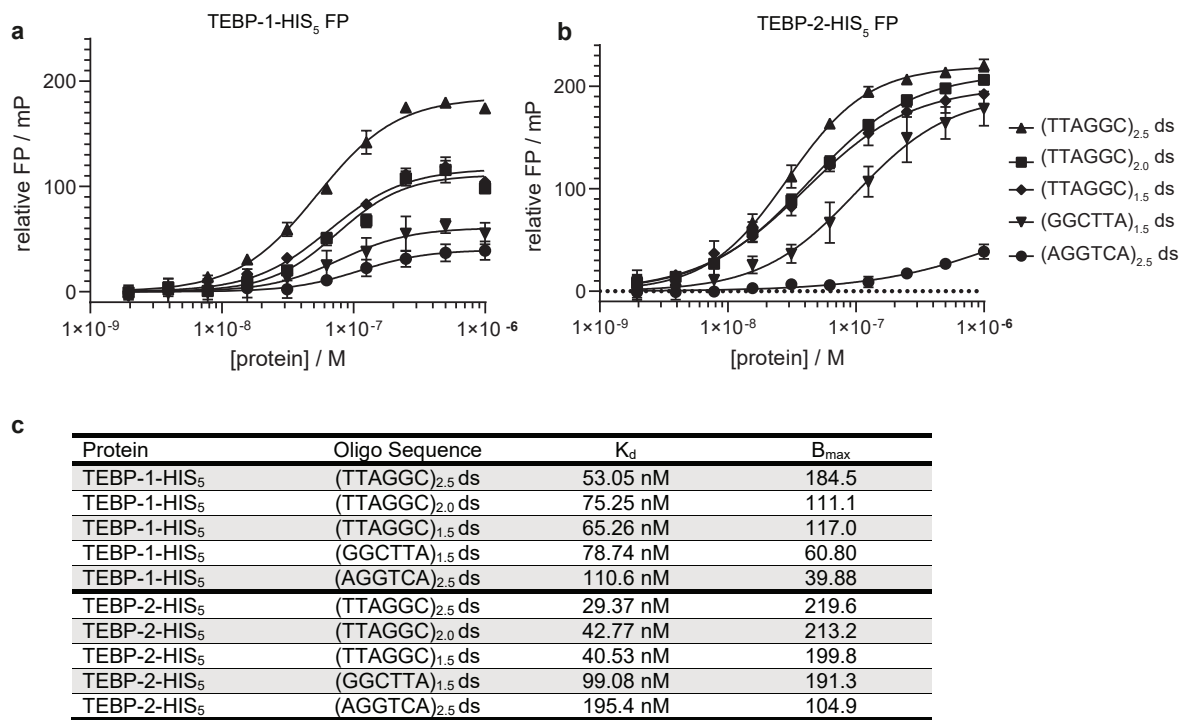
(b) Scheme representing the reductive dimethyl labeling workflow. Telomere $(TTAGGC)_n$, or control DNA $(AGGTCA)_n$ baits are incubated with nuclear extract in duplicates. Per condition each peptide gets labeled with either light methyl groups (CH_3) or heavy methyl groups (CD_3). Afterwards, the heavy sample of one condition is combined with the light sample of the other condition and vice-versa to achieve a forward and a reverse experiment. Forward and reverse experiments are measured and analyzed by comparing intensities of the proteins (calculated from their peptide intensities) in the respective channel.

(c) Pairwise sequence alignment of amino acid sequences of TEBP-1 and TEBP-2 using EMBOSS Needle, visualized using Jalview, showing the high sequence similarity between the two proteins. Amino acids are color coded according to the Clustal X colour scheme: blue – amino acids A, I, L, M, F, W, C and V; red: amino acids R and K; green – amino acids N, S, Q, T; pink – amino acid C; magenta – amino acids E and D; orange – amino acid G; cyan – amino acids H, Y; yellow – amino acid P. Conservation is shown in the yellow bars beneath the sequences, brighter yellow for higher conservation. Amino acid positions are indicated.

(d) Scheme of the *tebp-1* genomic locus. Below are indicated the positions with similarity to the homeodomain of human and yeast RAP1, as predicted by HHPred (3.2.0), deletions made by CRISPR-Cas9 genome editing (alleles *xf133* and *xf134*), as well as the locations of the tags (C-terminal GFP and 3xFLAG), also inserted by CRISPRCas9 genome editing.

(e) As in (d) but for the *tebp-2* locus.

(f-g) Chromatograms of Sanger sequencing of *tebp-1* and *tebp-2* deletion alleles compared to WT. Deletion sites are indicated with arrows. Colors indicate the different DNA bases: black – G; blue – C; red – T; green – A.
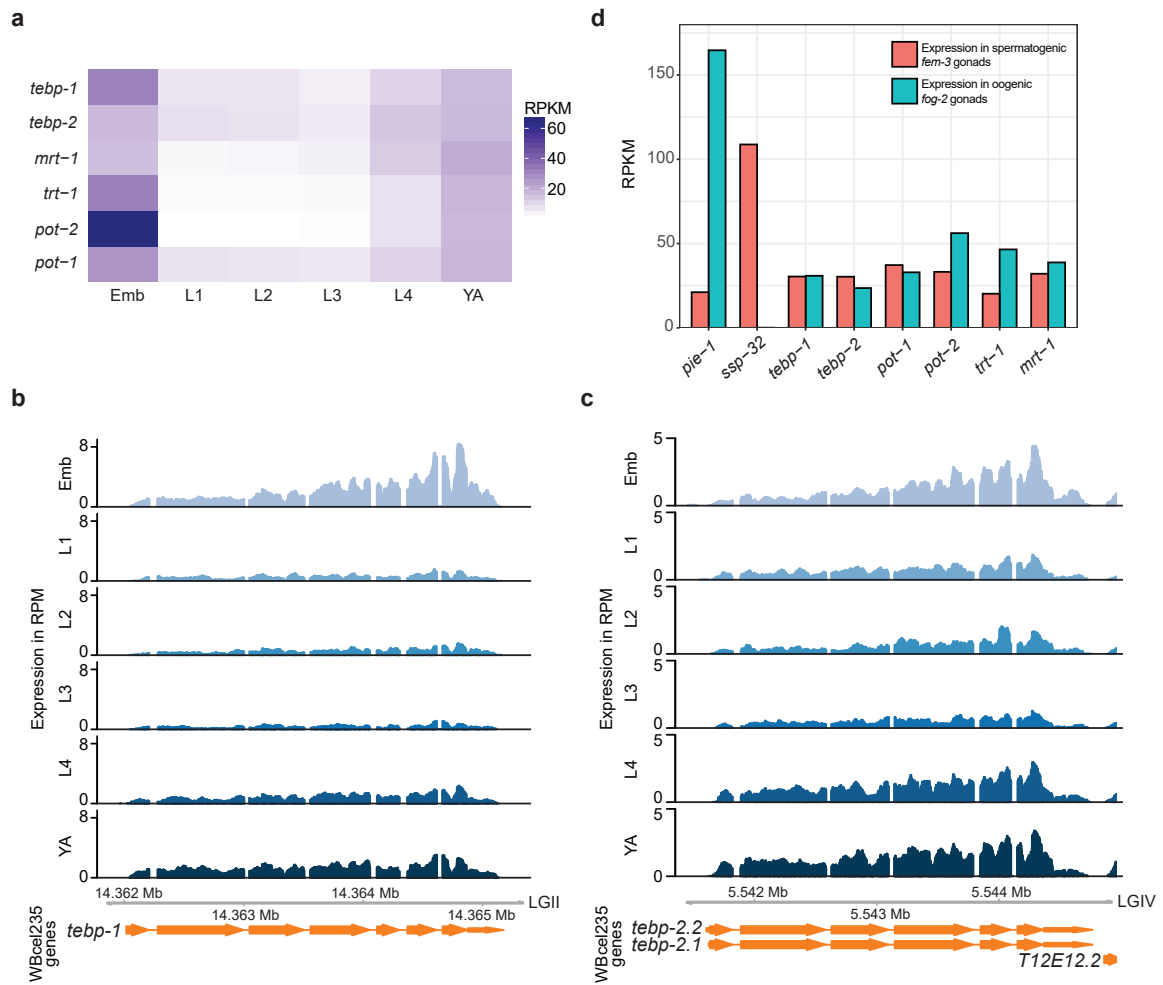
a TEBP-1-HIS$_5$ FP

b TEBP-2-HIS$_5$ FP

- ▲ (TTAGGC)$_{2.5}$ ds
- ■ (TTAGGC)$_{2.0}$ ds
- ♦ (TTAGGC)$_{1.5}$ ds
- ▼ (GGCTTA)$_{1.5}$ ds
- ● (AGGTCA)$_{2.5}$ ds

c

Protein	Oligo Sequence	K_d	B_{max}
TEBP-1-HIS$_5$	(TTAGGC)$_{2.5}$ ds	53.05 nM	184.5
TEBP-1-HIS$_5$	(TTAGGC)$_{2.0}$ ds	75.25 nM	111.1
TEBP-1-HIS$_5$	(TTAGGC)$_{1.5}$ ds	65.26 nM	117.0
TEBP-1-HIS$_5$	(GGCTTA)$_{1.5}$ ds	78.74 nM	60.80
TEBP-1-HIS$_5$	(AGGTCA)$_{2.5}$ ds	110.6 nM	39.88
TEBP-2-HIS$_5$	(TTAGGC)$_{2.5}$ ds	29.37 nM	219.6
TEBP-2-HIS$_5$	(TTAGGC)$_{2.0}$ ds	42.77 nM	213.2
TEBP-2-HIS$_5$	(TTAGGC)$_{1.5}$ ds	40.53 nM	199.8
TEBP-2-HIS$_5$	(GGCTTA)$_{1.5}$ ds	99.08 nM	191.3
TEBP-2-HIS$_5$	(AGGTCA)$_{2.5}$ ds	195.4 nM	104.9

Supplementary Fig. 2

Supplementary Fig. 2. Telomeric double-strand binding preferences of TEBP-1 (R06A4.2) and TEBP-2 (T12E12.3).

(a-b) Fluorescence polarization assays of 1 µM to 2 nM purified TEBP-1-His$_5$ (a) and TEBP-2-His$_5$ (b). Proteins were incubated with 2.5x, 2.0x, 1.5x T-rich, and 1.5x G-rich double-stranded telomeric FITC-labeled oligonucleotides, as well as 2.5x double-stranded control. Error bars represent +/- the standard deviation of the mean values. Per data point n=3 technical replicates. FP, fluorescence polarization; mP, millipolarization, upward triangle: 2.5x TTAGGC double-strand, downward triangle: 2.5x TTAGGC single-strand, square: 2x TTAGGC double-strand, diamond: 1.5x TTAGGC T-rich double-strand, downward triangle: 1.5x G-rich GGCTAA double-strand, circle: 2.5x shuffled control AGGTCA double-strand.

(c) Overview of K$_d$ and B$_{max}$ values from FP experiment (a-b).

a

	Emb	L1	L2	L3	L4	YA
tebp-1						
tebp-2						
mrt−1						
trt−1						
pot−2						
pot−1						

b

c

d

Expression in spermatogenic *fem-3* gonads
Expression in oogenic *fog-2* gonads

Supplementary Fig. 3

57

Supplementary Fig. 3. The expression profiles of *tebp-1* and *tebp-2* throughout development and in isolated gonads.

(a) Heatmap depicting mRNA expression levels, in Reads Per Kilobase Million (RPKM), of the known telomere binders *pot-1*, *pot-2*, and *mrt-1*, telomerase subunit *trt-1*, as well as *tebp-1* and *tebp-2*. Data from a previously published RNA-seq dataset[47].

(b-c) Genome browser tracks with the mRNA expression of *tebp-1* (b), and *tebp-2* (c), in reads per million (RPM), across the different life stages of *C. elegans*. Data from [47]. (a-c) Emb, embryos; L1-L4, first to fourth larval stages; YA, young adults.

(d) Expression of telomere factors in dissected *fem-3* mutant gonads (exclusively spermatogenic) and *fog-2* mutant gonads (exclusively oogenic), from previously published RNA-seq data[48]. *pie-1* and *ssp-32* are genes known to be expressed in oogenesis and in spermatogenesis, respectively, according to [48].

a

RPKM in WT

RPKM in *tebp-1(xf133)*

14.362 Mb   14.363 Mb   14.364 Mb   14.365 Mb   LGII

WBcel235 genes   *tebp-1*

xf133 deletion

b

RPKM in WT

RPKM in *tebp-2(xf131)*

5.542 Mb   5.543 Mb   5.544 Mb   LGIV

WBcel235 genes   *tebp-2.2*   *tebp-2.1*   *T12E12.2*

xf131 deletion

c *tebp-1(xf133)*

MERGE   FISH   DAPI

d *tebp-2(xf131)*

MERGE   FISH   DAPI

e *pot-2(tm1400)*

MERGE   FISH   DAPI

f N2

MERGE   FISH   DAPI

g

Quantification of qFISH signal

Relative integrated density

tebp-2(xf131) (n=6)   N2 (n=10)

0.04

7.8e-04

tebp-2(xf131) (n=6)   N2 (n=10)   *tebp-1(xf133)* (n=6)

Supplementary Fig. 4

Supplementary Fig. 4. TEBP-1 and TEBP-2 regulate telomere length in embryos.

(a-b) Genome browser tracks with the mRNA expression of *tebp-1* (a) and *tebp-2* (b), in Reads Per Kilobase Million (RPKM). RNA-seq data of wild-type, *tebp-1(xf133)*, and *tebp-2(xf131)* mutants.

(c-f) Representative maximum projection z-stacks of a qFISH assay using embryos of *C. elegans* mutant strains. The telomeres of these embryos were visualized by hybridization with a telomeric PNA-FISH-probe. Nuclei were stained with DAPI. Scale bars, 10 μm. *tebp-1(xf133)* and *tebp-2(xf131)* were grown for approx. 98/120 generations before the experiment. N = 3 biologically independent experiments with similar results.

(g) Barplot depicting analysis of qFISH images of the strains in (c-f), as indicated on the x-axis. Average telomere length is indicated by arbitrary units of relative integrated density on the y-axis, with wild-type N2 set to 1. The left hand plot is a zoomed-in inset of the N2 and *tebp-2(xf131)* values. n of analyzed independent embryos per strain: *tebp-2(xf131)*: n=6, N2: n=10, *tebp-1(xf131)*: n=6. Error bars represent the standard error of the mean (SEM) and p-values were calculated using Welch's t-test.

a

Genotype	Father	Mother	Synthetic sterile F2	Could grow double mutant homozygous line
tebp-2(xf131); tebp-1(xf133)	*tebp-2(xf131)*	*tebp-1(xf133)*	Yes	No
tebp-1(xf133); tebp-2(xf131); xfIs148(tebp-2::gfp MosSCI)	*tebp-1(xf133)*	*tebp-2(xf131); xfIs148(tebp-2::gfp MosSCI)*	No	Yes
tebp-2(xf131); tebp-1(xf133)	*tebp-1(xf133)*	*tebp-2(xf131); xfIs148(tebp-2::gfp MosSCI)*	Yes	No
tebp-2(xf131); pot-2(tm1400)	*tebp-2(xf131)*	*pot-2(tm1400)*	No	Yes
tebp-1(xf133); trt-1(ok410)	*tebp-1(xf133)*	*trt-1(ok410)*	No	Yes
tebp-1(xf133); mrt-1(tm1354)	*tebp-1(xf133)*	*mrt-1(tm1354)*	No	Yes
pot-2(tm1400); trt-1(ok410)	*pot-2(tm1400)*	*trt-1(ok410)*	No	Yes
tebp-2(xf131); trt-1(ok410)	*tebp-2(xf131)*	*trt-1(ok410)*	No	Yes
tebp-1(xf260); pot-2(tm1400)	N/A. Used CRISPR-Cas9 to introduce *tebp-1* mutation due to linkage	*pot-2(tm1400)*	N/A not a cross	Yes

b
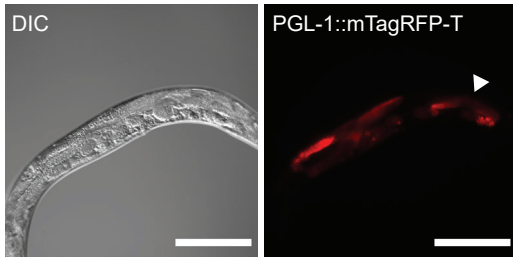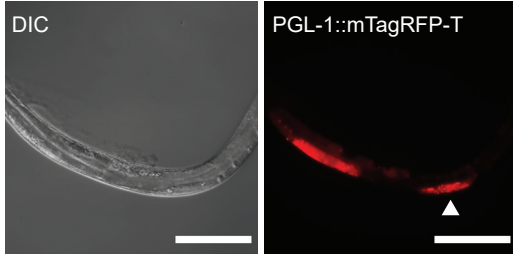
Category 2: one gonad arm atrophied   Category 3: both gonad arms atrophied

c

d   20°C

e   25°C

Supplementary Fig. 5

61
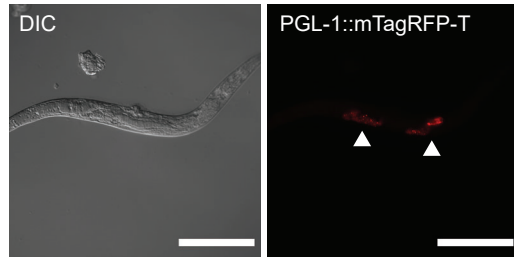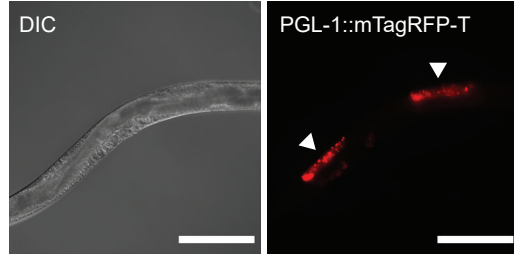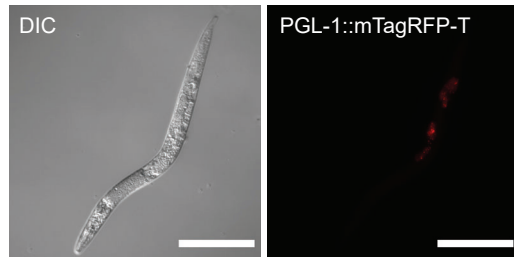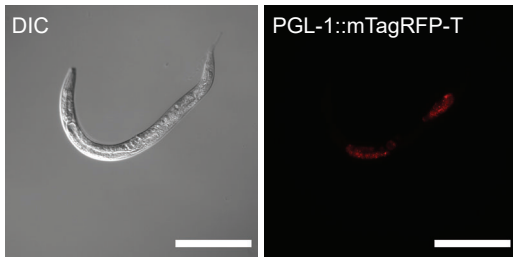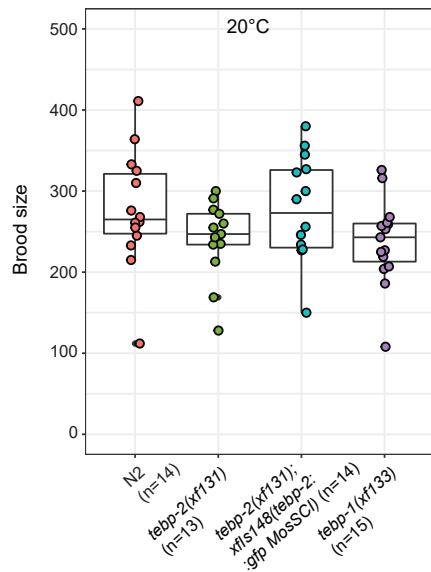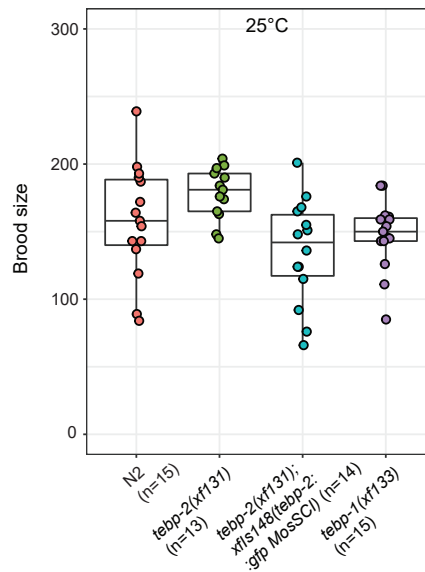
Supplementary Fig. 5. Dissecting the role of TEBP-1 and TEBP-2 in fertility.

(a) Overview of additional crosses performed to investigate distinct aspects of the synthetic sterility phenotype. For each cross, the columns indicate the genotype of the animals analyzed, the genotype of their parents, whether the animals have synthetic sterility, and if we could establish a homozygous line. The second row shows that the reciprocal cross between *tebp-1* and *tebp-2* also led to synthetic sterility. The third row shows that a *tebp-2::gfp* single-copy transgene rescues the synthetic sterility of *tebp-1; tebp-2* double mutants, while their transgene-less siblings still display synthetic sterility (fourth row). The following rows demonstrate that the synthetic sterility is specific to *tebp-1* and *tebp-2*, as it does not arise in crosses with other telomere-associated mutants.

(b) Additional representative widefield DIC and fluorescence pictures of worms with germlines of categories 2 (left panels) and 3 (right panels). Scale bars, 200 µm. Atrophied germlines are indicated with white arrowheads.

(c) Exemplary widefield DIC and fluorescence micrographs of worms showing growth defects and/or larval arrest. These animals were isolated concurrently to animals shown in (b), but did not reach adulthood. These two specific animals were offspring of *tebp-2(xf131); tebp-1(xf133)* +/-. Scale bars, 200 µm.

(d-e) Boxplots showing the brood sizes of wild-type N2, *tebp-1* or *tebp-2* single mutants, and *tebp-2(xf131); xfIs148(tebp-2::gfp)*. Central horizontal lines represent the median, the bottom and top of the box represent the 25th and 75th percentile, respectively. Whiskers represent the 5th and 95th percentile, dots represent the data points used to calculate the box plot. Experiments were carried out at 20°C (d) and 25°C (e). Statistical comparisons were performed with wildtype N2, calculated with two-sided and unpaired Mann–Whitney and Wilcoxon tests. N2 vs.*tebp-2(xf131)*: 20°C p-value=0.145, 25°C p-value= 0.097; N2 vs. *tebp-2(xf131);xfIs148(tebp-2::gfp MosSCI)*: 20°C p-value=0.91, 25°C p-value=0.183; N2 vs. *tebp-1(xf133)*: 20°C p-value=0.052, 25°C p-value=0.41. Analyzed individuals per strain are indicated as n on the x-axis labels.

a

Fractions of TEBP-2::GFP; TEBP-1::3xFLAG embryos

← Approximate MW of fraction

α-GFP
α-FLAG
α-GTSF-1

Fractions: 1   5   10   15   20

b α-FLAG TEBP-1 IP in young adults

c α-GFP TEBP-2 IP in young adults

d

e α-FLAG TEBP-1 IP in embryos (+Sm Nuclease)

f α-GFP TEBP-2 IP in embryos (+Sm Nuclease)

g

h

i

Supplementary Fig. 6

63
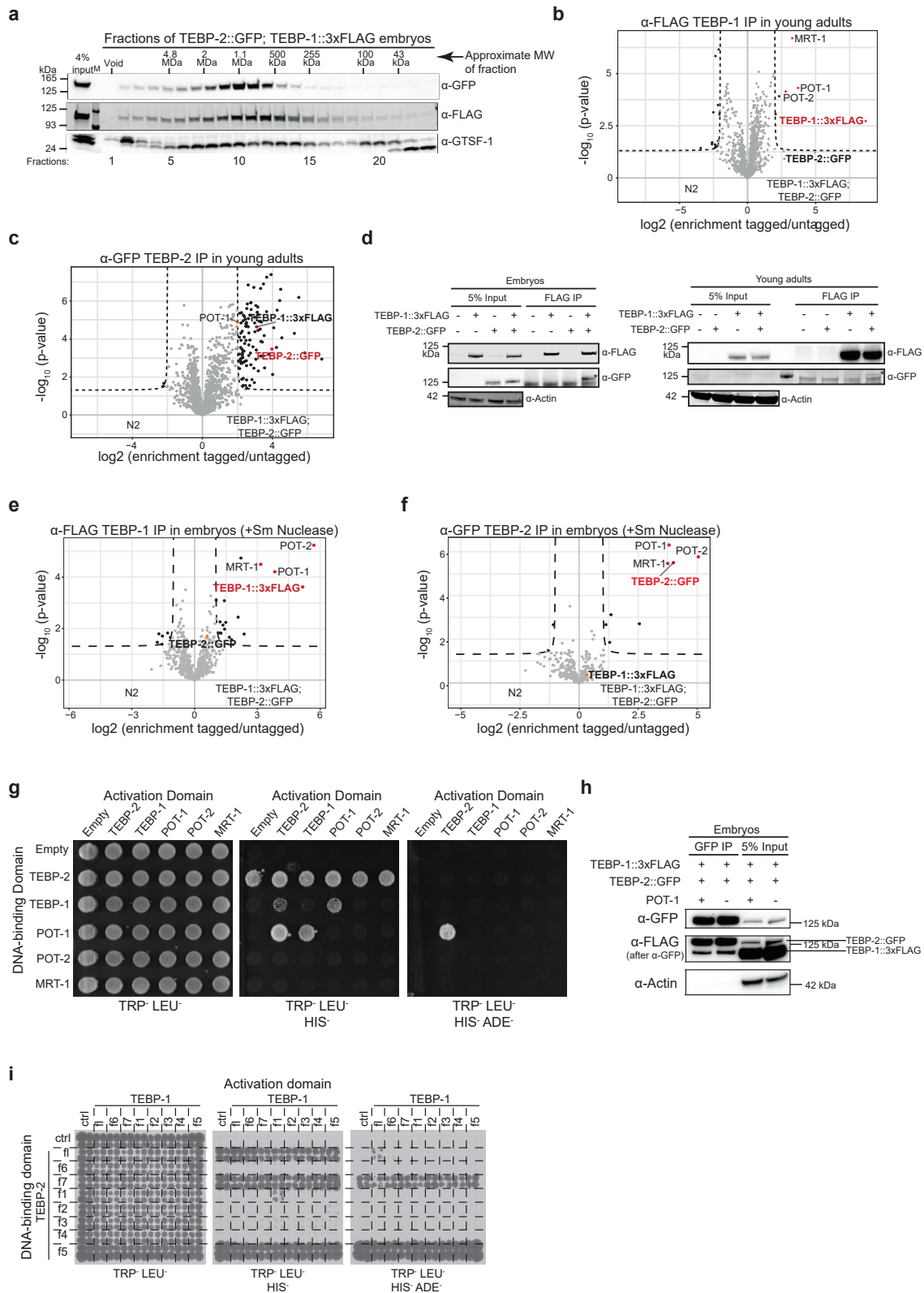
Supplementary Fig. 6. TEBP-1 and TEBP-2 interact with each other and with POT-1/MRT-1/POT-2.

(a) Western blot of the eluted fractions from size-exclusion chromatography of embryo extracts containing TEBP-1::3xFLAG and TEBP-2::GFP. The approximate molecular weight (MW) of the fractions is indicated above the blots. GTSF-1 was used as a control, as it has a known elution profile in size-exclusion chromatography[76]. Information about α-GTSF-1 can be found in [76]. N = 2 biologically independent experiments with similar results.

(b-c) Volcano plots with quantitative proteomic analysis of TEBP-1::3xFLAG (b) or TEBP-2::GFP (c) IPs in young adults. IPs were performed in quadruplicates. Enriched proteins (threshold: 4-fold, p-value<0.05) are shown as black dots, enriched proteins of interest are highlighted with red or orange dots, and the baits are named in red. Background proteins are depicted as grey dots.

(d) Co-IP western blot experiment of TEBP-1::3xFLAG and TEBP-2::GFP similar to Fig. 5e-f, except the IPs were performed with an α-FLAG antibody. Actin was used as loading control. IPs with embryo extracts in the left panel and with young adult extracts in the right panel. N = 3 biologically independent experiments with similar results for both experiments.

(e-f) Volcano plots showing quantitative proteomic analysis of either TEBP-1::3xFLAG (e) or TEBP-2::GFP (f) IPs in embryos. IPs were performed in quadruplicates and Sm nuclease was added to remove potential DNA-dependent interactions. Enriched proteins (threshold >2-fold, p-value<0.05) are shown as black dots. Enriched proteins of interest are highlighted with red or orange dots, and baits are named in red.

(g) Orthogonal grid of the Y2H spotting containing fusion constructs of the Gal4 activating or DNA-binding domains with the full length sequence of telomere factors. Left panel shows growth control in non-restrictive medium. Protein-protein interactions allow for growth on TRP- LEU- HIS- medium (middle panel). TEBP-2 bound to the Gal4 DNA-binding domain is self-activating, precluding the determination of interactions. The strongest interactions are permissive of growth on the highly stringent TRP- LEU- HIS- ADE- medium (right panel).

(h) Co-IP western blot experiments of TEBP-1::3xFLAG and TEBP-2::GFP in the presence and absence of POT-1, where absence of POT-1 refers to the *pot-1(tm1620)* mutation. The IPs were performed with an α-GFP antibody. Actin was used as loading control. IPs were performed with 800 µg of embryo extracts. Detection by ECL was performed sequentially, first for GFP and then for FLAG.

(i) Y2H spotting as in (g) with TEBP-1 and TEBP-2 partial constructs fused to the GAL4 activation or DNA-binding domain as in Fig. 6d,h. The full length, f7, and f5 TEBP-2 constructs fused to the Gal4 DNA-binding domain show self-activation. As in (g) the growth on the highly stringent TRP-LEU-HIS-ADE-medium (right panel) indicates the strongest interactions.

64
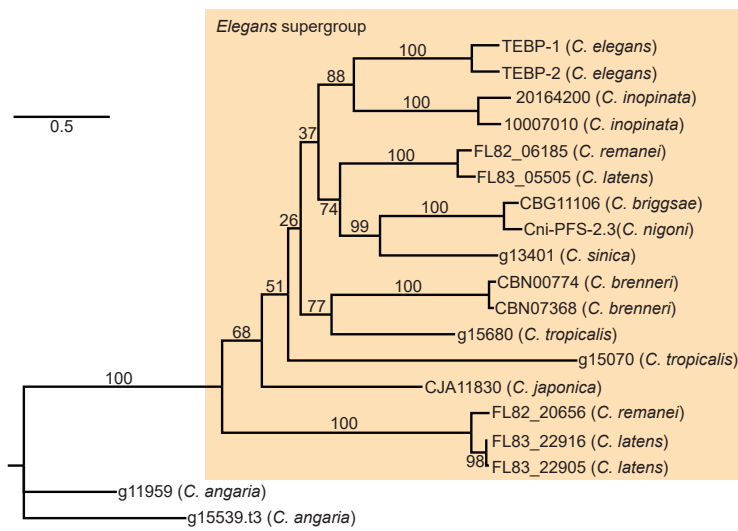
Supplementary Fig. 7

Supplementary Fig. 7. Phylogenetic analysis of the N-terminal region of TEBP-proteins.

Phylogenetic tree constructed as in Fig. 7a. The MAFFT protein alignment used for this tree comprised the first 600 alignment positions of the multiple sequence alignment in Supplementary Data 2 (sheet 2). Values on the nodes represent bootstrapping values of 10000 replicates, set to 100.

Supplementary Table 1: List of strains used and created in this study.

Listed are all strains with their respective genotype and source.

Strain Reference	Genotype	Source
	Wild-Type N2	CGC
YA1197	*ypIn2 [daz-1p::pot-1::mCherry::tbb-2 3'UTR + Cbr-unc-119(+)] II.*	A kind gift from Shawn Ahmed
tm1620	*pot-1(tm1620) III.*	National Bioresource Project for the nematode, Japan
tm1400	*pot-2(tm1400) II.*	National Bioresource Project for the nematode, Japan
YA1116	*mrt-1(tm1354) I.*	CGC
YA1059	*trt-1(ok410) I.*	CGC
EG6699	*ttTi5605 II; unc-119(ed3) III; oxEx1578*	CGC
RFK641	*tebp-2(xf131) IV.*	This study
RFK671	*tebp-1(xf133) II.*	This study
RFK672	*tebp-1(xf134) II.*	This study
RFK659	*TEBP-2(xfIs148[tebp-2(prm)::tebp-2::GFP::tebp-2(3'UTR)]) II; unc-119(ed9) III.*	This study
RFK1096	*tebp-2(xf235[TEBP-2::GFP]) IV.*	This study
RFK1022	*tebp-1(xf225[tebp-1::GFP]) II.*	This study
RFK958	*tebp-1(xf201[tebp-1::3xFLAG]) II.*	This study
RFK1173	*tebp-2(xf235[tebp-2::GFP]) IV; tebp-1(xf201[tebp-1::3xFLAG]) II.*	This study
RFK1174	*tebp-2(xf235[tebp-2::GFP]) IV; ypIn2[daz-1p::pot-1::mCherry::tbb-2 3'UTR + Cbr-unc-119(+)] II.*	This study
RFK1067	*tebp-1(xf225[tebp-1::GFP]) II; ypIn2[daz-1p::pot-1::mCherry::tbb-2 3'UTR + Cbr-unc-119(+)] II.*	This study
RFK1086	*pgl-1(xf233[pgl-1::mTagRFP-T]) IV.*	Jan Schreier, Ketting laboratory
RFK1327	*tebp-2(xf131) IV; pgl-1(xf233[pgl-1::mTagRFP-T]) IV.*	This study
RFK1328	*tebp-1(xf133) II; pgl-1(xf233[pgl-1::mTagRFP-T]) IV.*	This study
-	*tebp-2(xf131) IV; pot-2(tm1400) II.*	This study
-	*tebp-1(xf133) II; mrt-1(tm1354) I.*	This study
RFK1334	*trt-1(ok410) I; tebp-1(xf133) II.*	This study
RFK1309	*tebp-1(xf260) II; pot-2(tm1400) II.*	This study
-	*trt-1(ok410) I; pot-2(tm1400) II.*	This study
AF16	*C. briggsae* Wild-type	CGC

Supplementary Table 2. Fractions of the gel filtration runs and correlated molecular weight.

Separation range of the used column in red, fractions covered by the marker run in green. Fractions of the 96-well column marked in bold were concentrated and used for western blot detection (Figs. 5a and S6a respectively). MW: molecular weight.

Fraction	volume [ml]	log MW	calculated MW [kDa]	96 well
A1	1,0	8,982	960063,591	a1
A2	2,0	8,727	533212,105	a2
A3	3,0	8,472	296141,997	a3
A4	4,0	8,216	164475,040	a4
A5	5,0	7,961	91348,201	a5
A6	6,0	7,705	50734,105	a6
A7	6,5	7,578	37809,419	a7
A8	7,0	7,450	28177,340	**a8**
A9	7,5	7,322	20999,067	a9
A10	8,0	7,195	15649,483	**a10**
A11	8,5	7,067	11662,724	a11
A12	9,0	6,939	8691,605	**a12**
A13	9,5	6,811	6477,389	**b12**
A14	10,0	6,684	4827,252	**b11**
A15	10,5	6,556	3597,493	**b10**
A16	11,0	6,428	2681,020	**b9**
A17	11,5	6,301	1998,021	**b8**
A18	12,0	6,173	1489,018	**b7**
A19	12,5	6,045	1109,686	**b6**
A20	13,0	5,918	826,990	**b5**
A21	13,5	5,790	616,311	**b4**
A22	14,0	5,662	459,304	**b3**
A23	14,5	5,534	342,295	**b2**
A24	15,0	5,407	255,094	**b1**
A25	15,5	5,279	190,108	**c1**
A26	16,0	5,151	141,677	**c2**
A27	16,5	5,024	105,584	**c3**
A28	17,0	4,896	78,686	**c4**
A29	17,5	4,768	58,641	**c5**
A30	18,0	4,641	43,702	**c6**
A31	18,5	4,513	32,569	**c7**
A32	19,0	4,385	24,272	**c8**
A33	19,5	4,257	18,088	c9
A34	20,0	4,130	13,480	c10
A35	20,5	4,002	10,046	c11
A36	21,0	3,874	7,487	c12
A37	21,5	3,747	5,580	d12
A38	22,0	3,619	4,158	d11
A39	22,5	3,491	3,099	d10
A40	23,0	3,364	2,309	d9
A41	23,5	3,236	1,721	d8
A42	24,0	3,108	1,283	d7

Superose 6 column separation range (5-5000 kDa)

covered by marker run

DNA damage repair proteins across the tree of life

Summary

Here, we studied DNA damage recognition and repair across the tree of life (Nischwitz et al. 2023). Specifically, we investigated the 8-oxoguanine (8-oxoG) and an abasic lesions and uracil incorporated into DNA. These lesions are all repaired by excision repair pathways. 8-oxoG and abasic sites are DNA damage lesions which are specifically repaired by base excision repair (BER). 8-oxoG is formed with oxidative or alkylative damage and the abasic lesion can either occur as an independent lesion or a BER intermediate. The uracil incorporated base is primarily caused by improper DNA replication and is often repaired by ribonucleotide excision repair. These three DNA damage lesions were incorporated into synthetic oligos, and with an undamaged oligo quadruplicates of immunoprecipitations were performed with protein extracts from 11 different species. There were 337 proteins enriched over the undamaged oligo, 99 of which had been previously implicated in DNA repair. Intriguingly, of these 99 proteins there were high levels of unanticipated crosstalk between all three of the lesions, one exceptional example being that of the AP endonuclease protein family. Additionally, the enriched repair proteins held two unexpected groups of homologs, the photolyases and MUTYH-related glycosylases. While these are repair-related proteins, the particular lesion conditions have not been previously shown to involve these proteins. Additionally, we identified through a combination of orthology, network, and domain analysis 44 previously unconnected proteins to DNA repair.

Altogether, we provide a large dataset which can be used to study the conservation and divergence of DNA repair as well as propel new discoveries within DNA repair.

Zusammenfassung

Hier haben wir die Erkennung und Reparatur von DNA-Schäden im gesamten Baum des Lebens untersucht (Nischwitz et al. 2023). Insbesondere untersuchten wir das 8-Oxoguanin (8-oxoG) und eine abasische Läsion sowie den Einbau von Uracil in die DNA. Diese Läsionen werden alle

durch Exzisionsreparaturwege repariert. 8-OxoG- und abasische Stellen sind DNA-Schadensläsionen, die spezifisch durch Basenexzisionsreparatur (BER) repariert werden. 8-oxoG wird bei oxidativer oder alkylativer Schädigung gebildet und die abasische Läsion kann entweder als eigenständige Läsion oder als BER-Zwischenprodukt auftreten. Die eingebaute Uracil-Base wird hauptsächlich durch eine fehlerhafte DNA-Replikation verursacht und wird häufig durch Ribonukleotid-Exzisionsreparatur repariert. Diese drei DNA-Schadensläsionen wurden in synthetische Oligos eingebaut, und mit einem unbeschädigten Oligo wurden in vierfacher Ausführung Immunpräzipitationen mit Proteinextrakten aus 11 verschiedenen Arten durchgeführt. Mit dem unbeschädigten Oligo wurden 337 Proteine angereichert, von denen 99 bereits zuvor mit DNA-Reparatur in Verbindung gebracht worden waren. Interessanterweise gab es bei diesen 99 Proteinen ein hohes Maß an unerwarteten Überschneidungen zwischen allen drei Läsionen, ein außergewöhnliches Beispiel ist die Familie der AP-Endonuklease-Proteine. Darüber hinaus enthielten die angereicherten Reparaturproteine zwei unerwartete Gruppen von Homologen, die Photolyasen und MUTYH-verwandten Glykosylasen. Obwohl es sich hierbei um reparaturrelevante Proteine handelt, wurde bisher nicht nachgewiesen, dass diese Proteine an den jeweiligen Läsionsbedingungen beteiligt sind. Darüber hinaus identifizierten wir durch eine Kombination aus Orthologie-, Netzwerk- und Domänenanalyse 44 zuvor nicht mit der DNA-Reparatur in Verbindung stehende Proteine.

Insgesamt stellen wir einen großen Datensatz zur Verfügung, der zur Untersuchung der Erhaltung und Divergenz der DNA-Reparatur sowie zur Förderung neuer Entdeckungen im Bereich der DNA-Reparatur verwendet werden kann.

Statement of contribution

This project was led by myself and Vivien Schoonenberg, with the support of Falk Butter and Marion Scheibe. Vivien, Marion, Falk and I contributed to the experimental design and implementation. Vivien and I led the specific direction of the data analysis, data visualization, and provided in depth data interpretation. Albert Fradera Sola, Mario Dejung, and Michal Levin conducted bioinformatic data analysis. Olga Vydzhak and Brian Luke provided experimental support and offered critical feedback on the manuscript. Vivien and I wrote the initial draft and finalized version of the manuscript with support from Falk Butter and Marion Scheibe. All authors gave feedback on the final manuscript feedback.

Supervisor Confirmation: ____ *Falk Butter* _____

iScience

Article

DNA damage repair proteins across the Tree of Life
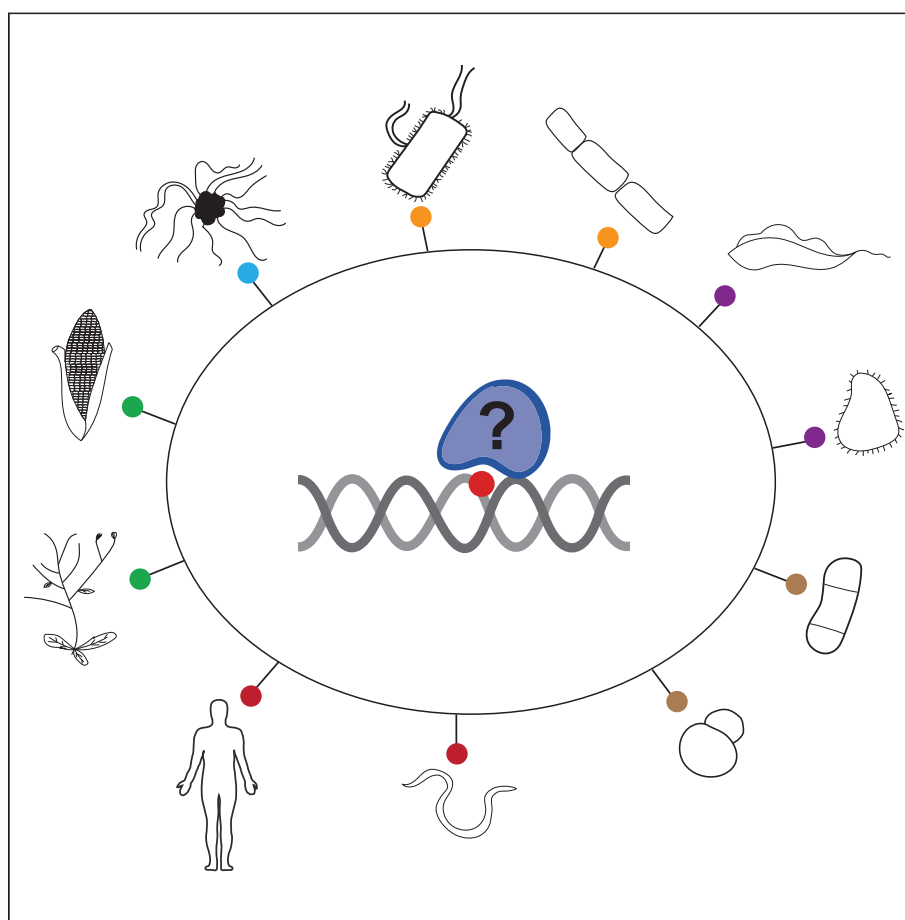
Emily Nischwitz, Vivien A.C. Schoonenberg, Albert Fradera-Sola, ..., Brian Luke, Falk Butter, Marion Scheibe

f.butter@imb.de (F.B.)
m.scheibe@imb.de (M.S.)

Highlights

Phylointeractomics resource of 3 DNA damage lesions, across 11 species

337 binding proteins identified with 99 known DNA damage repair factors

Linked 44 previously unconnected proteins to DNA damage repair

Resource for future study of crosstalk and evolutionary conservation

Article

DNA damage repair proteins across the Tree of Life

Emily Nischwitz,[1,3] Vivien A.C. Schoonenberg,[1,3] Albert Fradera-Sola,[1] Mario Dejung,[1] Olga Vydzhak,[2] Michal Levin,[1] Brian Luke,[1,2] Falk Butter,[1,4,*] and Marion Scheibe[1,*]

SUMMARY

Genome maintenance is orchestrated by a highly regulated DNA damage response with specific DNA repair pathways. Here, we investigate the phylogenetic diversity in the recognition and repair of three well-established DNA lesions, primarily repaired by base excision repair (BER) and ribonucleotide excision repair (RER): (1) 8-oxoguanine, (2) abasic site, and (3) incorporated ribonucleotide in DNA in 11 species: *Escherichia coli, Bacillus subtilis, Halobacterium salinarum, Trypanosoma brucei, Tetrahymena thermophila, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Caenorhabditis elegans, Homo sapiens, Arabidopsis thaliana,* and *Zea mays.* Using quantitative mass spectrometry, we identified 337 binding proteins across these species. Of these proteins, 99 were previously characterized to be involved in DNA repair. Through orthology, network, and domain analysis, we linked 44 previously unconnected proteins to DNA repair. Our study presents a resource for future study of the crosstalk and evolutionary conservation of DNA damage repair across all domains of life.

INTRODUCTION

The stability of the genome is constantly threatened by both exogenous and endogenous mutagens. These genotoxic stressors can damage the architecture of the DNA, causing single-stranded breaks, double-stranded breaks, or chemical modifications to individual bases. These alterations may prevent the successful storage of genetic information and its transmission from one generation to the next and may potentially affect cellular fitness. To maintain genome integrity, there is a carefully orchestrated DNA damage response that functions to identify and subsequently repair damaged DNA.[1] Base excision repair (BER) and ribonucleotide excision repair (RER) represent two pathways that are responsible for resolving some of the most frequently encountered DNA lesions.

BER is primarily responsible for removing nonhelix-distorting lesions.[2] Some of the most prevalent lesions removed via the BER pathway are alkylated or oxidized bases and uracil misincorporation. The most frequent oxidative base lesion is 7,8-dihydro-8-oxoguanine (8-oxoG/8-oxoGuanine), which has been reported to occur up to 1,500 times per mammalian cell per day.[3] There is strong conservation of the BER pathway in archaea, protozoa, fungi, metazoa, and plantae.[4–8] In higher eukaryotes, the repair process generally begins with damage recognition by a DNA glycosylase, which then removes the damaged base and creates an apurinic/apyrimidinic site (AP site/abasic site). Abasic sites can be formed not only as BER intermediates but also endogenously. It has been estimated that there are up to 10,000 abasic sites arising per day in a single mammalian cell.[9] When abasic sites are generated, a 5′-cleavage event is typically triggered by an AP endonuclease, resulting in a 3′-hydroxyl and 5′-deoxyribose phosphate. In single-nucleotide repair, the 5′-deoxyribose is removed primarily by DNA polymerase β and in some cases by DNA polymerase γ, and the resulting gap is then filled. If two or more nucleotides are repaired, the 3′-hydroxyl is used for strand displacement synthesis via either DNA polymerase β or δ and ε, usually in conjunction with PCNA (Proliferating Cell Nuclear Antigen).[10] The previously cleaved 5′-deoxyribose strand, often referred to as a 5′-flap, is removed by FEN1. In both instances, the nick is sealed with ligase I or III.[7]

Even more common than the generation of abasic sites is ribonucleotide misincorporation into double-stranded DNA during DNA replication. This occurs at a rate of one million sites per genome in mammalian cells, rendering it the most common endogenous DNA damage.[11] DNA polymerases have a highly conserved amino acid pocket that enforces sugar selectivity, referred to as a steric gate. While this steric gate helps polymerases prevent the entry of ribonucleotide triphosphates (rNTPs), there is still a large
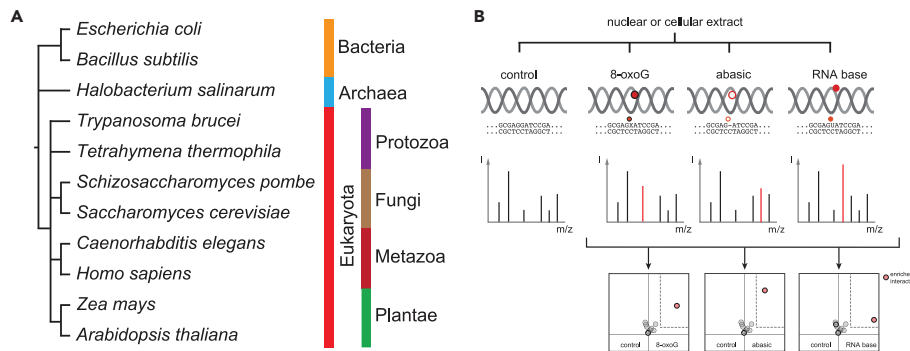
Figure 1. Overview of screen for proteins interacting with DNA damage marks
(A) Phylogenetic tree and overview of the eleven species included in this study.
(B) Experimental setup of the interactomics screen. Pull downs were performed for a control, and for an 8-oxoG, abasic, and RNA base lesion. Pull downs of the respective DNA damage lesion were compared to the common control to calculate enriched interaction partners passing a fold change threshold > 2 with Welch t-test p values < 0.05 (dashed gray line).

rate of ribonucleoside incorporation into DNA due to the imbalance of the nucleotide pools. For example, in *Saccharomyces cerevisiae,* there are 30- to 200-fold more ribonucleotides than nucleotides.[12] The *S. cerevisiae* replicative polymerases α, δ, and ε add approximately 1,900, 2,200, and 9,600 ribonucleotides per round of replication, respectively.[13] Across different organisms, there is a variable bias within the type of ribonucleotides incorporated into DNA. In this study, we selected rU, which in *S. cerevisiae* and *Schizosaccharomyces pombe* has comparable incorporation rates to rC and rA in nuclear genomes[14] but has thus far been studied less. When misincorporated ribonucleoside monophosphate (rNMP), also known as DNA-incorporated rNTPs, are integrated into DNA, they are most frequently repaired by RNase H2-mediated RER. RNase H2 recognizes the rNMP and incises at the 5′-side of the ribonucleoside, leaving a 3′-hydroxyl and 5′-phosphate. As in BER, the 3′-hydroxyl is used for strand displacement DNA synthesis via either DNA polymerase δ supported by PCNA or by DNA polymerase ε. The flap that is formed, beginning with the 5′-phosphate, is removed by FEN1 or EXO1, after which the repaired strand is ligated.[15,16]

Previously, we used a phylointeractomic screen to study the evolution of proteins binding telomeres across the vertebrate lineage.[17] Here, we revisit this concept, investigating the phylogenetic diversity in the recognition and repair of three well-established DNA lesions, primarily repaired by BER or RER: (1) 8-oxoguanine, (2) an abasic site, and (3) incorporated ribonucleotide in DNA. Previous literature has highlighted strong conservation among fundamental proteins in both of these pathways.[18] However, only by studying these pathways across the Tree of Life can the conservation and divergence of these different repair machinery be elucidated. Including organisms across all three domains of life, this study recapitulates previous findings and reveals new candidate proteins with the potential to be involved in DNA damage repair. We provide a large resource dataset that can be used to propel new discoveries within these specific DNA repair pathways and model organisms.

RESULTS AND DISCUSSION

Wide-scale identification of proteins interacting with DNA damage marks

In this study, we selected 11 species from a broad phylogenetic range encompassing all three domains of life: *Escherichia coli* and *Bacillus subtilis* (bacteria); *Halobacterium salinarum* (archaea); *Trypanosoma brucei* and *Tetrahymena thermophila* (eukaryota, protists); *S. pombe* and *S. cerevisiae* (eukaryota, fungi); *Caenorhabditis elegans* and *Homo sapiens* (eukaryota, metazoa); *Zea mays* and *Arabidopsis thaliana* (eukaryota, plantae) (Figure 1A). We used oligonucleotides that were 79 bases long with three different site-specific synthesized DNA alterations, to which a biotinylated counterstrand was annealed (Table S1). These double-stranded nucleic acid baits were immobilized on paramagnetic streptavidin beads and then incubated with protein lysates from the different species. Bound proteins were eluted from the beads and prepared for mass spectrometry measurements on a high-resolution orbitrap platform (Figure 1B). We quantified between 1,357 and 3,615 protein groups per species (Figure S1A). The replicates
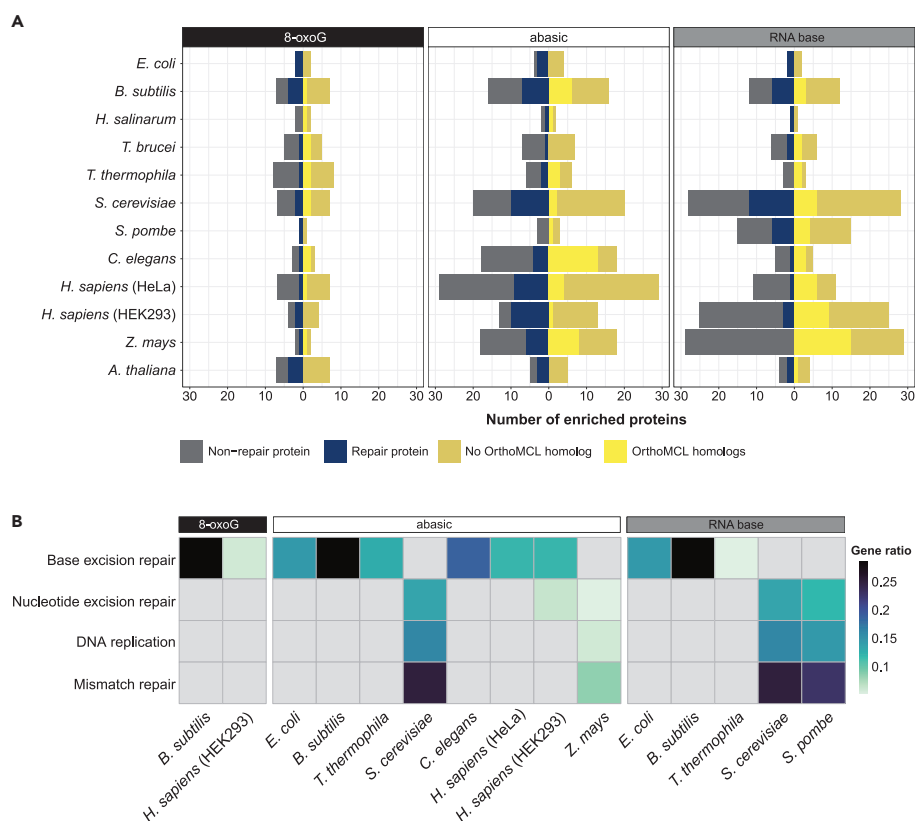
A

B

Figure 2. Interactors of the DNA damage lesions per species

(A) Number of proteins enriched at each lesion in each species highlighted for Gene Ontology annotation "DNA repair" (GO:0006281) (blue) and presence of orthologs in OrthoMCL (yellow).

(B) KEGG term overrepresentation of enriched proteins at each lesion across species. Conditions with no enriched KEGG terms are not shown or presented in gray. Gene ratio refers to genes in the dataset (enriched proteins at lesion) over genes in the background (whole genome).

of every single experiment showed good technical reproducibility covering a similar range of protein intensities (Figure S1B). Each of the three DNA lesions, 8-oxoG, abasic, and RNA, was compared to a common nonmodified oligonucleotide with four replicates per condition to allow the calculation of an average enrichment value (fold change) and a p value for the reproducibility of the enrichment (Welch t-test) (Figure 1). Those proteins that had a $\log_2$ fold change > 2 and a p value < 0.05 were considered enriched. Overall, we enriched 337 proteins across all lesions and species.

Functional enrichment and network analysis reveal novel insights into the enriched interactors

We classified the 337 enriched proteins as either "DNA repair" or "non-DNA repair" using the Gene Ontology (GO) term GO:0006281 (Figure 2A). Of the 337 proteins, 99 were related to DNA repair, and 13 proteins were orthologs of DNA repair proteins (Figure 2A and Table 1, proteins with asterisks). Thus, our experimental conditions allowed for the identification of both known direct and indirect binders to the DNA damage lesions. Next, we used OrthoMCL to trace protein orthologies between species (Tables S2, S3, and S4).[19] The orthology group predictions are based on sequence similarity (reciprocal BLAST) and normalization of interspecies differences, followed by Markov clustering. In total, the OrthoMCL database contains 70,388 ortholog groups across more than 55 species.[20] Proteins detected in our DNA damage interactome screen across eleven species belonged to 10,329 of these groups. We identified 82 proteins that possessed no OrthoMCL orthology with the other 10 species included within

Table 1. Overview of enriched interactors of each DNA damage lesion, per species (fold change > 2, Welch t-test p value < 0.05)

Species	8-oxoG	abasic	RNA base
E. coli	**mutY**, **phrB**	fadJ, **nfo**, **phrB**, polA	**nfo**, polA
B. subtilis	**exoA**, **mutY**, **nfo**, *ydaT*, yhaZ, yisX, **yxlJ**	dinG[a], *disA*, **exoA**, hupA, **mutM**, **nfo**, *parC*, parE, **priA**, topB[a], *ydaT*, *ydeI*, *yfjM*, *yhaZ*, *yqxK*, **yxlJ**	dinG[a], **exoA**, **mutM**, **nfo**, topB[a], *ydcG*, *ydeI*, *yfjM*, *yhaZ*, yisX, yusI, **yxlJ**
H. salinarum	cydB, *VNG_2525H*	**ogg**, *VNG_2498H*	**ogg**
T. brucei	GLE2, *Tb927.11.14995*, Tb927.7.1290, *Tb927.8.4240*, **Tb927.8.5510**	DRBD9, GLE2, PPL2, Tb927.10.6550, Tb927.3.5150, **Tb927.8.5510**, TOP2	DRBD9, NST4, *SET30*, *Tb927.2.6100*, **Tb927.6.1580**, **Tb927.8.5510**
T. thermophila	PHR2[a], TTHERM_000530789, *TTHERM_00145210*, TTHERM_00147470, TTHERM_00361370, TTHERM_00463150, *TTHERM_00614680*, TTHERM_00852850	APN2[a], *PARP4*, PARP6, *PCP1*, *PCP2*, PHR2[a]	PARP6, *PCP1*, *TTHERM_00013250*
S. pombe	**myh1**	sac11, *SPAC3H8.08c*, top2	alp5, *hmo1*, *hpz1*, kin1, *mca1*, mlo3, *moc3*, nop12, **rfc1**, **rfc2**, **rfc3**, **rfc4**, **rfc5**, *SPAC3H8.08c*, *SPCC126.11c*
S. cerevisiae	**APN1**, ASG1, MYO4, *NUT1*, **PHR1**, POL5, *RNQ1*	**APN1**, ASG1, CMR1, **INO80**, MAK5, MYO4, *PDR1*, **PHR1**, POL5, **RFC1**, **RFC2**, **RFC3**, **RFC4**, **RFC5**, **RSC1**, *RSC58*, RSC6, **SNF2**, SWI6, TOP2	APL4, **APN1**, ASG1, CMR1, *HAP1*, **INO80**, MBP1, **MGM101**, MYO4, *OAF3*, *PDR1*, POL5, **RFC1**, **RFC2**, **RFC3**, **RFC4**, **RFC5**, **RSC1**, *RSC30*, *RSC58*, RSC6, RSC9, SFH1, **SNF2**, **STH1**, SWI6, TOP2, *YPL245W*
C. elegans	*col-143*, **exo-3**, hmg-5	**apn-1**, *col-119*, *col-140*, *col-143*, *dpy-17*, **exo-3**, *F07A5.2*, *F07H5.8*, his-74, **K07C5.3**, obr-1, **parp-2**, *perm-2*, *phat-1*, *phat-2*, *T01E8.8*, *Y14H12B.2*, *Y37D8A.19*	*C27D8.2*, **exo-3**, *F07A5.2*, hmg-12, *T01E8.8*
H. sapiens (HeLa)	**FANCI**, FERMT2, KPNA6, MYL12A, *NACC1*, PPWD1, RTRAF	**APTX**, ATP5MG, *BEND3*, **BLM**, BOP1, COQ6, DNAJC13, EXOSC3, GATAD2A, HNRNPF, HNRNPH2, **HPF1**, *ISG20L2*, **LIG3**, MRTO4, MYL12A, NAP1L1, NIP7, **NOP53**, **PARP1**, **POLB**, PPIG, *RIOX1*, RPL21, RPLP1, RPS26, *S100A8*, **UBE2N**, **XRCC1**	AHCTF1, CENPV, CHD2, *FXR1*, *KAT6A*, *MECP2*, **MPG**, *PCGF1*, *SAP130*, ZMYND11, *ZNF512B*
H. sapiens (HEK293)	MAX, **MUTYH**, NTHL1, SEPTIN11	**APTX**, CMSS1, **DDB1**, **DDB2**, DNAJC13, **LIG3**, NOC3L, **PARP2**, **PNKP**, **POLB**, WRN, **XPC**, **XRCC1**	AHCTF1, *APOBEC3C*, *BCOR*, *BCORL1*, BRPF1, CENPV, CHD1, CHD2, *CTCF*, GLYR1, *KAT6A*, KRI1, KRR1, **MPG**, *MSANTD7*, NIP7, NOC3L, *NSD2*, NUP205, *PCGF1*, PITX2, RNF2, SUB1, **TRIP12**, *ZNF512B*

Table 1. *Continued*

Species	8-oxoG	abasic	RNA base
Z. mays	B4FTT9[a], *P06678*	A0A1D6F6W7[a], A0A1D6JZF1[a], A0A1D6K922, *A0A1D6LV91*, *A0A1D6NSE6*, A0A1D6P5Y9, *A0A804P6S3*, *B4FDA0*, B4FER3[a], **B4FJC2**, **B4FQT5**, *B4FRR3*, *B4FWP8*, *B4FX14 B6SNB5*, , B6U4F1, *K7UTP1*, K7VBU4[a]	A0A1D6F4B6, *A0A1D6GRJ8*, A0A1D6HK01, A0A1D6HW59, *A0A1D6LV91*, A0A1D6LVY7, *A0A1D6MYU1*, *A0A1D6N2N7*, A0A1D6NSE6, A0A1D6QEP6, *A0A804MH07*, *A0A804MT25*, A0A804NRM4, *A0A804R2N8*, *B4FDA0*, B4FDW2, *B4FRR3*, *B4FX14*, B4G1M3, B4G1W8, *B6SNB5*, *B6UA70*, *C0P7N5*, C0P9C9, C4J4W6, C4J9R0, *C4JC33*, *K7UTP1*, Q6R9L4
A. thaliana	**ARP**, At1g09150, At4g32105, **At5g16990**, CRYD, PHR1, TRE1	At1g06260, At1g07080, CRYD, **MOC1**, PHR1	**ARP**, HON5, **MOC1**, TRE1

[a]Indicates orthology to known DNA damage repair factor, bold indicates previously known role in DNA damage repair, italics indicates no OrthoMCL orthology with the other 10 species included in the study.

the study (Table 1, italicized protein names), four of which were repair proteins (Figure 2A). This suggests that in addition to finding conserved and previously established DNA repair factors, we also enriched species-specific DNA repair proteins.

To determine which functionalities were overrepresented, in addition to general "DNA repair", within the interactors of 8-oxoG, abasic, and RNA lesions, we utilized both the Kyoto Encyclopedia for Genes and Genomes (KEGG) and GO (Table S5).[21,22] We found an overrepresentation of the KEGG term "base excision repair" for all lesions. There was additional enrichment of "nucleotide excision repair", "mismatch repair", and "DNA replication" (Figure 2B). Further interrogation of the enriched interactors of 8-oxoG showed enrichment of the GO biological processes "base-excision repair", "base-excision repair ap site formation", and "photoreactive repair" (Table S5, Figure S2). Within the interactors of the abasic lesion, there was enrichment of "DNA repair" annotated proteins in multiple species, and there were seven more terms belonging to the parent term of "DNA repair". Four DNA repair-related GO terms ("UV damage excision repair", "double-strand break repair", "DNA repair", and "base excision repair") were overrepresented among the interactors of the RNA base lesion.

To investigate the context of our enriched proteins at each of the lesions, we created lesion- and species-specific networks using previously established interactions and proteins included in the STRING database (Tables S6, S7, and S9, Figure S3A).[23] We found a total of 339 interactions across our enriched proteins and species (Figure S3B). Of these enriched protein sets (3 lesions, 12 conditions, 36 total), ~61% had previously reported interactions among them. The largest number of known interactions (90) was found for the RNA lesion in *S. cerevisiae*. The 8-oxoG, abasic, and RNA-enriched proteins exhibited 7, 187, and 151 previously established interactions, respectively. This indicates the relative specificity of the 8-oxoG recognition and a more complex response resolving abasic and RNA lesions.

Interactors of 8-oxoG, abasic, and RNA lesions across phylogenetic branches

To compare the overlap of enriched orthologs across the included species at the 8-oxoG lesion, abasic lesion, and RNA base, we again used orthology group predictions by OrthoMCL (Table S2), only counting proteins that surpassed our enrichment threshold (Figure 3, Tables S4 and S10). Within the interactors of the 8-oxoG lesion, we identified protein families that were conserved in up to four species (Figures 3A, 3B, and S4). The most conserved protein families were photolyases, MUTYH, and ExoIII-like and EndoIV-like AP endonucleases. Photolyases are critical repair proteins in bacteria, archaea, plantae, fungi, and animals. Despite their importance, they lost all DNA repair functionality in placental mammals.[24] The five enriched photolyases were grouped into two orthology groups (hsap_CRY1/OG6_100453 and atha_PHR1/OG6_104135). The divergence in these orthology groups indicates a specialization of the photolyases between species. It was unanticipated that photolyases would be enriched at 8-oxoG, as
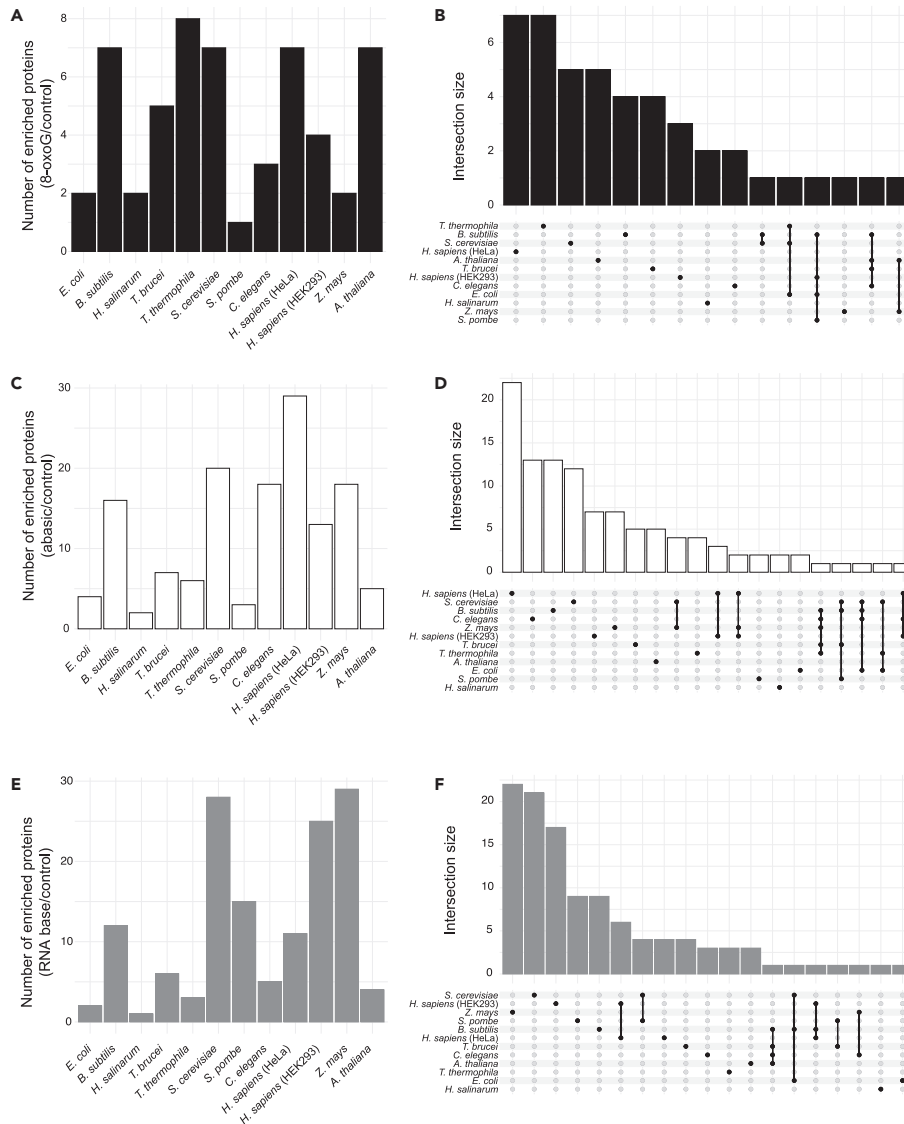
Figure 3. Interactors of the different lesions across phylogenetic branches

(A) Bar plot of the total number of enriched proteins at 8-oxoG across species.

(B) UpSet plot showing overlap of enriched proteins at the 8-oxoG lesion for the different species based on assigned orthology groups via OrthoMCL.

(C) Bar plot of the total number of enriched proteins at abasic lesions per species.

(D) UpSet plot showing overlap of enriched proteins at the abasic lesion for the different species based on assigned orthology groups via OrthoMCL.

(E) Bar plots of the total number of enriched proteins at the uracil RNA base per species.

(F) UpSet plot showing overlap of enriched proteins at the RNA base lesion for the different species based on assigned orthology groups via OrthoMCL.

typically these proteins recognize and resolve pyrimidine dimers. However, with the enrichment traversing five different species, there is a strong argument to suggest that a base conversion or lesion intermediate interacts with these photolyases and is resolved similarly across the Tree of Life. Other conserved interactors enriched at the 8-oxoG lesions were four members of the hsap_MUTYH group (OG6_102506). This

enrichment was specific to 8-oxoG in *B. subtilis, S. pombe*, and *H. sapiens*, whereas mutY in *E. coli* was also bound to the abasic lesion. Although this is a well-characterized BER glycosylase, it has thus far been shown primarily to bind 8-oxoG:A as opposed to the 8-oxoG:C used here. It is possible that the MUTYH orthologs generally bind to 8-oxoG due to their strong affinity, or they bind to a shared intermediate state of 8-oxoG:A and 8-oxoG:C.[25]

At the abasic lesion, we found a higher degree of overlapping proteins with seven instances of three or more orthologs enriched in two or more species (hsap_DNAJC13, hsap_TOP2B, scer_PHR1, hsap_LIG3, scer_APN1, hsap_APTX1, and hsap_APEX1) (Figures 3C, 3D, and S5, Table S10). Two anticipated groups were the hsap_APEX1 (ExoIII-like) and scer_APN1 (EndoIV-like) AP endonucleases (OG6_101139 and OG6_104339, respectively), which are critical to the removal of abasic sites. Members of hsap_LIG3 and hsap_APTX1 are also critical to the BER pathway.[2] While LIG3 has been well studied in *H. sapiens*, the enriched ortholog in *C. elegans* has not been studied in the context of BER (K07C5.3, UniProt ID: Q19138). It is still unclear which ligase is involved in BER in *C. elegans*.[26] There were three homologs enriched in the hsap_APTX1 group in HeLa and HEK cell lines and in *Z. mays*. APTX removes AMP from BER intermediates to form 3'-OH utilized by repair polymerases. A similar enrichment pattern was present in the hsap_DNAJC13 group. DNAJC13 is a heat shock protein that is critical to the heat stress response and has been associated with Parkinson's disease.[27,28] DNAJC13 has not been studied in the context of BER.

Among the enriched proteins interacting with rU across species, members of the RFC (Replication factor C) complex were enriched in both *S. cerevisiae* and *S. pombe* (Table S10). RFC is critical to the loading of PCNA, which is a well-established interactor of RNase H2, an initiator of RER. There was significant enrichment of the hsap_APEX group in *B. subtilis, T. brucei, C. elegans,* and *A. thaliana*. Additionally, proteins of the scer_APN group in *E. coli, B. subtilis,* and *S. cerevisiae* were enriched at rU. While the striking amount of enrichment of AP endonuclease was expected at the abasic and 8-oxoG lesions, this was unanticipated for the RNA lesion. There was also a noticeable enrichment of chromatin remodelers (Figures 3E, 3F, and S6). In both HeLa and HEK293 cells, PCGF1 and CHD2 were enriched. PCGF1 is part of the polycomb repressive complex 1, which is critical to epigenetic alterations repressing gene expression. In HEK293 cells, two interactors of the polycomb repressive complex were enriched, BCOR and BCORL1.[29] Additionally, within HEK293 cells, CHD1 and CTCF, which also mediate chromatin architecture in the presence of damage, were enriched.[30,31] In *S. cerevisiae*, we observed enrichment of the chromatin remodelers Ino80, Snf2, Swi6, and seven members of the remodels the structure of chromatin (RSC) family (Sfh1, Sth1, and Rsc1/6/9/30/58) (Table S4). All of the described chromatin remodelers have not yet been characterized in the misincorporated uracil from DNA but have been directly linked to the promotion of BER.[32]

DNA damage interactors conserved across lesions

In this study, we observed potential DNA repair crosstalk through preferential binding of the same proteins at multiple lesions (Figure 4). We included two DNA damage lesions that are canonical substrates for BER, 8-oxoG, and abasic lesions, as well as a uracil ribonucleotide incorporated into DNA. As 8-oxoG is a common trigger for BER and abasic lesion is a common BER intermediate, we anticipated finding joint interactors between these two lesions. Of the 55 8-oxoG interactors, 19 overlapped with the abasic interactors (Table S11). Within this overlap, we unexpectedly found four instances of photolyases (Figures 4A–4C, Table S11). Additionally, in *B. subtilis*, ydaT was shared between the 8-oxoG and the abasic lesion (Figure 4D). This is an uncharacterized stress response protein that increases resistance to ethanol and low temperatures.[33]

There were 47 instances in which a protein was enriched both at the abasic site and rU. Such a large degree of overlap between the RNA base and abasic lesion was not initially expected. However, there has been evidence that abasic sites can occur within RNA and are primarily resolved by APE1 and MPG.[34] In HEK and HeLa cells, we enriched MPG and its ortholog yxlJ in *B. subtilis*. Additionally, APE1 and APN1 orthologs were enriched in 6 of the 11 species. Thus, the removal of abasic sites from RNA may share mechanisms with uracil and abasic site removal when incorporated into DNA. Our data also suggest that in *S. cerevisiae*, the chromatin remodeling mechanisms that are needed to repair abasic sites are shared for the repair of rU (Ino80, Rsc1, Rsc6, Rsc58, Swi6, and Snf2) (Figure 4B), in line with chromatin state being a critical factor for the removal of both ribonucleotides and BER intermediates.[16,32,35] Beyond the overlaps of enriched proteins between two lesions, we also observed a notable overlap between all three lesions. In

A   *A. thaliana*

B   *S. cerevisiae*
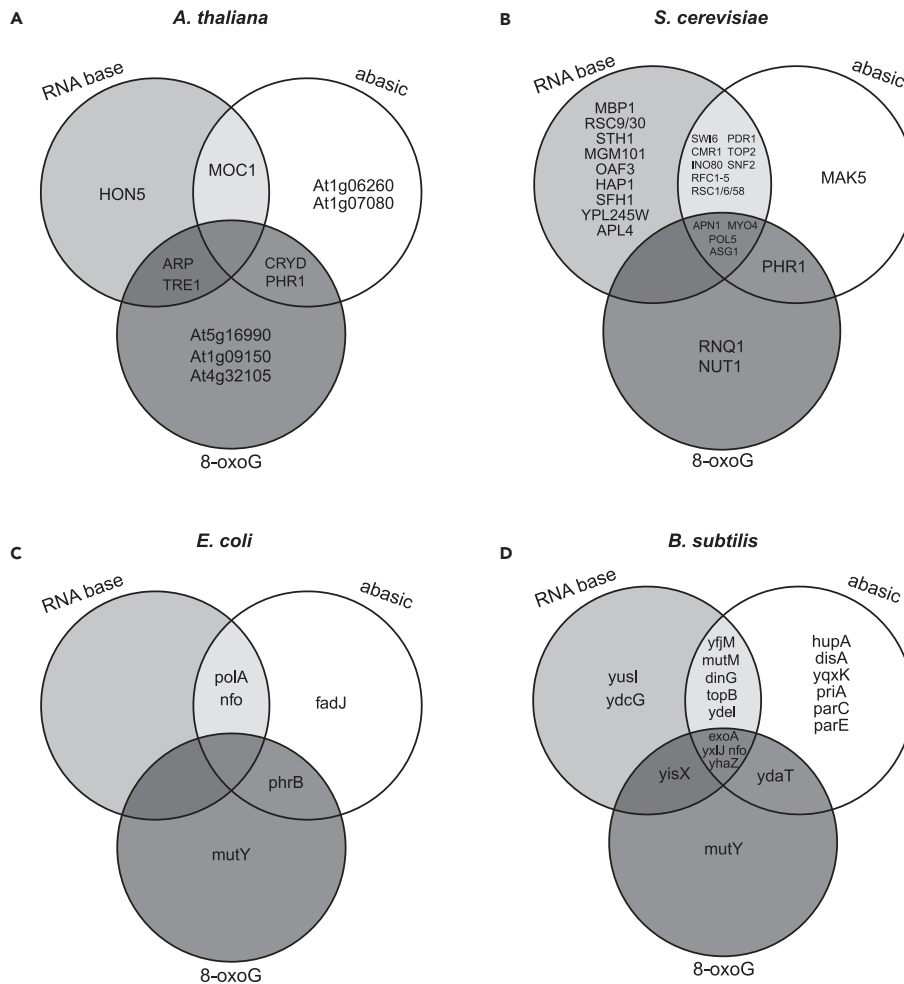

C   *E. coli*

D   *B. subtilis*

Figure 4. Conserved interaction partners across the lesions
Venn diagrams showing the overlapping enriched proteins at the RNA base, abasic site, and 8-oxoG lesions for (A) *A. thaliana*, (B) *S. cerevisiae*, (C) *E. coli*, and (D) *B. subtilis*. Overlap in other species is detailed in Table S11.

B. subtilis, T. brucei, S. cerevisiae, and *C. elegans,* AP endonuclease orthologs are enriched at all three lesions. In *B. subtilis*, we observed two uncharacterized glycosylases, yhaZ, and yxlJ, at all three lesions (Figure 4D). Although ASG-1, POL5, and MYO4 are not characterized as DNA repair proteins, they were also found to interact with all three lesions in *S. cerevisiae* (Figure 4B). Taken together, our screen reiterates a broader profile for DNA repair factors in the repair of 8-oxoG, abasic, and RNA lesions and a potential crosstalk between the different repair pathways (Figure 4, Table S11).

Binding patterns by DNA repair factors are evolutionarily conserved across all domains of life

As the maintenance of genome stability is critical in each organism, many DNA damage factors are conserved in both sequence and functionality across species.[18] Across species and lesions, we enriched for classical BER-related proteins, including orthologs of the glycosylases MUTYH and MPG, deadenylase APTX, LIG3 and XRCC1, PCNA clamp loader RFC1-4, POLB, and the AP endonucleases APEX1 and Apn1 (Figure 5A, Table S4). The APEX1/APE1 and Apn1 orthology groups represent the ExoIII-like AP exonucleases and EndoIV-like AP endonucleases, respectively. These groups of conserved AP endonucleases have been studied at length due to their evolutionary history.[2,36,37] Using a maximum likelihood phylogenetic
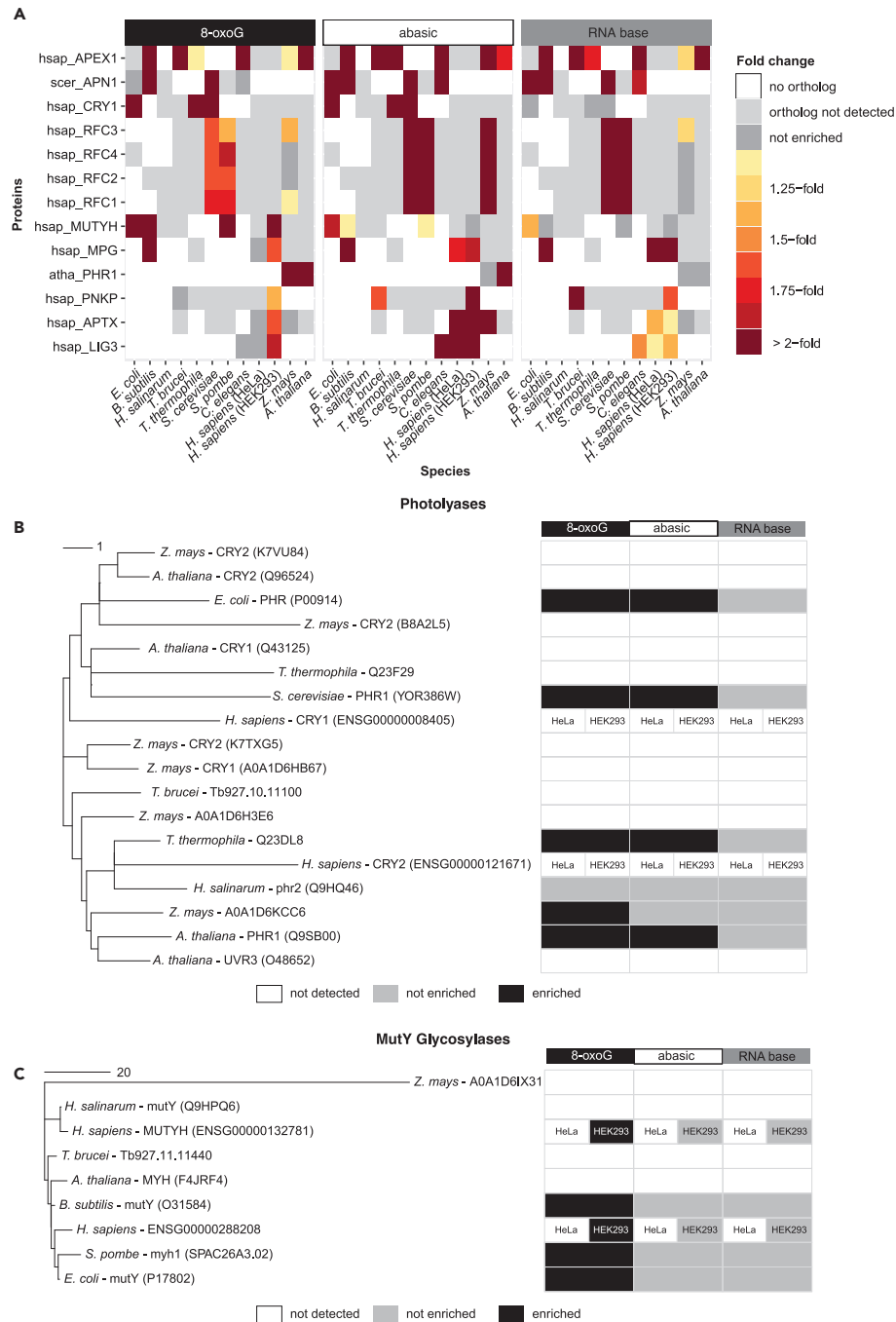
Figure 5. Conservation of DNA repair orthologs across the Tree of Life

(A) Heatmap representing enrichment levels of OrthoMCL orthology groups with GO annotation "DNA repair" (GO:0006281) with two or more enriched proteins across eleven species and 8-oxoG (black), abasic (white) and RNA base (gray) lesions. The color scale represents the fold change in comparison to control samples. Abbreviations: hsap, *Homo sapiens*; scer, *S. cerevisiae*; cele, *C. elegans*; atha, *A. thaliana*; spom, *S. pombe*.

Figure 5. *Continued*

(B) Maximum likelihood phylogenetic tree of the photolyase gene family including information on detection and enrichment (fold change > 2, Welch t-test p value < 0.05) for the different lesions. White boxes represent proteins that were not detected in the respective experiment. The scale bar in the plots indicates the number of amino acid substitutions per site.

(C) Maximum likelihood phylogenetic tree of the MUTY glycosylase gene family. Same as (B).

tree including all AP endonucleases across the 11 species, we demonstrate the potential enrichment differences between the two groups (Figure S7). For both groups of endonucleases, we found 2-fold or greater binding to 8-oxoG and abasic lesions in eight of the eleven species. Additionally, more unexpectedly spanning both groups was the enrichment of AP endonucleases at the RNA base in six of the eleven species. While AP endonucleases have been well characterized within BER, thus far, they have been shown to play a more minor role in RER.[16] It is possible that both types of AP endonucleases play a larger role than originally anticipated.

Two additional protein families that had highly conserved enrichment patterns were the photolyases (scer_PHR1 and atha_PHR1) and MUTYH-related glycosylases (hsap_MUTYH). Despite both being DNA repair proteins, the binding of these proteins was unexpected in this particular context (Figures 5B and 5C). Photolyases are known to have specific repair activity for cyclobutane pyrimidine dimers and 6-4 pyrimidine-pyrimidone photoproducts caused by UV light.[38] However, the *S. cerevisiae* PHR1 orthologs in *E. coli*, *T. thermophila*, and *S. cerevisiae* were significantly enriched at both the 8-oxoG and abasic lesions (Figure 5B). Both orthologs in the atha_PHR1 group were also significantly enriched at the 8-oxoG lesion. The enrichment at the abasic lesion in *A. thaliana* and for *Z. mays* was 1.9-fold, just below our threshold. For these orthology groups, the maximum likelihood phylogenetic tree showed a clear divergence of the plant photolyases, despite their similar *in vitro* binding characteristics. We did not observe the enrichment of any orthologs of PHR1 (atha_PHR1 and scer_PHR1) at the RNA lesion, which extended across all species regardless of evolutionary relation (Figure 5B).

MutY-related glycosylases are well characterized in the removal of 8-oxoG:A, but there are few studies showing their binding to 8-oxoG:C, which was used in this study. In an *in vitro* setting when the diffusion rate was measured, MUTYH orthologs would reside much longer at 8-oxoG:A but also have moderate stalling at 8-oxoG:C.[39] MUTYH orthologs were found to bind specifically to 8-oxoG in *E. coli*, *B. subtilis*, *S. pombe*, and *H. sapiens* (Figure 5C). There were no instances of detection of an MUTYH ortholog without enrichment at 8-oxoG, indicating highly specific binding that was independent of the evolutionary relation of the protein sequences. MUTYH has recently been suggested to facilitate the overall DNA damage response as a scaffolding protein.[40] While this function has been primarily explored within vertebrates, our findings indicate that its multiple functionalities might have emerged far earlier in evolution than originally estimated (Figure 5C).

Identification of uncharacterized DNA repair proteins across multiple species

In addition to the known DNA repair proteins, one-third of the enriched proteins were previously not associated with the "DNA repair" GO term (GO:0006281). We found enrichment of 35, 85, and 105 non-DNA repair classified proteins at the 8-oxoG lesion, abasic lesion, and RNA base, respectively. To investigate these proteins further, we created species-specific networks for all three DNA damage lesions using the STRING database (Figure S8 and Tables S6, S7, S8, and S9). Within these networks, we marked proteins categorized as repair (triangle) and non-DNA repair proteins (circle) and indicated at which lesion they were enriched. Here, we highlighted the *S. cerevisiae*, *C. elegans*, and *T. thermophila* networks (Figure 6A). Within *S. cerevisiae*, all enriched proteins contained in STRING interacted and formed one large network (Figure 6A). Included are five chromatin remodelers (RSC6, RSC9, RSC58, SFH1, and SWI6) that, although not characterized as DNA repair proteins, had a prominent number of interactions with both repair and non-DNA repair proteins. Within the RSC family, RSC1, RSC30, and STH1 have been classified as DNA repair proteins and are specifically linked to BER.[32,41] This suggests that the other RSC proteins likely play a role in chromatin remodeling surrounding DNA repair. Additionally, the non-DNA repair protein CMR1 had 11 interaction partners, five of which were "DNA repair" proteins. Notably, although not included in the "DNA repair" GO term, CMR1 has been shown to be needed to resolve genotoxic stress and has a preference for binding UV lesions *in vitro*.[42,43] For the *C. elegans* interactors, we identified three different subnetworks (Figures 6A and S8). Within one subnetwork, the "DNA repair" proteins parp-2,
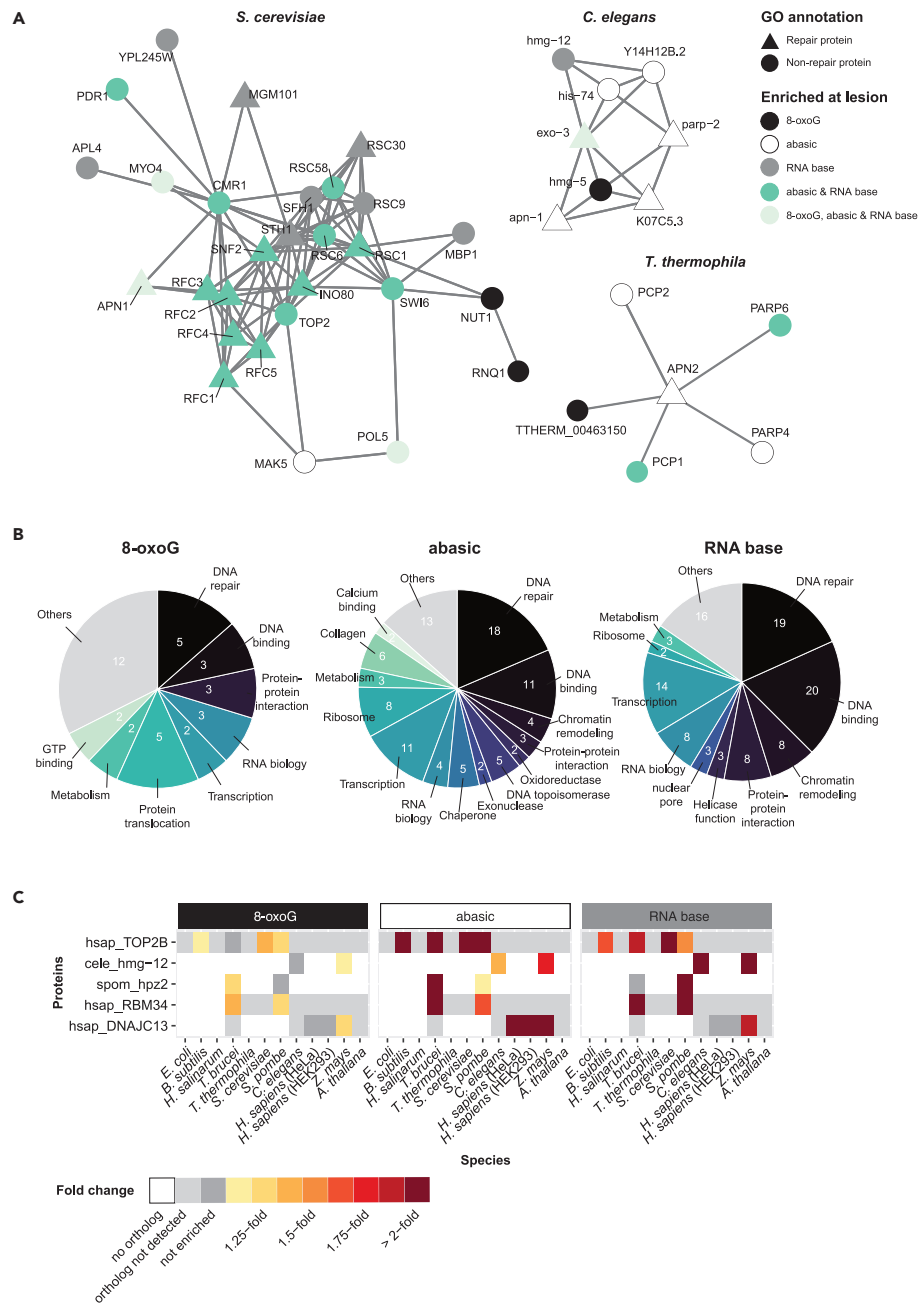
Figure 6. Network, domain, and phylogenetic analyses implicate novel proteins in DNA repair

(A) Networks of enriched proteins across lesions for *S. cerevisiae*, *C. elegans*, and *T. thermophila*. Interactions as established in the STRING database.

(B) Classification of non-DNA repair proteins based on Pfam domain annotation. The total number of proteins classified at 8-oxoG was 29, at abasic 75, and at the RNA base 74.

(C) Heatmap representing enrichment levels of OrthoMCL orthology groups without GO annotation "DNA repair" (GO:0006281) with two or more enriched proteins across all eleven species and 8-oxoG (black), abasic (white) and RNA base (gray) lesions. The color scale represents the fold change in comparison to control samples. Abbreviations: hsap, *Homo sapiens*; cele, *C. elegans*; spom, *S. pombe*.

exo-3, and apn-1 interacted with 3, 4, and 1 non-DNA repair proteins, respectively. All three proteins were mutually linked to hmg-5. Hmg-5 was studied in a *C. elegans* Parkinson's disease model, and together with nth-1, BER glycosylase, and other associated proteins reduced mitochondrial stress and oxidative damage.[44] Within the *T. thermophila* network, there were mutual interactions between APN2, identified as a DNA repair protein based on its orthology to the *S. cerevisiae* AP endonuclease, and four different PARP-related proteins, as well as TTHERM_00463150, which has not been characterized. This indicates that APN2 might orchestrate the recruitment of PARP-related proteins or that PARP-related proteins are needed for APN2 to access DNA.

We evaluated the Pfam domains found among the enriched non-DNA repair proteins to elucidate more of their potential functionalities (Table S12).[45] The two most frequently identified domains were DNA-binding domains: (1) "protein of unknown function, DUF573" (corresponding to Interpro protein family "GLABROUS1 enhancer-binding protein family"), which is often part of proteins associated with plant stress response, and (2) "Fungal Zn(2)-Cys(6)" often involved in growth and metabolism.[46,47] We assigned each Pfam domain into one of 15 categories to summarize its primary function (Table S12). In all three lesions, the majority of domains were related to DNA repair and DNA binding (Figure 6B). Thus, despite the lack of categorization as DNA repair genes under the GO term "DNA repair", there was a clear link to DNA repair functionality within these proteins. For example, we identified the "poly(ADP-ribose) polymerase" and "DNA-Ligase Zn-finger region" in four different proteins. These included hpz1 in *S. pombe* and Tb927.10.6550 in *T. brucei*, which both belong to the same orthology group. The other two proteins are PARP-related proteins in *T. thermophila*, PCP1 and PCP2. We also detected the "PARP-associated WGR domain" and a "PARP catalytic domain" in PARP4 and PARP6 in *T. thermophila*.

Furthermore, we examined the conservation of enrichment of non-DNA repair proteins across species to further support a role in DNA damage repair and recognition of lesions. We found at least five instances in which non-DNA repair genes were enriched in multiple species (Figure 6C). Intriguingly, some of these proteins were also identified within our domain analysis. For example, both enriched proteins in the spom_hpz2 orthology group in *T. brucei* and *S. pombe* contained a PARP-related domain. Furthermore, there was specific enrichment of the *T. brucei* ortholog (Tb927.10.6550) at the abasic lesion and the *S. pombe* ortholog (hpz1) at the RNA base. Additionally, all three proteins enriched within the hsap_DNAJC13 orthology group have Pfam 'DnaJ domains'. These proteins preferentially bound to the abasic lesion in both HEK293 and HeLa cell lines as well as the two paralogs in *Z. mays* (UniProt: A0A1D6K922 and A0A1D6P5Y9). The conservation of enrichment across various species in both cases suggests a very likely role in DNA repair.

Through the use of network, domain, and phylogenetics analysis, we have identified proteins that, despite not being classified as DNA repair proteins, likely have a role in the DNA damage response.

Conclusions

Performing a mass spectrometry-based phylointeractomics screen across 11 species, we compared the binding capabilities of three well-established DNA damage lesions, an 8-oxoG modification, abasic site, and ribonucleotide base incorporation. We enriched 337 proteins across all lesions and selected species (Table 1). Of these 337 proteins, 99 were related to DNA repair, which in a proteome-wide generic screen with thousands of possible proteins strongly indicates the specificity of the experiment. Supporting the specificity even further, DNA repair-related KEGG and GO terms were overrepresented in the enriched group of proteins. Through phylogenetic analysis, we established that the enrichment of particular DNA damage proteins extends through many species.

In addition to DNA repair genes, we identified two other intriguing groups of interactors in our screen. Namely, we detected an enrichment of 82 species-specific proteins as well as proteins that have not been implicated previously in DNA repair. This group of proteins presents an avenue to study potentially unique aspects of repair or damage response in their corresponding model organism. To elucidate

functionality and connection to DNA damage repair for originally non-DNA repair proteins, we utilized network, domain, and phylogenetics analysis. With this, we indicated an additional 44 proteins to potentially play a role in the DNA damage response.

Our study systematically evaluates *in vitro* binding partners in both BER lesions and an RNA lesion in eleven model species across the Tree of Life. We recapitulate previous findings and nominate putative unknown candidates to be involved in the resolution of these lesions. Through the use of network, domain, and phylogenetics analysis, we identified a subset of non-DNA repair classified proteins to likely be involved in DNA repair. Overall, this study opens avenues for further investigation of newly identified candidates to explore key factors in the crosstalk between BER and RER DNA damage pathways.

Limitations of the study

In some cases, we do not identify or enrich all expected interaction partners at the included lesions, which can be caused by a variety of reasons. For instance, preparation from a large range of different tissues and cellular material can lead to variation in the pool of proteins available for measurement. The lack of *in vivo* conditions, such as pH, salt concentrations, temperature, post-translational modifications, and many other cellular conditions, affects DNA-protein interactions. As we did not perform cross-linking mass spectrometry, it is possible that some more transient interactions were not maintained. Furthermore, it is important to highlight the likely creation of repair intermediates in the *in vitro* pull-down assays. The ability to repair 8-oxoG, abasic sites, and uracil residues *in vitro* has been previously demonstrated with human cell extract.[48,49] However, we did find that unrepaired lesions existed in our experiment; for example, 11 out of the 24 canonical DNA repair-related proteins were uniquely enriched at the 8-oxoG lesion, suggesting that unrepaired lesions persisted.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - ○ Lead contact
  - ○ Materials availability
  - ○ Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - ○ Cultivation and extract preparation
  - ○ DNA pull-down experiments
  - ○ Mass spectrometry sample preparation
  - ○ Mass spectrometry measurement
  - ○ Mass spectrometry data analysis
  - ○ Bioinformatics analysis and statistical analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106778.

REFERENCES

1. Ciccia, A., and Elledge, S.J. (2010). The DNA Damage Response: making it safe to play with knives. Mol. Cell 40, 179–204. https://doi.org/10.1016/j.molcel.2010.09.019.

2. Beard, W.A., Horton, J.K., Prasad, R., and Wilson, S.H. (2019). Eukaryotic base excision repair: new approaches shine light on mechanism. Annu. Rev. Biochem. 88, 137–162. https://doi.org/10.1146/annurev-biochem-013118-111315.

3. Klungland, A., Rosewell, I., Hollenbach, S., Larsen, E., Daly, G., Epe, B., Seeberg, E., Lindahl, T., and Barnes, D.E. (1999). Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. Proc. Natl. Acad. Sci. USA 96, 13300–13305. https://doi.org/10.1073/pnas.96.23.13300.

4. Marshall, C.J., and Santangelo, T.J. (2020). Archaeal DNA repair mechanisms. Biomolecules 10, 1472. https://doi.org/10.3390/biom10111472.

5. Genois, M.-M., Paquet, E.R., Laffitte, M.-C.N., Maity, R., Rodrigue, A., Ouellette, M., and Masson, J.-Y. (2014). DNA repair pathways in trypanosomatids: from DNA repair to drug resistance. Microbiol. Mol. Biol. Rev. 78, 40–73. https://doi.org/10.1128/MMBR.00045-13.

6. Yao, S., Feng, Y., Zhang, Y., and Feng, J. (2021). DNA damage checkpoint and repair: from the budding yeast Saccharomyces cerevisiae to the pathogenic fungus Candida albicans. Comput. Struct. Biotechnol. J. 19, 6343–6354. https://doi.org/10.1016/j.csbj.2021.11.033.

7. Robertson, A.B., Klungland, A., Rognes, T., and Leiros, I. (2009). DNA repair in mammalian cells: base excision repair: the long and short of it. Cell. Mol. Life Sci. 66, 981–993. https://doi.org/10.1007/s00018-009-8736-z.

8. Córdoba-Cañero, D., Morales-Ruiz, T., Roldán-Arjona, T., and Ariza, R.R. (2009). Single-nucleotide and long-patch base excision repair of DNA damage in plants. Plant J. 60, 716–728. https://doi.org/10.1111/j.1365-313X.2009.03994.x.

9. Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. Biochemistry 11, 3610–3618. https://doi.org/10.1021/bi00769a018.

10. Gredilla, R., Garm, C., and Stevnsner, T. (2012). Nuclear and mitochondrial DNA repair in selected eukaryotic aging model systems. Oxid. Med. Cell. Longev. 2012, 282438. https://doi.org/10.1155/2012/282438.

11. Reijns, M.A.M., Rabe, B., Rigby, R.E., Mill, P., Astell, K.R., Lettice, L.A., Boyle, S., Leitch, A., Keighren, M., Kilanowski, F., et al. (2012). Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. Cell 149, 1008–1022. https://doi.org/10.1016/j.cell.2012.04.011.

12. Nick McElhinny, S.A., Watts, B.E., Kumar, D., Watt, D.L., Lundström, E.B., Burgers, P.M.J., Johansson, E., Chabes, A., and Kunkel, T.A. (2010). Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. Proc. Natl. Acad. Sci. USA 107, 4949–4954. https://doi.org/10.1073/pnas.0914857107.

13. Williams, J.S., Lujan, S.A., and Kunkel, T.A. (2016). Processing ribonucleotides incorporated during eukaryotic DNA replication. Nat. Rev. Mol. Cell Biol. 17, 350–363. https://doi.org/10.1038/nrm.2016.37.

14. Balachander, S., Gombolay, A.L., Yang, T., Xu, P., Newnam, G., Keskin, H., El-Sayed, W.M.M., Bryksin, A.V., Tao, S., Bowen, N.E., et al. (2020). Ribonucleotide incorporation in yeast genomic DNA shows preference for cytosine and guanosine preceded by deoxyadenosine. Nat. Commun. 11, 2447. https://doi.org/10.1038/s41467-020-16152-5.

15. Sassa, A., Yasui, M., and Honma, M. (2019). Current perspectives on mechanisms of ribonucleotide incorporation and processing in mammalian DNA. Gene Environ. 41, 3. https://doi.org/10.1186/s41021-019-0118-7.

16. Kellner, V., and Luke, B. (2020). Molecular and physiological consequences of faulty eukaryotic ribonucleotide excision repair. EMBO J. 39, e102309. https://doi.org/10.15252/embj.2019102309.

17. Kappei, D., Scheibe, M., Paszkowski-Rogacz, M., Bluhm, A., Gossmann, T.I., Dietz, S., Dejung, M., Herlyn, H., Buchholz, F., Mann, M., and Butter, F. (2017). Phylointeractomics reconstructs functional evolution of protein binding. Nat. Commun. 8, 14334. https://doi.org/10.1038/ncomms14334.

18. Kovalchuk, I. (2016). Chapter 38 - conserved and divergent features of DNA repair: future perspectives in genome instability research. In Genome Stability, I. Kovalchuk and O. Kovalchuk, eds. (Academic Press), pp. 651–666. https://doi.org/10.1016/B978-0-12-803309-8.00038-0.

19. Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189. https://doi.org/10.1101/gr.1224503.

20. Chen, F., Mackey, A.J., Stoeckert, C.J., and Roos, D.S. (2006). OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res. 34, D363–D368. https://doi.org/10.1093/nar/gkj123.

21. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 51, D587–D592. https://doi.org/10.1093/nar/gkac963.

22. Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 49, D325–D334. https://doi.org/10.1093/nar/gkaa1113.

23. Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 49, D605–D612. https://doi.org/10.1093/nar/gkaa1074.

24. Mei, Q., and Dvornyk, V. (2015). Evolutionary history of the photolyase/cryptochrome superfamily in eukaryotes. PLoS One 10,

e0135940. https://doi.org/10.1371/journal. pone.0135940.

25. Yudkina, A.V., Shilkin, E.S., Endutkin, A.V., Makarova, A.V., and Zharkov, D.O. (2019). Reading and misreading 8-oxoguanine, a paradigmatic ambiguous nucleobase. Crystals 9, 269. https://doi.org/10.3390/cryst9050269.

26. Elsakrmy, N., Zhang-Akiyama, Q.-M., and Ramotar, D. (2020). The base excision repair pathway in the nematode Caenorhabditis elegans. Front. Cell Dev. Biol. 8, 598860. https://doi.org/10.3389/fcell.2020.598860.

27. Besemer, A.S., Maus, J., Ax, M.D.A., Stein, A., Vo, S., Freese, C., Nalbach, K., von Hilchen, C., Pfalzgraf, I.F., Koziollek-Drechsler, I., et al. (2021). Receptor-mediated endocytosis 8 (RME-8)/DNAJC13 is a novel positive modulator of autophagy and stabilizes cellular protein homeostasis. Cell. Mol. Life Sci. 78, 645–660. https://doi.org/10.1007/s00018-020-03521-y.

28. Gorenberg, E.L., and Chandra, S.S. (2017). The role of Co-chaperones in synaptic proteostasis and neurodegenerative disease. Front. Neurosci. 11, 248. https://doi.org/10.3389/fnins.2017.00248.

29. Astolfi, A., Fiore, M., Melchionda, F., Indio, V., Bertuccio, S.N., and Pession, A. (2019). BCOR involvement in cancer. Epigenomics 11, 835–855. https://doi.org/10.2217/epi-2018-0195.

30. Stadler, J., and Richly, H. (2017). Regulation of DNA repair mechanisms: how the chromatin environment regulates the DNA damage response. Int. J. Mol. Sci. 18, 1715. https://doi.org/10.3390/ijms18081715.

31. Luijsterburg, M.S., de Krijger, I., Wiegant, W.W., Shah, R.G., Smeenk, G., de Groot, A.J.L., Pines, A., Vertegaal, A.C.O., Jacobs, J.J.L., Shah, G.M., and van Attikum, H. (2016). PARP1 links CHD2-mediated chromatin expansion and H3.3 deposition to DNA repair by non-homologous end-joining. Mol. Cell 61, 547–562. https://doi.org/10.1016/j.molcel.2016.01.019.

32. Tanwar, V.S., Jose, C.C., and Cuddapah, S. (2019). Role of CTCF in DNA damage response. Mutat. Res. 780, 61–68. https://doi.org/10.1016/j.mrrev.2018.02.002.

33. Czaja, W., Mao, P., and Smerdon, M.J. (2014). Chromatin remodelling complex RSC promotes base excision repair in chromatin of Saccharomyces cerevisiae. DNA Repair 16, 35–43. https://doi.org/10.1016/j.dnarep.2014.01.002.

34. Pedreira, T., Elfmann, C., and Stülke, J. (2022). The current state of Subti Wiki, the database for the model organism Bacillus subtilis. Nucleic Acids Res. 50, D875–D882. https://doi.org/10.1093/nar/gkab943.

35. Liu, Y., Rodriguez, Y., Ross, R.L., Zhao, R., Watts, J.A., Grunseich, C., Bruzel, A., Li, D., Burdick, J.T., Prasad, R., et al. (2020). RNA abasic sites in yeast and human cells. Proc. Natl. Acad. Sci. USA 117, 20689–

20695. https://doi.org/10.1073/pnas. 2011511117.

36. Redrejo-Rodríguez, M., Vigouroux, A., Mursalimov, A., Grin, I., Alili, D., Koshenov, Z., Akishev, Z., Maksimenko, A., Bissenbaev, A.K., Matkarimov, B.T., et al. (2016). Structural comparison of AP endonucleases from the exonuclease III family reveals new amino acid residues in human AP endonuclease 1 that are involved in incision of damaged DNA. Biochimie 128–129, 20–33. https://doi.org/10.1016/j.biochi.2016.06.011.

37. Daley, J.M., Zakaria, C., and Ramotar, D. (2010). The endonuclease IV family of apurinic/apyrimidinic endonucleases. Mutat. Res. 705, 217–227. https://doi.org/10.1016/j.mrrev.2010.07.003.

38. Kavakli, I.H., Baris, I., Tardu, M., Gül, Ş., Öner, H., Çal, S., Bulut, S., Yarparvar, D., Berkel, Ç., Ustaoğlu, P., and Aydın, C. (2017). The photolyase/cryptochrome family of proteins as DNA repair enzymes and transcriptional repressors. Photochem. Photobiol. 93, 93–103. https://doi.org/10.1111/php.12669.

39. Nelson, S.R., Kathe, S.D., Hilzinger, T.S., Averill, A.M., Warshaw, D.M., Wallace, S.S., and Lee, A.J. (2019). Single molecule glycosylase studies with engineered 8-oxoguanine DNA damage sites show functional defects of a MUTYH polyposis variant. Nucleic Acids Res. 47, 3058–3071. https://doi.org/10.1093/nar/gkz045.

40. Raetz, A.G., and David, S.S. (2019). When you're strange: unusual features of the MUTYH glycosylase and implications in cancer. DNA Repair 80, 16–25. https://doi.org/10.1016/j.dnarep.2019.05.005.

41. Bohm, K.A., Hodges, A.J., Czaja, W., Selvam, K., Smerdon, M.J., Mao, P., and Wyrick, J.J. (2021). Distinct roles for RSC and SWI/SNF chromatin remodelers in genomic excision repair. Genome Res. 31, 1047–1059. https://doi.org/10.1101/gr.274373.120.

42. Gallina, I., Colding, C., Henriksen, P., Beli, P., Nakamura, K., Offman, J., Mathiasen, D.P., Silva, S., Hoffmann, E., Groth, A., et al. (2015). Cmr1/WDR76 defines a nuclear genotoxic stress body linking genome integrity and protein quality control. Nat. Commun. 6, 6533. https://doi.org/10.1038/ncomms7533.

43. Choi, D.-H., Kwon, S.-H., Kim, J.-H., and Bae, S.-H. (2012). Saccharomyces cerevisiae Cmr1 protein preferentially binds to UV-damaged DNA in vitro. J. Microbiol. 50, 112–118. https://doi.org/10.1007/s12275-012-1597-4.

44. SenGupta, T., Palikaras, K., Esbensen, Y.Q., Konstantinidis, G., Galindo, F.J.N., Achanta, K., Kassahun, H., Stavgiannoudaki, I., Bohr, V.A., Akbari, M., et al. (2021). Base excision repair causes age-dependent accumulation of single-stranded DNA breaks that contribute to Parkinson disease pathology. Cell Rep. 36, 109668. https://doi.org/10.1016/j.celrep.2021.109668.

45. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S.,

Richardson, L.J., et al. (2021). Pfam: the protein families database in 2021. Nucleic Acids Res. 49, D412–D419. https://doi.org/10.1093/nar/gkaa913.

46. Huang, J., Zhang, Q., He, Y., Liu, W., Xu, Y., Liu, K., Xian, F., Li, J., and Hu, J. (2021). Genome-wide identification, expansion mechanism and expression profiling analysis of GLABROUS1 enhancer-binding protein (GeBP) gene family in gramineae crops. Int. J. Mol. Sci. 22, 8758. https://doi.org/10.3390/ijms22168758.

47. Tianqiao, S., Xiong, Z., You, Z., Dong, L., Jiaoling, Y., Junjie, Y., Mina, Y., Huijuan, C., Mingli, Y., Xiayan, P., et al. (2021). Genome-wide identification of Zn2Cys6 class fungal-specific transcription factors (ZnFTFs) and functional analysis of UvZnFTF1 in ustilaginoidea virens. Rice Sci. 28, 567–578. https://doi.org/10.1016/j.rsci.2021.03.001.

48. Parsons, J.L., and Dianov, G.L. (2012). In vitro base excision repair using mammalian cell extracts. In DNA Repair Protocols Methods in Molecular Biology, L. Bjergbæk, ed. (Humana Press), pp. 245–262. https://doi.org/10.1007/978-1-61779-998-3_17.

49. Squillaro, T., Finicelli, M., Alessio, N., Del Gaudio, S., Di Bernardo, G., Melone, M.A.B., Peluso, G., and Galderisi, U. (2019). A rapid, safe, and quantitative in vitro assay for measurement of uracil-DNA glycosylase activity. J. Mol. Med. 97, 991–1001. https://doi.org/10.1007/s00109-019-01788-8.

50. de Albuquerque, B.F.M., Luteijn, M.J., Cordeiro Rodrigues, R.J., van Bergeijk, P., Waaijers, S., Kaaij, L.J.T., Klein, H., Boxem, M., and Ketting, R.F. (2014). PID-1 is a novel factor that operates during 21U-RNA biogenesis in Caenorhabditis elegans. Genes Dev. 28, 683–688. https://doi.org/10.1101/gad.238220.114.

51. Scherer, M., Levin, M., Butter, F., and Scheibe, M. (2020). Quantitative proteomics to identify nuclear RNA-binding proteins of Malat1. Int. J. Mol. Sci. 21, E1166. https://doi.org/10.3390/ijms21031166.

52. Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. 2, 1896–1906. https://doi.org/10.1038/nprot.2007.261.

53. Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinforma. Oxf. Engl. 21, 3439–3440. https://doi.org/10.1093/bioinformatics/bti525.

54. Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191. https://doi.org/10.1038/nprot.2009.97.

55. R Core Team (R Foundation for Statistical Computing. 2022). R: A Language and Environment for Statistical Computing.

56. Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. Protein Sci. *28*, 1947–1951. https://doi.org/10.1002/pro.3715.

57. Csárdi, G., and Nepusz, T. (InterJournal, complex systems. 2006). The Igraph Software Package for Complex Network Research.

58. Tyner, S., Briatte, F., and Hofmann, H. (2017). Network visualization with ggplot2. Rom.

Jahrb. *9*, 27–59. https://doi.org/10.32614/RJ-2017-023.

59. Sievers, F., and Higgins, D.G. (2018). Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. *27*, 135–145. https://doi.org/10.1002/pro.3290.

60. Paradis, E., and Schliep, K. (2019). Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*,

526–528. https://doi.org/10.1093/bioinformatics/bty633.

61. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. *50*, D543–D552. https://doi.org/10.1093/nar/gkab1038.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E. coli: DH5α	NEB	C2987H
B. subtilis: DSM10	DSMZ	DSM10
Chemicals, peptides, and recombinant proteins		
Dynabeads Streptavidin C1	Thermo Scientific	65002
Complete protease inhibitor cocktail tablets	Roche	04693116001
Protein assay dye reagent concentrate 5x	Biorad	#500-0006
NuPAGE LDS Sample Buffer (4x)	Life Technologies	NP0008
DL-DITHIOTHREITOL, >=98% (TLC), >=99.0%	Sigma	D0632
4-10% NuPage NOVEX PAGE gel	Novex	NP0321BOX
Iodoacetamide >=99% (NMR)	Sigma	I6125
Acetonitrile	VWR	20048.320
Trypsin, MS approved, from porcine pancreas	Serva	37286.03
ReproSil-Pur 120 C18-AQ, 1.9 µm 15 % C endc.	Dr. Maisch GmbH	r119.aq.0001
Deposited data		
Original code	This paper (Mario Dejung)	Github: https://github.com/mariodejung/DNAdamage_phylointeractome
Original code	This paper (Albert Fradera Sola)	Github: https://github.com/AFraderaSola/DNADamage_Phylointeracome
Mass spectrometry proteomics data	This paper	ProteomeXchange; PXD036040
Experimental models: Cell lines		
H. sapiens: HeLa	ECACC	HeLa S3
H. sapiens: HEK293	ECACC	HEK293
Experimental models: Organisms/strains		
H. salinarum: NRC-1	Vencio lab (University of Sao Paulo)	NRC-1
S. cerevisiae: BY4742α	Luke lab (JGU)	BY4742α
S. pombe: pp265	Baumann lab (JGU)	pp265
T. thermophila: SB210	Tetrahymena Stock Center	Stock ID: SD00703
T. brucei:	Engstler lab (University of Wuerzburg)	Lister 427
C. elegans: N2	Caenorhabditis Genetics Center	Strain Name: N2 Genotype: *C. elegans* wild isolate
A. thaliana: Columbia	Wachter lab (JGU)	Columbia
Z. mays	LIDL	N/A
Oligonucleotides		
Control without Lesion: AGAGTAAGGGCCT GCGGCGAGGATCCGACCACGATTCGCGC AGAAGGGGCCGAAATTCGCCGTGGACTC CCTCAGTAAT	Bio-synthesis	N/A

(Continued on next page)

89

REAGENT or RESOURCE	SOURCE	IDENTIFIER
8-oxoG lesion: AGAGTAAGGGCCTGCGGC GAG(8-Oxo-dG) ATCCGACCACGATTCGCG CAGAAGGGGCCGAAATTCGCCGTGGACT CCCTCAGTAAT	Bio-synthesis	N/A
abasic lesion: AGAGTAAGGGCCTGCGGCG AG(dSpacer) ATCCGACCACGATTCGCGCA GAAGGGGCCGAAATTCGCCGTGGACTCC CTCAGTAAT	Bio-synthesis	N/A
RNA lesion: AGAGTAAGGGCCTGCGGCG AG(rU) ATCCGACCACGATTCGCGCAGAA GGGGCCGAAATTCGCCGTGGACTCCCT CAGTAAT	Bio-synthesis	N/A
Annealed strand (reverse control): (Biotin)ATTACTGAGGGAGTCCAC GGCGAATTTCGGCCCCTTCTGCG CGAATCGTGGTCGGATCCTCGCC GCAGGCCCTTACTCT	Metabion	N/A
Software and algorithms		
MaxQuant	Cox and Mann[1]	1.6.5.0
R	The R core team	4.2.0

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Falk Butter (f.butter@imb.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD036040.

- All original code has been deposited into the GitHub repository used for the proteomics and STRING database analysis, which is available at https://github.com/mariodejung/DNAdamage_phylointeractome and https://github.com/AFraderaSola/DNADamage_Phylointeracome.

- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All cultivation and growth conditions relevant for *B. subtilis* (DSM10), *E. coli* (DH5α), *H. salinarum* (NRC-1), *S. cerevisiae* (BY4742α), *S. pombe* (pp265), *T. thermophila* (SB210), *C. elegans* (N2), *T. brucei* (Lister 427), *H. sapiens* cell lines (HeLa and HEK293), *A. thaliana* and *Z. mays* are included within the 'method details' section.

METHOD DETAILS

Cultivation and extract preparation

Bacteria: *B. subtilis (DSM10)* and *E. coli (DH5α)* were grown at 37°C in LB medium (IMB media lab) and harvested at OD_{600}=0.7. Cell pellets were resuspended in PBB buffer (150 mM NaCl, 50 mM Tris/HCl pH 8.0, 0.5% Igepal CA-630, 10 mM $MgCl_2$, Pierce protease inhibitor EDTA free) and sonicated with a sonifier 450

(Branson) 3 times for 45 s (cycle=70%, output level 2) with 2-minute breaks. The lysate was centrifuged at 4°C for 15 min at 20,200 x g. The supernatant was supplemented with 10% (f.c.) glycerol (Roth), shock-frozen in liquid nitrogen and stored at −80°C.

Archaea: *H. salinarum* strain NRC-1 was cultivated in Complex Media (4.3 M NaCl, 81 mM $MgSO_4$ x 7 H_2O, 27 mM KCl, 12 mM sodium citrate, 1% w/v oxoid peptone) at 37°C and in light for ∼52h/2.5 days and harvested at OD_{600}=0.5. The cells were pelleted at 3,500 x g for 30 min at 4°C and washed twice in Basic Salt Solution (4.3 M NaCl, 81 mM $MgSO_4$ x 7 H_2O, 27 mM KCl, 12 mM sodium citrate) to remove the medium. After washing, the cells were resuspended in 10 ml of lysis buffer (2.1M NaCl, 50 mM Tris/HCl pH 7.5, 10 mM $MgCl_2$) and sonicated on ice using a Branson 450 sonifier 6 times for 30 s (cycle=50%, output level 2) with 1 min breaks. The sonicated lysate was cleared by centrifugation at 3,500 x g for 30 min at 4°C and supplemented with 10% (f.c.) glycerol (Sigma) before shock-freezing in liquid nitrogen and stored at −80°C.

Yeast: *S. cerevisiae (BY4742a)* was grown in YP medium containing 20% glucose (IMB media lab) at 37°C until OD_{600} = 0.5 and harvested by centrifugation at 20,200 x g. *S. pombe (pp265)* was cultivated in YES media at 32°C until OD_{600} = 1.0 and harvested by centrifugation. For both species, cells were lysed using 0.5 mm zirconia glass beads (Roth) in lysis buffer (100 mM NaCl, 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 0.01% Igepal CA-630, 1x PMSF) at 4°C with 3 cycles alternating between 30 s milling and 30 s cooling using a FastPrep-24 system (MP Biomedicals). The supernatant was transferred to a new tube, shock-frozen in liquid nitrogen and stored at −80°C.

T. thermophila: A mid-log SB210 culture of 3x10^7 cells was grown in 2% proteose peptone (BD Biosciences), 0.2% yeast extract (BD Biosciences), 12 μM ferric chloride, and 1x penicillin/streptomycin/fungizone (Hy-Clone) at 30 °C at 100–120 rotations per minute. Cells were pelleted at 1,500 x g for 3 minutes and washed in 10 mM Tris-HCl pH 7.4. Cells were transferred to a 1.5 ml centrifuge tube and centrifuged at 1,500 x g for 2 min, and the supernatant was removed. Cells were resuspended in 1.2 ml lysis buffer (350 mM NaCl, 40 mM Hepes pH 7.5, 1% Triton X-100, 10% glycerol, freshly added 1 mM DTT, and 1x complete protease inhibitors [Roche]), and approximately 200 μl zirconia glass beads (Roth) were added and vortexed for 3 minutes at 4°C. The tube was centrifuged at ≥16,000 x g at 4°C for 5 min, and the supernatant was transferred to a new tube. The sample was centrifuged at ≥16,000 x g at 4°C for 15 minutes. The supernatant was transferred to a new tube, shock-frozen in liquid nitrogen and stored at −80°C.

C. elegans: Nuclear extraction was performed with N2 gravid adult worms as in.[50] Worms were synchronized and grown on egg plates until they reached the gravid adult stage. Then, worms were washed with M9 buffer 3 times, pelleted, and frozen into pellets in Extraction Buffer (40 mM NaCl, 20 mM MOPS pH 7.5, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 10% glycerol, 2 mM DTT, and 1x complete protease inhibitors, Roche). Pellets were ground into a fine powder with a mortar and pestle. The powder was transferred to a precooled glass douncer (Kimble), and the samples were ruptured with piston B over 30 strokes. The debris was cleared twice at 200 x g for 5 minutes at 4°C. The nuclear pellet was isolated by centrifuging at 2,000 x g for 5 minutes at 4°C. This pellet was washed in extraction buffer twice. The nuclear pellet was resuspended in 200 μL Buffer C+ (420 mM NaCl, 20 mM Hepes/KOH pH 7.9, 2 mM $MgCl_2$, 0.2 mM EDTA, 20% glycerol, and freshly added 0.1% Igepal CA-630, 0.5 mM DTT, 1x complete protease inhibitors [Roche]). The lysate was centrifuged at 4°C for 15 min at 20,200 x g. The supernatant was supplemented with 10% (f.c.) glycerol (Roth), shock-frozen in liquid nitrogen and stored at −80°C.

Plants: *Z. mays* and *A. thaliana* (Columbia) were ground, frozen in liquid nitrogen and transferred to a liquid nitrogen precooled 50 ml steel container for cryomilling with an MM400 (Retsch) at 30 Hz for 4 min. *Z. mays* powder was resuspended in 35 ml PBB buffer (150 mM NaCl, 50 mM Tris/HCl pH 8.0, 0.5% IGEPAL-CA630, 10 mM $MgCl_2$, Pierce protease inhibitor EDTA free) and incubated on ice for 10 min. For *A. thaliana*, powder was resuspended in 30 ml Buffer A (10 mM Hepes KOH pH 7.9, 1.5 mM $MgCl_2$ 10 mM KCl), incubated on ice for 10 min, and subsequently dounced with 40 strokes in a glass douncer using pestle B (Kimble). After centrifugation at 3,640 x g at 4°C, the pellet was washed with 1x DPBS (Gibco), centrifuged again and incubated in 4-6 ml Buffer C+ (420 mM NaCl, 20 mM Hepes/KOH pH 7.9, 2 mM $MgCl_2$, 0.2 mM EDTA, 20% glycerol, and freshly added 0.1% Igepal CA-630, 0.5 mM DTT, 1x complete protease inhibitors [Roche]) for 1 hour at 4°C on a rotation wheel. Cell fragments were removed by centrifugation at 20,200 x g and 4°C for 60 min. The supernatant was shock-frozen in liquid nitrogen and stored at −80°C.

91

Cultured cells: HeLa and HEK293 cells were grown in DMEM (Gibco) with 10% FBS (Gibco) and PennStrep (Sigma) at 37°C with 75% relative humidity and 5% CO_2 in an incubator (Thermo). Cells were harvested, washed in 1x DPBS (Gibco), resuspended in buffer A (10 mM Hepes KOH pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl) and incubated on ice for 10 min. Cells were centrifuged at 500 x g for 5 min and resuspended in Buffer A+ (10 mM Hepes KOH pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, Roche protease inhibitor EDTA free, 0.1% Igepal CA-630, 0.5 mM DTT) and then dounced with 40 strokes in a glass douncer using pestle B (Kimble). Cells were centrifuged at 2,640 x g for 15 min, and the cell pellet was washed with 1x DPBS (Gibco) prior to in-cubation of the pellet in buffer C+ (420 mM NaCl, 20 mM HEPES/KOH pH 7.9, 2 mM $MgCl_2$, 0.2 mM EDTA, 20% glycerol, and freshly added 0.1% Igepal CA-630, 0.5 mM DTT, 1x complete protease inhibitors [Roche]) for 1 hour at 4°C on a rotation wheel. Cell fragments were removed by centrifugation at 20,200 x g and 4°C for 60 min. The supernatant was shock-frozen in liquid nitrogen and stored at −80°C.

DNA pull-down experiments

Chemically synthesized oligonucleotides (Table S1) were ordered HPLC-purified from BioSynthesis (Lewis-ville) and Metabion (Planegg). For pull-downs, 1 nmol of single-stranded DNA lesion (or nondamaged control) oligonucleotide was annealed with 1 nmol of 5'-biotinylated counterstrand with annealing buffer (20 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 100 mM KCl) by first heating to 85°C for 5 min and slowly cooling to RT. The double-stranded oligonucleotides were immobilized on 250 μg streptavidin Dynabeads C1 (Thermo) and incubated with different amounts of protein extract ranging from 200-1,000 μg (200 μg: *C. elegans*, *Z. mays* and *A. thaliana*; 400 μg: HEK293 and HeLa; 500 μg: *H. salinarum*, *T. thermophila*; 800 μg: *S. cerevisiae* and 1,000 μg: *B. subtilis*, *E. coli*, *S. pombe* and *T. brucei*) in 1x PBB buffer (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 0.5% Igepal CA-630, 5 mM $MgCl_2$ and 1x protease inhibitor cocktail [Roche]) rotating at 4°C for 90 min. Protein concentrations were determined using Protein Assay Dye Reagent (Bio-Rad). All samples were prepared in quadruplicate. After incubation, unbound proteins were removed by 3 washes with PBB buffer. The Dynabeads were ultimately resuspended in 25 μl 1x LDS (Thermo) containing 100 mM DTT (Sigma) and heated to 70°C for 10 min.

Mass spectrometry sample preparation

LDS supernatant was loaded on a 4-10% NuPage NOVEX PAGE gel (Thermo) and run for 10 min at 180 V. Samples were processed as previously described.[51] In short, gel pieces were cut, destained with 50% EtOH/50 mM ammonium bicarbonate (ABC), dehydrated with acetonitrile (VWR), reduced with 10 mM DTT (Sigma), alkylated using iodoacetamide (Sigma) and subsequently again dehydrated with acetonitrile (VWR) and digested with 1 μg of MS-grade trypsin (Sigma) at 37°C overnight. The peptides were eluted from the gel pieces, loaded onto a StageTip[52] and stored at 4°C until measurement.

Mass spectrometry measurement

Peptides were eluted from the StageTips using 80% acetonitrile/0.1% formic acid and concentrated prior to loading either on an uHPLC nLC-1000 system coupled to a Q Exactive Plus mass spectrometer (Thermo) or an uHPLC nLC-1200 system coupled to an Exploris 480 mass spectrometer (Thermo). The peptides were loaded on a 20 cm (Q Exactive Plus) or 50 cm (Exploris 480) column (75 μm inner diameter) in-house packed with Reprosil C18 (Dr. Maisch GmbH) and eluted with a 73- or 88-min optimized gradient increasing from 2% to 40% mixture of 80% acetonitrile/0.1% formic acid at a flow rate of 225 nl/min or 250 nl/min. The Q Exactive Plus was operated in positive ion mode with a data-dependent acquisition strategy of one MS full scan (scan range 300 - 1,650 m/z; 70,000 resolution; AGC target 3e6; max IT 20 ms) and up to ten MS/MS scans (17,500 resolution; AGC target 1e5, max IT 120 ms; isolation window 1.8 m/z) with peptide match preferred using HCD fragmentation. The Exploris 480 was operated in positive ion mode with a data-dependent acquisition strategy of one MS full scan (scan range 300 - 1,650 m/z; 60,000 resolution; normalized AGC target 300%; max IT 28 ms) and up to twenty MS/MS scans (15,000 resolution; AGC target 100%, max IT 40 ms; isolation window 1.4 m/z) with peptide match preferred using HCD fragmentation.

Mass spectrometry data analysis

MaxQuant (Version 1.6.5.0) was used to search and quantify the raw mass spectrometry files for each spe-cies individually.[53] Individual protein databases used as search space for MaxQuant can be found in Table S3. Oxidation and acetylation were set as variable modifications, and carbamidomethylation was set as a fixed modification. Label-free quantification (LFQ) was used to calculate and normalize intensities without activating fast LFQ. The minimum ratio count used was 2. Match between runs was used to match

within each lesion (control, abasic, 8-oxoG, RNA base), with a match time window of 0.7 min, match ion mobility window of 0.05, alignment time window of 20 min, and alignment ion mobility of 1. Matching of unidentified features was deactivated. For protein quantification, we used a label minimum ratio count of 2 and unique + razor peptides for quantification.

Bioinformatics analysis and statistical analysis

MaxQuant proteinGroup results files of all species were combined into a single file, with a column "species" indicating the individual species and cell type (Table S4). The complete dataset was filtered by removing reverse database binders, potential contaminants or proteins identified only on a modification site. Additionally, all protein groups with fewer than 2 peptides (1 unique) were filtered out. Missing LFQ values were treated as if they were below the detection limit of the mass spectrometer. Imputation was performed for each replicate of a condition individually from a beta distribution, within a range of the 0.2 and 2.5 percentile of measured intensities of the replicate. Only proteins that were present in ≥ 2 replicates of 4 per pull-down condition were used to calculate enrichment values ($\log_2$ fold change, p value by Welch t-test) (Table S4). Gene information and annotations were downloaded[54,55] and used to assign detected proteins to orthology groups, as per OrthoMCL.[20] Labeling of specific orthology groups for Figure 6 was performed based on the following hierarchy of species: hsap, scer, spom, cele, ecol, atha, bsub, halo, tbrt, tetr, and zmay. In other words, if an orthology group contained a human gene, it would be referred to as this. If not, the *S. cerevisiae* gene was taken, and so forth according to the listed hierarchy. If multiple genes of one species were present in the orthology group, the first one from the list was selected. Heatmap clustering was performed on a numerical matrix, where 1 was an enriched protein ($\log_2$-fold change >2, p value <0.05), 0 a detected protein (i.e., not enriched but measured), and -1 a protein not detected within a species at all. To find similar clusters of proteins, we applied the complete linkage method (default setting) in hclust from the stats package in the R framework.[56] For functional enrichment analysis, terms were queried in the Gene Ontology (GO)[22] and the Kyoto Encyclopedia of Genes and Genomes (KEGG)[57] databases. Terms for a particular group of enriched proteins were tested for overrepresentation (adjusted p value [FDR] < 0.05; Fisher's exact test) against all terms found in the background (whole genome). The top three most overrepresented terms in each database were selected for graphical representation. To determine known and predicted interactions, enriched proteins were queried in the STRING database version 11.5.[23] Hits from text mining and co-occurrence interaction sources were excluded. Hits with a score >150 in any of the remaining interaction sources (experiments, databases, coexpression, gene fusion and neighborhood) were included in the downstream analysis. Thus, protein-protein networks were generated with in-house scripts based on an R framework incorporating igraph,[58] with the Fruchterman-Reingold force-directed layout algorithm implementation, and ggnetwork.[59] Enriched proteins were illustrated as nodes, where color indicates their associated experimental lesion and their shape indicates whether they are known repair proteins or not. STRING known and predicted interactions were visualized as edges. All networks were drawn with the spoke model.

For phylogenetic tree construction, the amino acid sequences of all orthologs from the respective OrthoMCL groups were extracted from the species-specific protein sequence FASTA files (Table S3). For AP endonucleases, the OrthoMCL groups OG6_101139 and OG6_104339 were chosen to represent the group. OG6_104135 and OG6_100453 contain the Photolyase family, and OG6_102506 contains the MutY Glycosylase family. The evolutionary history was inferred by using the maximum likelihood method and JTT matrix-based model.[60] The tree with the highest log likelihood is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model and then selecting the topology with superior log likelihood value. Evolutionary analyses were conducted in MEGA X.[61]

Pfam analysis for proteins with no previous DNA repair associations was conducted using Pfam domain annotations downloaded from OrthoMCL.[20] To enable broader categorizations, Pfam terms were classified into more general terms based on text mining of the Pfam term description (Table S12). These classifiers were used to detect the distribution of Pfam functions across the proteins that have not been previously annotated as DNA repair proteins.

QUANTIFICATION AND STATISTICAL ANALYSIS

All quantification and statistical analysis details and associated citations can be found in the method details in the 'mass spectrometry data analysis' and 'bioinformatics analysis and statistical analysis' sections. In

short, the pull downs performed in the analysis were performed in quadruplicate, and the p value was determined by Welch's t-test with an enrichment threshold of $\log_2$ fold change >2 and p value <0.05. Utilizing both GO and KEGG databases, enriched proteins were tested for overrepresentation using Fisher's exact test, determining an adjusted p value (false discovery rate) < 0.05. The STRING database was used to determine previously established interactors to proteins of interest. To create phylogenetic trees, the evolutionary history was inferred in MEGA X[61] by using the maximum likelihood method and JTT matrix-based model.[60] Pfam domain annotations were downloaded from OrthoMCL.[20]

Supplemental information

DNA damage repair proteins across the Tree of Life

Emily Nischwitz, Vivien A.C. Schoonenberg, Albert Fradera-Sola, Mario Dejung, Olga Vydzhak, Michal Levin, Brian Luke, Falk Butter, and Marion Scheibe
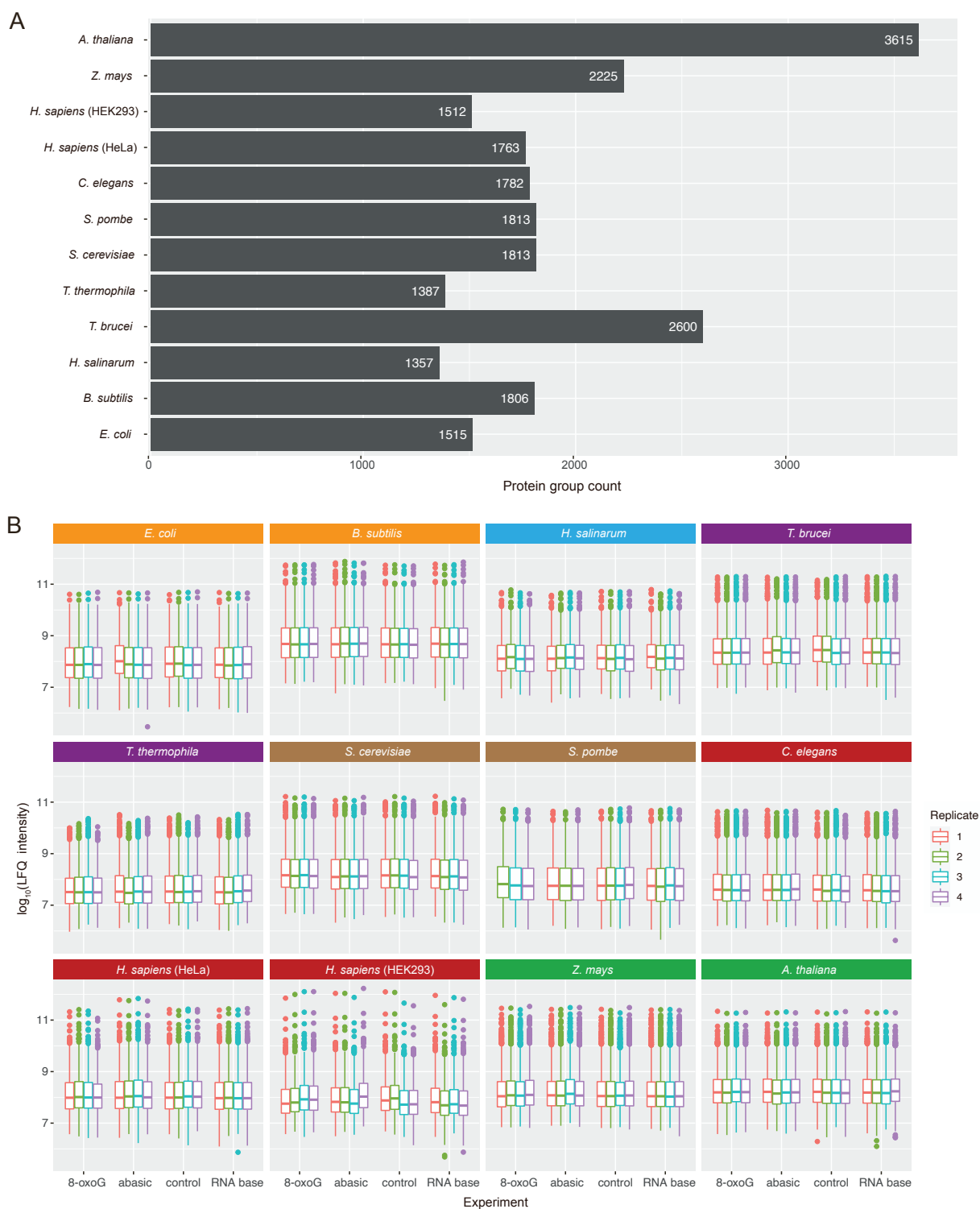
Figure S1. **Protein quantification.** Related to Figure 1. **A)** Bar plot showing the total amount of identified, quantifiable proteins per species (see methods). **B)** Boxplot showing $\log_{10}$ LFQ intensity per replicate (y-axis) and experiment (x-axis) of each included species.
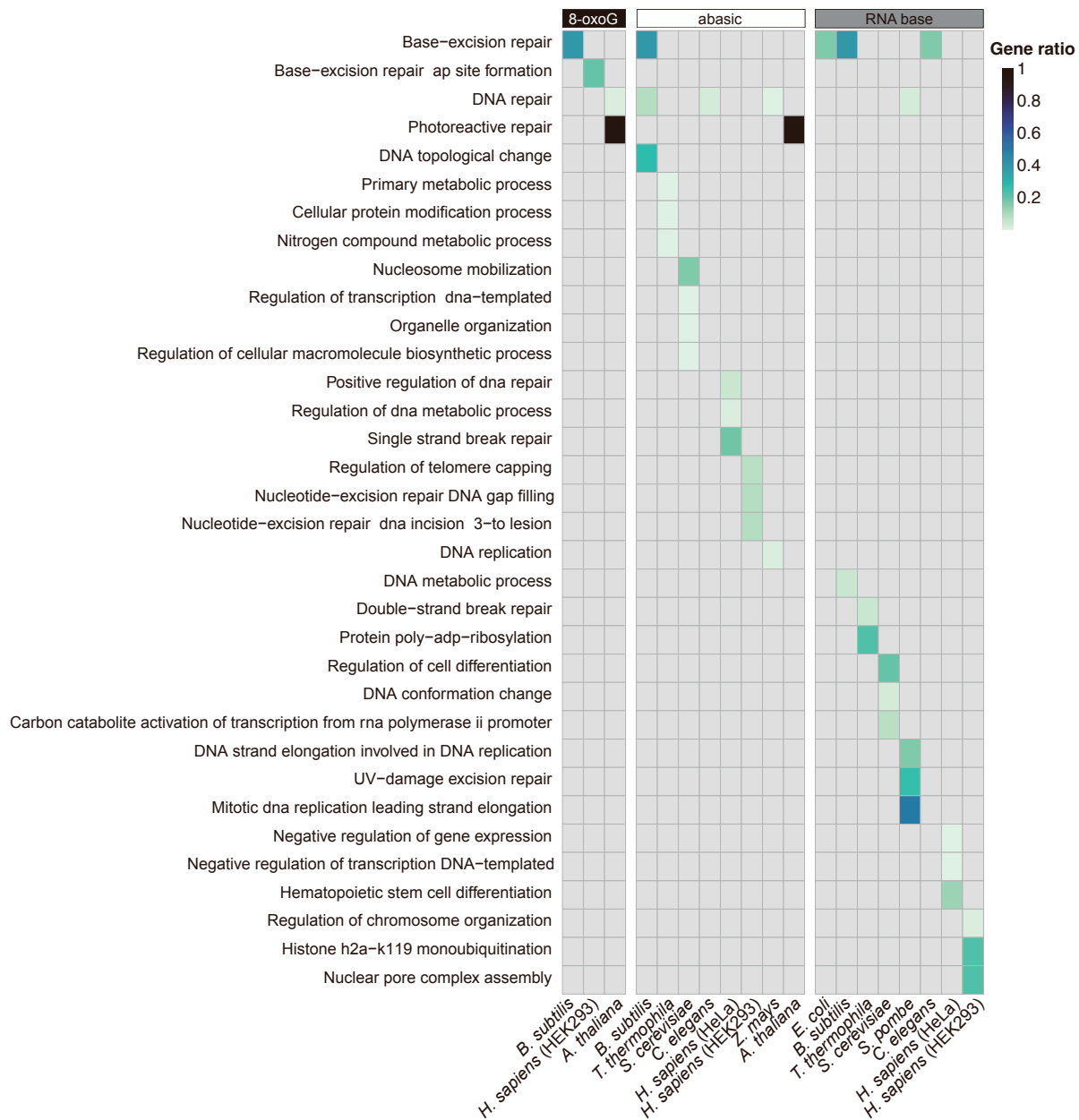
Figure S2. Overrepresentation of GO terms biological processes among enriched proteins. Related to Figure 2. Overrepresentation of GO terms biological processes among enriched proteins at each lesion across species. Conditions with no enriched GO terms are not shown, or presented in gray. 'Gene ratio' refers to genes in the dataset (enriched proteins at lesion) over genes in the background (whole genome).
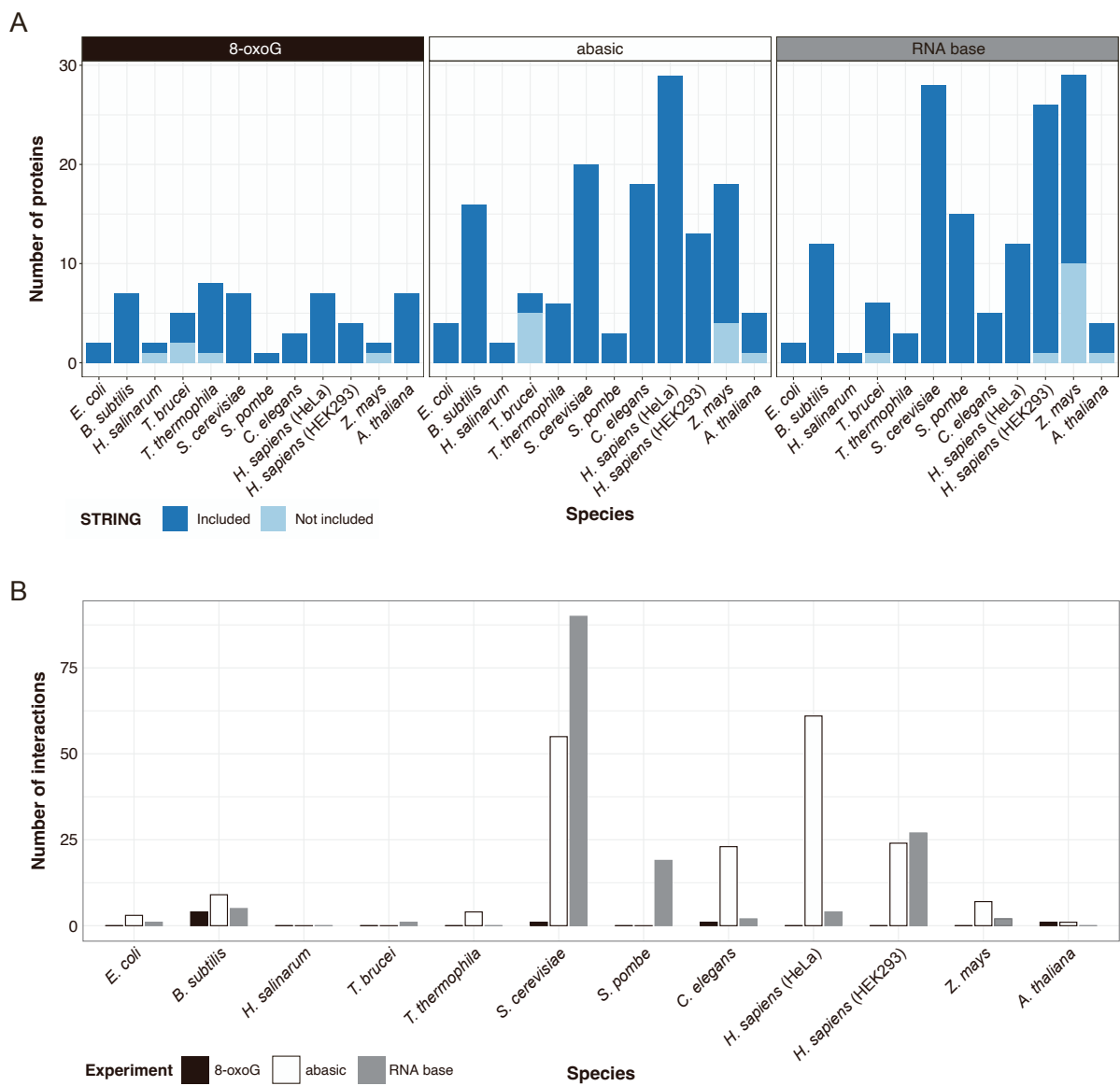
Figure S3. Quantification (STRING) interactions of enriched proteins.
Related to Figure 2. **A)** Bar plot of enriched proteins per lesion and species
that are included in the STRING database. **B)** Bar plot of number of
interactions per lesion and species.

8-oxoG
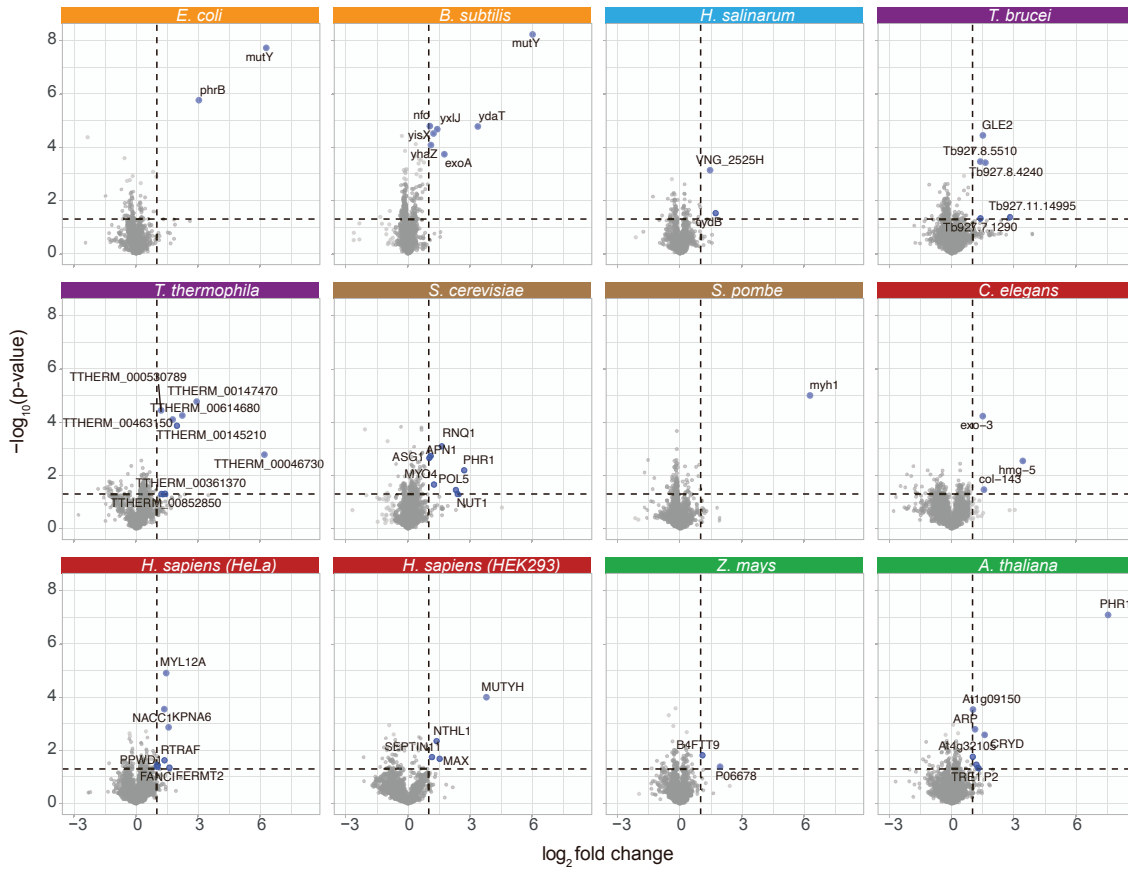
Figure S4. Protein enrichment at 8-oxoG lesion. Related to Figure 3. Volcano plots showing the fold change (x-axis) and p-value (y-axis) of proteins binding to the 8-oxoG lesion compared to the control for each species. Proteins with $\log_2$ fold change > 1 and Welch t-test p-value < 0.05 are highlighted and labeled.
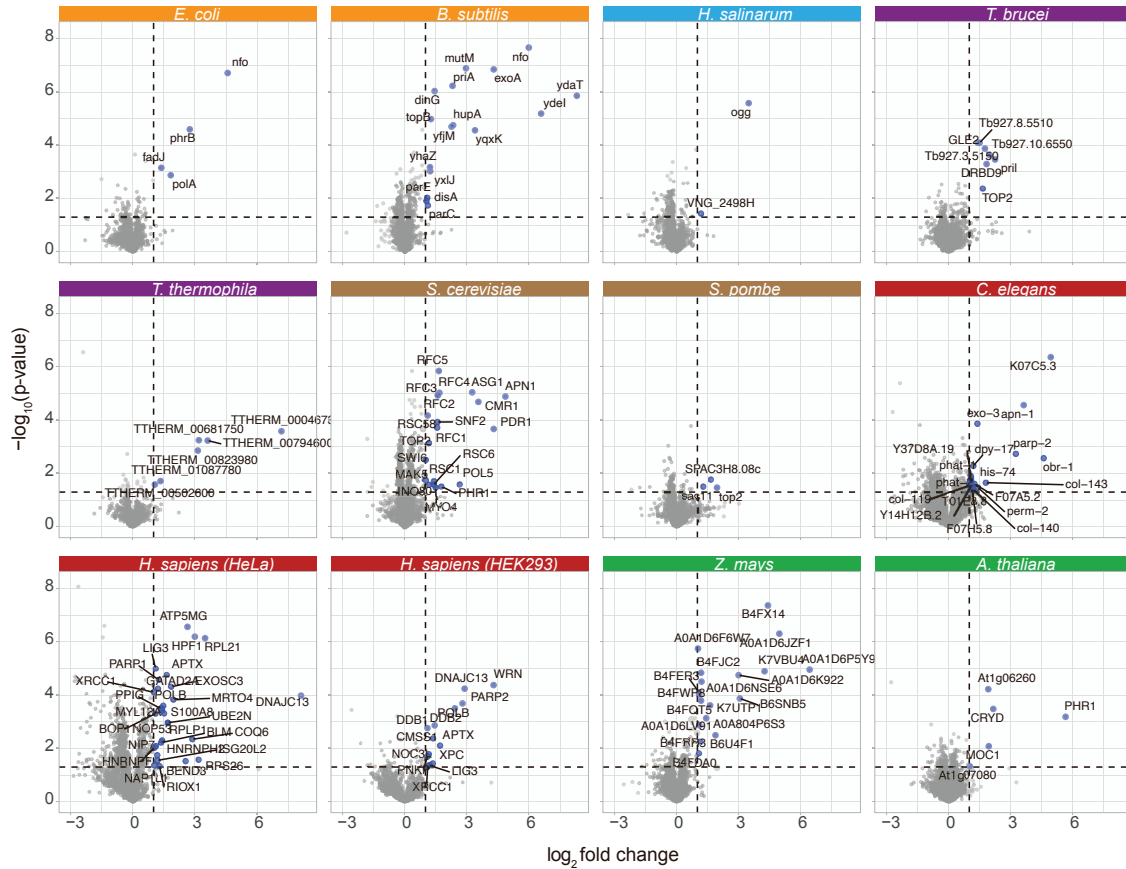
Figure S5. Protein enrichment at abasic lesion. Related to Figure 3. Volcano plots showing the fold change (x-axis) and p-value (y-axis) of proteins binding to the abasic lesion compared to the control for each species. Proteins with $\log_2$ fold change > 1 and Welch t-test p-value < 0.05 are highlighted and labeled.

RNA base
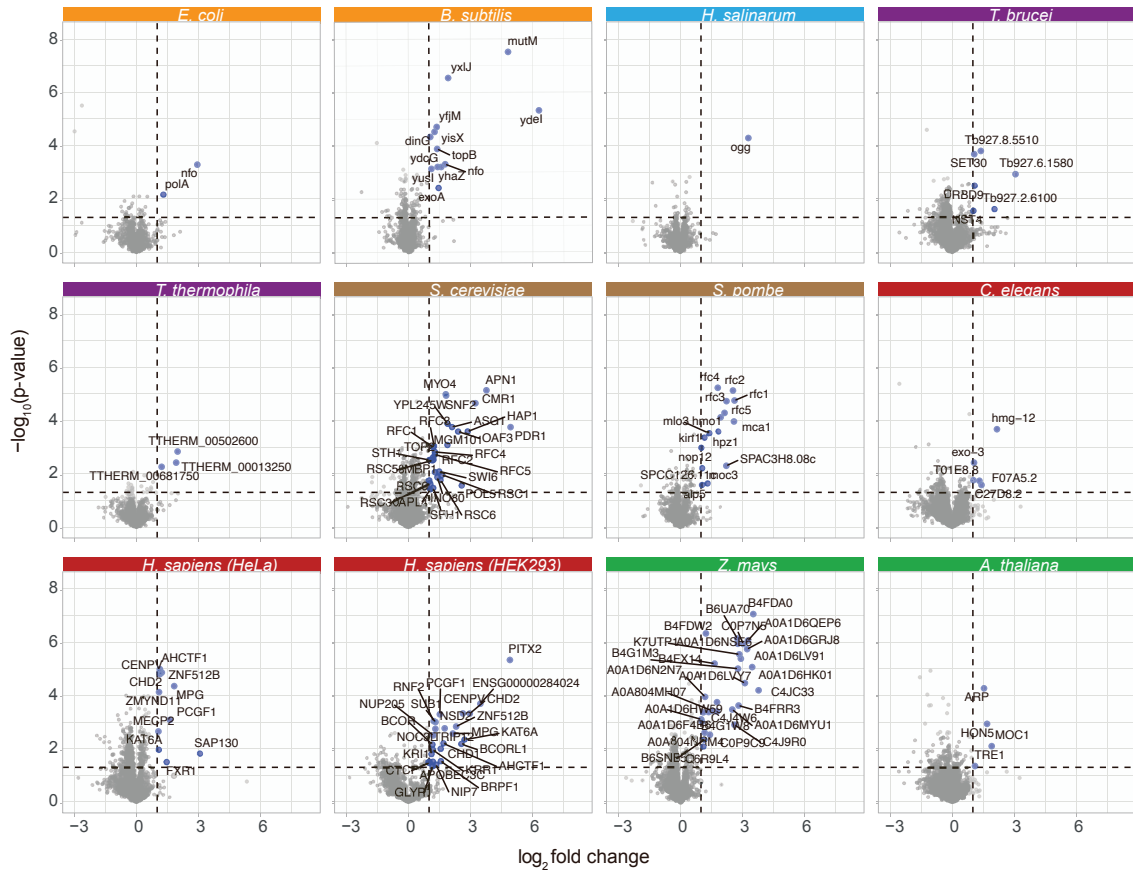
Figure S6. Protein enrichment at RNA base lesion. Related to Figure 3. Volcano plots showing the fold change (x-axis) and p-value (y-axis) of proteins binding to the RNA base lesion compared to the control for each species. Proteins with $\log_2$ fold change > 1 and Welch t-test p-value < 0.05 are highlighted and labeled.
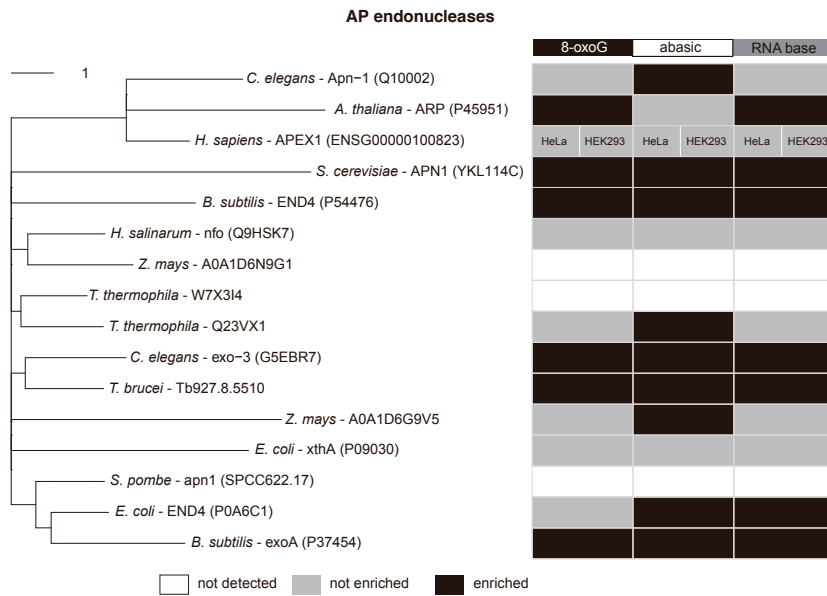
Figure S7. Maximum likelihood phylogenetic tree of AP endonucleases.
Related to Figure 5. Maximum likelihood phylogenetic tree of AP endonucleases within the scer_APN1 and hsap_APEX1 orthology groups. Includes information on detection and enrichment (fold change > 2, Welch t-test p-value < 0.05) for the different lesions. White boxes represent proteins that were not detected in the respective experiment. The scale bar in the plots indicates the number of amino acid substitutions per site.
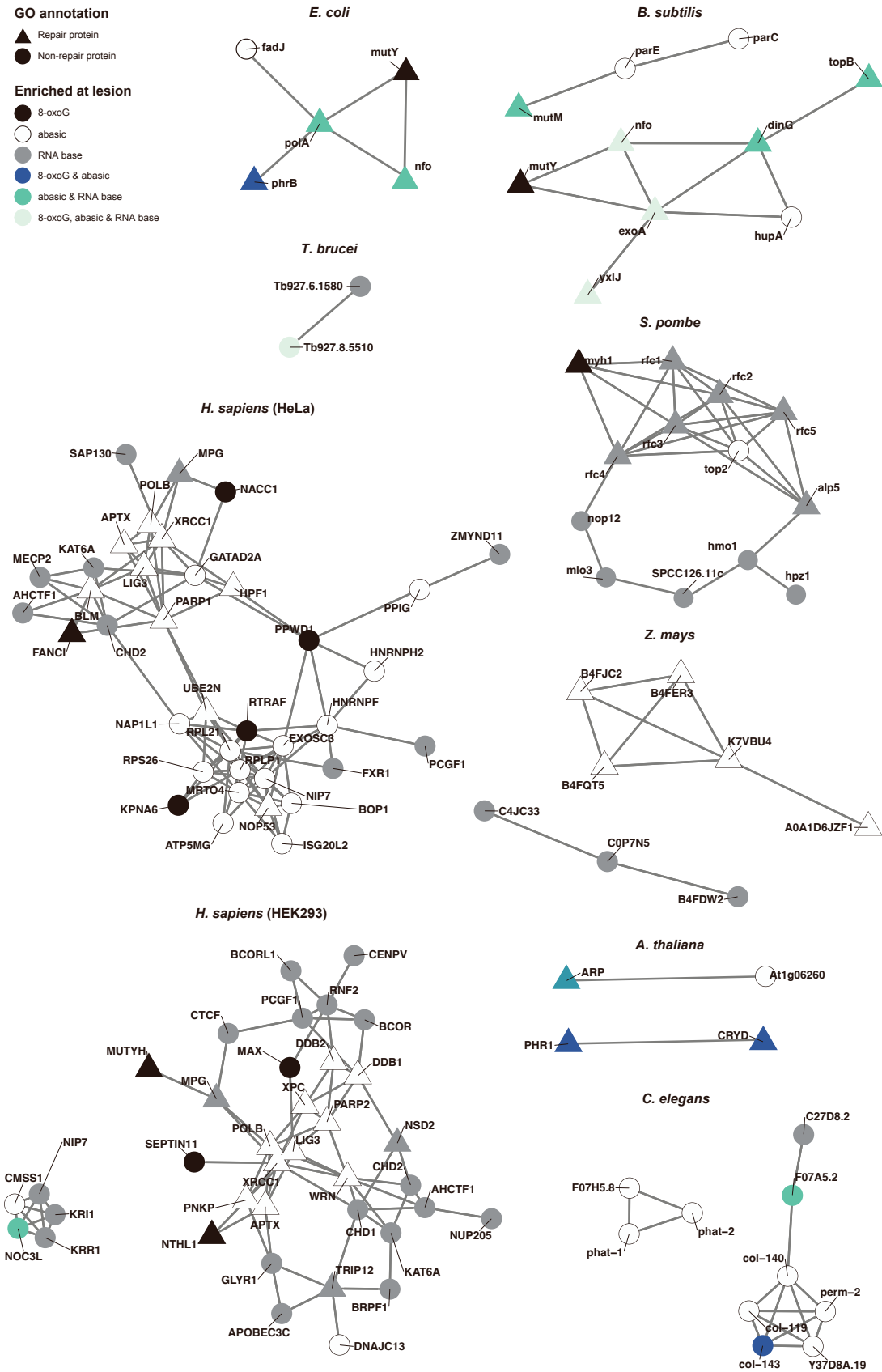
Figure S8. Networks of enriched proteins across lesions. Related to Figure 6. Networks of enriched proteins across lesions for *E. coli, B. subtilis, T. brucei, S. pombe*, *H. sapiens* (HeLa and HEK293), *Z. mays, A. thaliana*, and *C. elegans*. Interactions as reported in the STRING database.
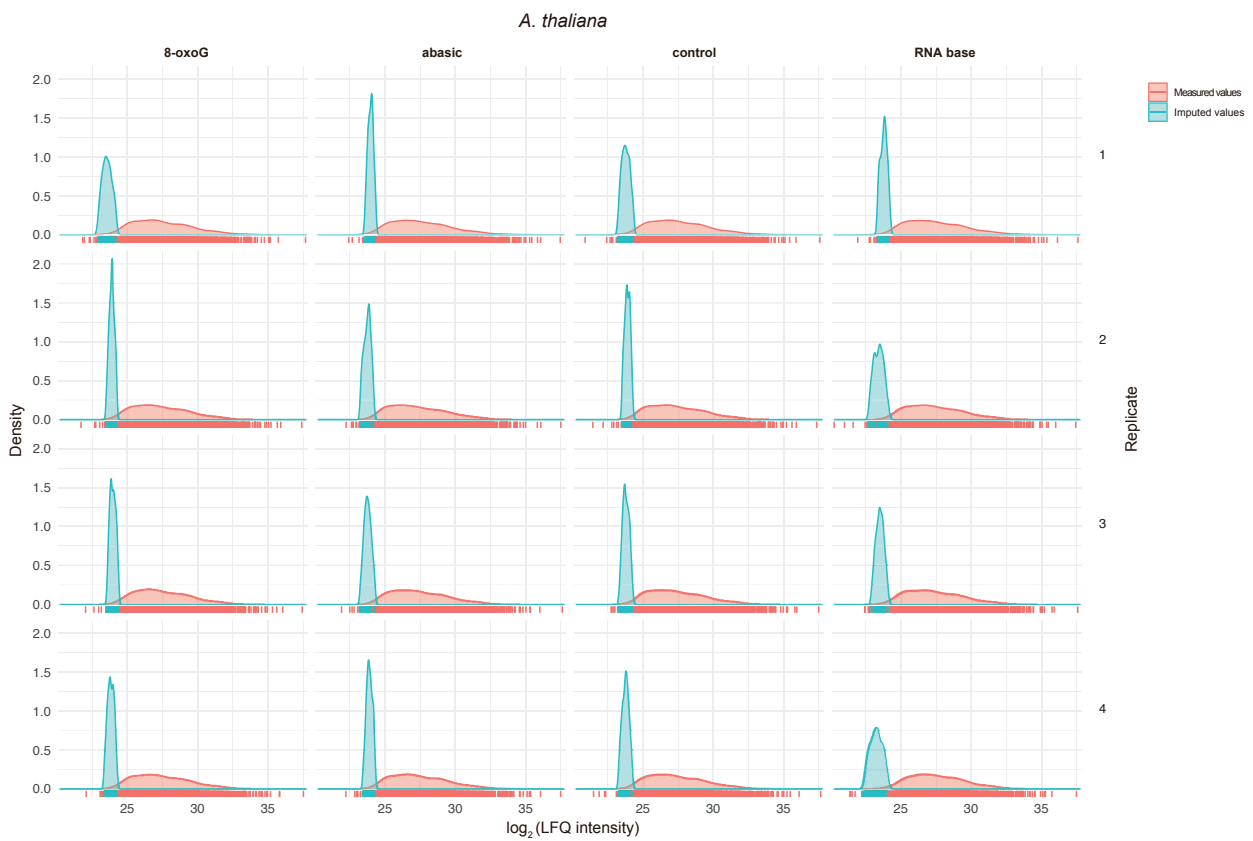
Figure S9. Example imputation values. Related to STAR methods. Density plot showing representative example (*A. thaliana*) of distribution of imputed $\log_2$ LFQ intensity values. Imputation was done for each replicate of the experiment (8-oxoG, abasic, control, RNA base), from a beta distribution within a range of the 0.2 and 2.5 percentile of measured intensities of the replicate (see methods). Density of imputation values shown in light blue, original measured intensities in red.

Table S1: Oligo sequences used in this study. Related to Figure 1 and STAR methods.

Name of Oligo	5` to 3` Sequence
Control without lesion	AGAGTAAGGGCCTGCGGCGAGGATCCGACCACGATTCGCGCAGAAGGGGCCGAAATTCGCCGTGGACTCCCTCAGTAAT
8-oxoG lesion	AGAGTAAGGGCCTGCGGCGAG(8-Oxo-dG) ATCCGACCACGATTCGCGCAGAAGGGGCCGAAATTCGCCGTGGACTCCCTCAGTAAT
abasic lesion	AGAGTAAGGGCCTGCGGCGAG(dSpacer)ATCCGACCACGATTCGCGCAGAAGGGGCCGAAATTCGCCGTGGACTCCCTCAGTAAT
RNA lesion	AGAGTAAGGGCCTGCGGCGAG(rU)ATCCGACCACGATTCGCGCAGAAGGGGCCGAAATTCGCCGTGGACTCCCTCAGTAAT
Annealed strand (reverse control)	(Biotin)ATTACTGAGGGAGTCCACGGCGAATTTCGGCCCCTTCTGCGCGAATCGTGGTCGGATCCTCGCCGCAGGCCCTTACTCT

Table S3: Overview of the databases used. Related to STAR methods.

SPECIES	ABBREV	RESOURCE_NAME	RESOURCE_URL	NO_SEQUENCES
A. thalania	atha	Uniprot	https://www.uniprot.org/proteomes/UP000006548	39328
B. subtilis	bsub	Uniprot	https://www.uniprot.org/proteomes/UP000001570	4260
C. elegans	cele	Uniprot	https://www.uniprot.org/proteomes/UP000001940	26548
E. coli	ecol	Uniprot	https://www.uniprot.org/proteomes/UP000000625	4448
H. salinarum	halo	Uniprot	https://www.uniprot.org/proteomes/UP000000554	2426
H. sapiens	hsap	Ensembl (version 102)	http://nov2020.archive.ensembl.org/index.html	113656
S. cerevisiae	scer	FungiDB	https://fungidb.org/fungidb/app/downloads/Current_Release/ScerevisiaeS288C/fasta/data/	5907
S. pombe	spom	FungiDB	https://fungidb.org/fungidb/app/downloads/Current_Release/Spombe972h/fasta/data/	5139
T. thermophila	tetr	Uniprot	https://www.uniprot.org/proteomes/UP000009168	26972
T brucei	tbrt	TriTrypDB	https://tritrypdb.org/tritrypdb/app/downloads/Current_Release/TbruceiTREU927/fasta/data	9788
Z. mays	zmay	Uniprot	https://www.uniprot.org/proteomes/UP000007305	137157
Orthology groups		OrthoMCL	https://orthomcl.org/common/downloads/release-6.8/groups_OrthoMCL-6.8.txt.gz	("current", downloaded Jan 2022)

Table S11: Overlap of interaction partners across lesions. Related to Figure 4.

Species	RNA base and abasic	RNA base and 8-oxoG	abasic and 8-oxoG	all 3 lesions
E. coli	polA, nfo		phrB	
B. subtilis	yfjM, mutM, dinG, topB, ydeI	yisX	ydaT	yhaZ, exoA, nfo, yxlJ
H. salinarum	ogg			
T. brucei	DRBD9		GLE2	Tb927.8.5510
T. thermophila	PCP1, PARP6		PHR2	
S. cerevisiae	RFC5, RSC6, CMR1, PDR1, INO80, RSC1, SNF2, RFC2, RSC58, SWI6, TOP2, RFC3, RFC4, RFC1		PHR1	MYO4, POL5, ASG1, APN1
S. pombe	SPAC3H8.08c			
C. elegans	F07A5.2, T01E8.8		col-143	exo-3
H. sapiens (HeLa)			MYL12A	
H. sapiens (HEK293)	NOC3L			
Z. mays	A0A1D6LV91, A0A1D6NSE6, B4FDA0, B4FRR3, B4FX14, B6SNB5, K7UTP1			
A. thaliana	MOC1	ARP, TRE1	CRYD, PHR1	

A systems view on DNA damage response kinetics in *Tetrahymena thermophila*

Summary

Here, we took a systems view of the kinetics of the DNA damage response in *Tetrahymena thermophila*. Prior to the 8-hour time course, we exposed Tetrahymena to six common DNA mutagens. The damaging agents used were ultraviolet light (inducing nucleotide excision repair), hydrogen peroxide (inducing base and nucleotide excision repair), methyl methanesulfonate (inducing base and nucleotide excision repair; Campbell and Romero, 1998, Smith, et al., 2004), hydroxyurea (halting replication; Sandoval et al., 2015), ionizing radiation (inducing double-stranded break repair; Loidl and Mochizuki, 2009), and cisplatin (inducing nucleotide excision repair and intercrosslink repair; Loidl and Mochizuki, 2009). Over the time course, we collected protein and RNA samples at 0, 1, 2, 3, 4, 6, and 8 hours. The protein samples were later processed with in-gel digestion and measured using label-free quantitative mass spectrometry. RNA was isolated, libraries were prepared that were subsequently using QuantSeq.

To assess whether our treatments induced the expected DNA repair pathways, we initially examined the dynamics of known DNA damage throughout the time course. All six treatments elicited both pathway-specific and generalized DNA damage response. To comprehensively evaluate the dataset, we evaluated all transcripts and proteins for significant fold change and dynamicity score. We employed self-organizing maps, an unsupervised machine learning approach, to cluster all dynamic genes and proteins. Within both the RNA and protein clusters, we observed co-clustering of large known complexes. The majority of RNA clusters showed enrichment for 'protein phosphorylation', indicating the diverse but overall crucial role of protein phosphorylation in the DDR of *Tetrahymena*. We observed intriguing conservation and divergence of transcript and protein regulation within certain gene families, which we will further explore through correlation data analysis. We are currently establishing knockdown cell lines to investigate some of these dynamics in general DNA damage responders and members of the PARP family, which exhibited unique responses to each DNA damaging agent.

This data set provides a unique view into DNA damage kinetics, and we hope that it will serve as a valuable resource for future investigations in *Tetrahymena*. This data set and analysis will be available at https://butterlab.imb-mainz.de/Tt_DDR/.

Zusammenfassung

Hier haben wir die Kinetik der DNA-Schadensantwort (DNA damage response, DDR) in *Tetrahymena thermophila* aus einer systemischen Perspektive untersucht. Vor dem achtstündigen Zeitverlauf setzten wir Tetrahymena sechs gängigen DNA-Mutagenen aus. Dabei handelte es sich um ultraviolettes Licht (induziert die Nukleotid-Exzisionsreparatur), Wasserstoffperoxid (induziert die Basen- und Nukleotid-Exzisionsreparatur), Methylmethansulfonat (induziert die Basen- und Nukleotid-Exzisionsreparatur; Campbell und Romero, 1998, Smith, et al, 2004), Hydroxyharnstoff (stoppt die Replikation; Sandoval et al., 2015), ionisierende Strahlung (induziert die Reparatur von Doppelstrangbrüchen; Loidl und Mochizuki, 2009) und Cisplatin (induziert die Nukleotid-Exzisionsreparatur und die Intercrosslink-Reparatur; Loidl und Mochizuki, 2009). Während des Zeitverlaufs entnahmen wir Protein- und RNA-Proben nach 0, 1, 2, 3, 4, 6 und 8 Stunden. Die Proteinproben wurden daraufhin mit einem In-Gel-Verdau verarbeitet und mit markierungsfreier quantitativer Massenspektrometrie gemessen. Die RNA wurde isoliert und für die Herstellung von Bibliotheken verwendet, die anschließend mit QuantSeq analysiert wurden.

Um zu beurteilen, ob unsere Behandlungen die erwarteten DNA-Reparaturwege induzierte, untersuchten wir zunächst die Dynamik der bekannten DNA-Schäden im Zeitverlauf. Alle sechs Behandlungen lösten sowohl pfadspezifische als auch allgemeine DNA-Schadensreaktionen aus. Um den Datensatz umfassend auszuwerten, untersuchten wir alle Transkripte und Proteine auf signifikante Fold Changes und Dynamicity Score. Mit Hilfe von selbstorganisierenden Karten und einem unüberwachten maschinellen Lernverfahren wurden alle dynamischen Gene und Proteine geclustert. Sowohl bei den RNA- als auch bei den Protein-Clustern beobachteten wir eine gemeinsame Clusterbildung großer bekannter Komplexe. Die Mehrheit der RNA-Cluster zeigte eine Anreicherung für "Proteinphosphorylierung", was auf die vielfältigeund entscheidende Rolle der Proteinphosphorylierungen in der DDR von *Tetrahymena* hinweist. Wir beobachteten eine faszinierende Erhaltung und Divergenz der Transkript- und Proteinregulation innerhalb bestimmter Genfamilien, die wir durch eine Korrelationsdatenanalyse weiter untersuchen werden. Derzeit etablieren wir Knockdown-Zelllinien, um einige dieser Dynamiken bei allgemeinen DNA-Schadens-Respondern und Mitgliedern der PARP-Familie zu untersuchen, die spezifische Reaktionen auf jedes der DNA-Mutagene zeigten.

Dieser Datensatz bietet einen einzigartigen Einblick in die Kinetik von DNA-Schäden, daher hoffen wir, dass er als wertvolles Hilfsmittel für künftige Untersuchungen in *Tetrahymena* dienen wird. Dieser Datensatz und die Analyse werden unter https://butterlab.imb-mainz.de/Tt_DDR/ verfügbar sein.

Statement of contribution

This project was led by myself and Vivien Schoonenberg, with the support of Falk Butter. I was responsible for the experimental design, implementation, and execution. I guided Rachel Mullner and Susanne Zimbelmann through their experimental contributions to this data set. Vivien led all aspects of the data analysis. Both Vivien and I directed the data visualization and provided in-depth data interpretation. Currently, Vivien and I are writing the initial draft version of the manuscript with support from Falk Butter, and we are completing lingering analysis and experiments. We plan to submit in Fall 2023.

Supervisor Confirmation: _____ *Falk Butter* _____

A systems view on DNA damage response kinetics

Emily Nischwitz[1,4], Vivien A.C. Schoonenberg[1,3,4], Emily Nischwitz[1,4], Rachel Mullner[1,2], Susanne Zimbelmann [1,4], Michal Levin[1], and Falk Butter[1,2,*]

[1] Institute of Molecular Biology (IMB), 55128 Mainz, Germany
[2] Institute of Molecular Virology & Cell Biology, Friedrich Loeffler Institute, 17493, Greifswald, Germany
[3] Present department:
Division of Hematology/Oncology, Boston Children's Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Broad Institute, Harvard Medical School, Boston, MA, 02115, USA
Molecular Pathology Unit & Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114, USA
[4] These authors contributed equally.
[*] Correspondence: falk.butter@fli.de

Abstract

The tightly regulated DNA damage response is critical to the overall integrity of the genome. Here, we combine proteomics and transcriptomics to study DNA damage kinetics across various treatments. We performed a large-scale transcriptome and proteome screen of six well-established treatments, invoking different DNA damage repair pathways in the ciliate Tetrahymena thermophila. This extensive data set of 9 conditions, 7 time points, and over 250 transcriptome and proteome measurements provides the first comprehensive kinetic transcriptome and proteome DNA damage response analysis and resource in eukaryotes to date. We observed upregulation of known DNA repair proteins as well as a global dynamic transcript and protein response. Using unsupervised machine learning clustering, we found different expression profile trends. PARP and PARP-related proteins had damage specific responses. We utilized a novel knockdown system investigating the influence of the PARP family. In addition to the comprehensive analysis presented here, the data can be explored via an accessible user interface at https://butterlab.imb-mainz.de/Tt_DDR/. Ultimately, our study will help identify novel candidates to study in the DNA damage response, as well as provide new insights on current proteins of interest.

1. Introduction

Environmental genotoxic stressors create DNA damage that pose a threat to the stability and integrity of the genome. It is therefore critical to have a carefully regulated orchestra of DNA damage response factors and pathways (Ciccia and Elledge 2010). Primary DNA repair pathways include nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR), homologous recombination (HR), non-homologous end joining (NHEJ), and interstrand crosslink repair (ICL) (Schärer 2013; Beard et al. 2019; Z. Li, Pearlman, and Hsieh 2016; Chapman, Taylor, and Boulton 2012; Deans and West 2011). DNA damage repair activity is required in all living organisms, and the dysregulation of any of these pathways has been correlated with disease (Kovalchuk 2016; Jackson and Bartek 2009).

Exogenous mutagens can induce damage lesions that associate with particular repair pathways. UV exposure typically results in pyrimidine (6-4) pyrimidone photoproducts ((6-4)PPs) and cis-syn cyclobutane pyrimidine dimers (Spivak 2015), repaired by NER. Cisplatin (CPT) causes covalent bonds between base pairs on different DNA strands referred to as interstrand crosslinks (ICLs) (H. Huang et al. 1995); this damage is often repaired by NER, however there is a cell cycle-dependent compilation of various repair pathways to address this damage, including HR, NER, translesion synthesis (TLS), and the fanconi anemia (FA) pathway (Duan et al. 2020; Deans and West 2011). Hydrogen peroxide (HP) and methyl methanesulfonate (MMS) cause oxidative and alkylative damage, respectively (Ransy et al. 2020). Previously it was thought that BER was the primary repair pathway for both these damages, however there is growing evidence of the interdependence of BER and NER for the repair of these lesions (Fayyad et al. 2020;

Kumar et al. 2022). IR causes direct DSBs, which are repaired by either homologous recombination (HR) and non-homologous end joining (NHEJ) (Chapman, Taylor, and Boulton 2012). In addition to these direct DSBs, a large amount of reactive oxygen species are created by IR which can be repaired by BER. If two nearby lesions are simultaneously excised by BER repair proteins, this can lead to an inadvertent DSB. HU does not damage DNA through direct interaction with DNA or creation of a deleterious byproduct, but rather inhibits the enzyme ribonucleotide reductase (RR). The inhibition of RR drastically reduces the available amount of deoxynucleotide triphosphate pools, in turn inhibiting replication (Agrawal et al. 2014). This leads to stalled replication forks which ultimately can lead to DSBs (Petermann et al. 2010), then repaired by DSBR.

There are a limited number of studies addressing DNA damage repair from a global transcriptomic and proteomic perspective (von Stechow and Olsen 2017; Stokes, Zhu, and Farnsworth 2018). Here, we use *Tetrahymena thermophila* (*Tetrahymena*) (Howard-Till et al. 2022). One of the most exceptional phenotypes of Tetrahymena is its unique nuclear architecture containing a macronucleus (MAC), and a transcriptionally silent micronucleus (MIC), passed during sexual reproduction. During meiosis in *Tetrahymena* the ~45N 180 MAC chromosomes undergo ~200 regulated double-stranded DNA breaks (Loidl 2021).

Even within well-studied unicellular eukaryotic, such as *S. cerevisiae*, there have been few studies of the proteomic or transcriptomic response to DNA damage, and far fewer which have addressed the temporal response to DNA damage (D. R. Kim et al. 2011; Kubota et al. 2012; Rodríguez-Lombardero et al. 2014; Zhou et al. 2016; Suter, Auerbach, and Stagljar 2006; Hauser et al. 2019; Bharati, Kumari, and Akhtar 2020). To obtain a systematic comparative overview of the kinetics of the DNA damage repair in a eukaryotic organism, we performed transcriptome and proteome measurement over 8 hours, with six well-established genotoxic treatments invoking different DNA damage repair pathways.

2. Results

1.1 Known DNA damage repair factors are differentially regulated in response to genotoxic stressors. To study DNA damage response kinetics, we treated *Tetrahymena* over the course of 8 hours with 6 different treatments. The damaging agents were 254 nm ultraviolet light (UV), cis-diamineplatinum (II) dichloride (CPT), hydrogen peroxide (HP), methyl methanesulfonate (MMS), ionizing radiation (IR), and hydroxyurea (HU). The treatment conditions were determined either by establishment of EC50 or from previous DNA damage studies of Tetrahymena (Loidl and Mochizuki 2009; Sandoval et al.

2015; Campbell and Romero 1998) (Supplemental table 1). To obtain transcriptome and proteome expression information, we harvested samples at 0, 1, 2, 3, 4, 6, and 8 hours (H0-H8) in quadruplicate, and performed mRNA sequencing (RNA-seq) and high-resolution mass spectrometry (MS) (Figure 1A). We measured all transcriptomes and proteomes in sets of three, two treatments paired with a non-treatment, which were collected and processed together. With this we could calculate changes in transcript expression and protein intensity in each drug treatment condition over a matched non-treated condition, thereby correcting for any potential batch effect ($\log_2$ fold change values) After stringent filtering across treatments and time points, we were able to quantify 20,443 transcripts and 6,551 protein groups with 99.6% of proteins associated to transcriptome data (Figure 1B).

To verify each treatment induced the anticipated DNA damage, we examined differential upregulation of primary DNA repair pathways (Figure 1C). We compiled gene lists for five DNA repair pathways: NER (Tatum et al. 2011; Schärer 2013), BER (Kelley, Kow, and Wilson 2003; Beard et al. 2019), MMR (Chakraborty and Alani 2016; Kunkel and Erie 2015; Bowen et al. 2013), DSBR (X. Li and Heyer 2008; Mathiasen and Lisby 2014; Pannunzio, Watanabe, and Lieber 2018; Scully et al. 2019), ICL (Lehoczký, McHugh, and Chovanec 2007), and general DDR (Ciccia and Elledge 2010; Pizzul et al. 2022). These lists containing 130 DNA repair genes are not intended to be comprehensive, but highlight previously established key genes involved in the different DNA repair pathways. Amongst the five selected DNA repair pathways, we found between 81.4-100% of the genes had upregulated transcripts and between 82.4-96% were upregulated proteins. Of the detected global DNA damage response transcripts and proteins, 81.4% and 88.9% were upregulated in the transcriptome and proteome dataset, respectively. For each treatment, we examined the amount of upregulated transcripts and proteins with its commonly associated DNA repair pathway. Across all treatments, 68.4-96.6% of the transcripts and 33.3-92% of the proteins associated with the respective pathways were upregulated (Figure 1C).

Our set of DNA repair genes were further examined for each treatment over time (Figure 1D). Through hierarchical clustering we found a group of 10 transcripts (MSH6L3, RAD51, RAD4, SNML1, RAD5, RLP1, RAD5L4, RFA1, TTHERM_00316410 (Rad3 homolog), and TTHERM_00391570 (Rev3 homolog)) with a minimum 2.3 $\log_2$ fold change across all treatments, indicating a general DNA damage response across treatments. However, there is variability between protein expression for each treatment. This indicates the dynamics occurring within the transcriptome and proteomes are unique and differentially regulated.
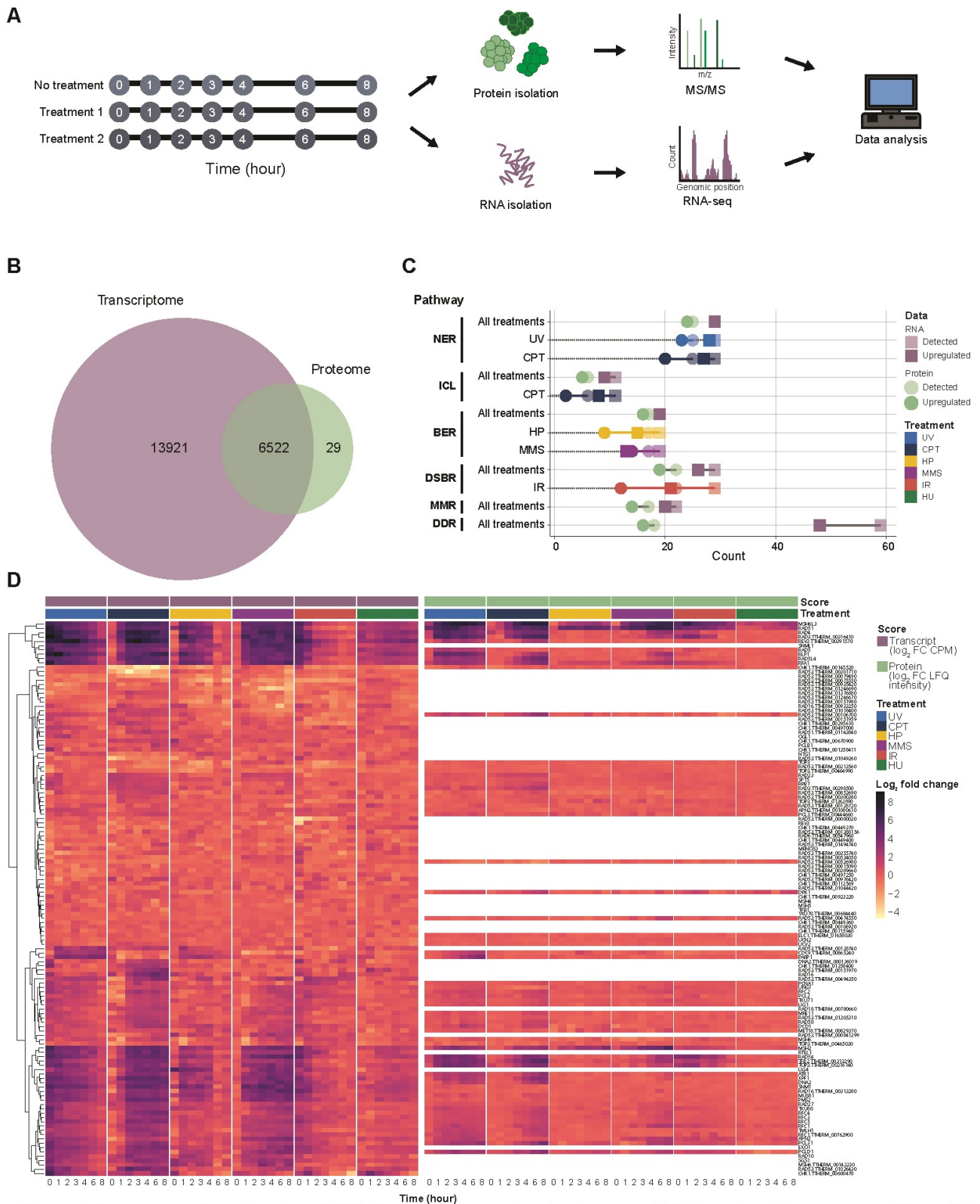
Figure 1. Screen to explore kinetics of DNA damage response in *Tetrahymena*. A) Schematic of the screen workflow. Cells were with a mutagenic agent, and samples were harvested incrementally over eight hours. At each timepoint, samples were collected for processing for both RNA-sequencing and quantitative mass spectrometry. B) Venn diagram depicting the overlap between the identified transcripts and proteins. C) Lollipop plot of enriched DNA Damage repair factors. D) Hierarchial clustering was used to create a heatmap of DNA repair genes. NER: nucleotide excision repair, ICL: intercrosslink repair, BER: base excision repair, DSBR: double-strand break repair, MMR: mismatch repair, DDR: DNA damage response, UV: ultra violet light, CPT:cisplatin, HP: hydrogen peroxide, MMS: methyl methanesulfonate, IR: ionizing radiation, HU: hydroxyurea

1.2 All genotoxic stressors induce a global dynamic response. We examined expression dynamics of both the transcriptome and proteome over time by calculating the Gini coefficient for every quantified transcript and protein (Figure 2). We applied a Gini coefficient filter of the 60th quantile (Gini score > 0.042) to the transcriptome, to separate dynamic and stable transcripts (Figure 2A). In addition to our dynamicity filter, to select transcripts up- and downregulated in response to the DNA damaging agents we applied a $\log_2$ fold change filter, where the biggest change over the time course is more than 1 or less than -1 ($\log_2$ fold change),and a significance filter, of adjusted p-value < 0.05 (FDR). Of the 20,443 transcripts, 8,815 surpassed these thresholds. Amongst them were the previously characterized DSBR gene RAD51 and MSH6 homolog, MSH6L3 (Campbell and Romero 1998; Marsh et al. 2000; Marsh, Cole, and Romero 2001), whereas TTLL6B was found to be a stable, non - or downregulated transcript (Figure 2B). There was a distinctive treatment specific bias in the amount of dynamic up- or downregulated transcripts found, UV treatment having the most and HP treatment the least.

The highest degree of overlap between dynamic transcripts was conditional on the amount of dynamic transcripts, regardless of up or downregulation. However, there was significantly more overlap than expected (Fisher's exact test) whenever three or more treatments overlapped, in nearly every instance (Figure 2C). While a large-scale response is notable, as overlaps narrow, unique shared responses emerged within subsets of treatments. In the upregulated dynamic transcripts eight-related overlaps showed overrepresentation for the Gene Ontology terms 'cellular detoxification', 'cellular response to stress', and 'DNA repair', including the core overlap of transcripts found in all treatments (Gene Ontology Consortium 2021). This reinforces that despite some overlap being driven by the amount of transcripts in each group, we do find a significant core and treatment specific response. Additionally, four or more overlaps had overrepresentation of the MMR, NER, BER, and HR KEGG terms. Many of these overlaps included treatments that are not largely associated with these pathways, indicating that *Tetrahymena* has a strong and diverse DDR at the transcriptome level.

We performed the same analysis with the 6,551 quantified proteins quantified across the six treatments. Here, the Gini coefficient threshold for dynamicity was also set at the 60th quantile (Gini score > 0.021 with a minimum $\log_2$ fold change > 1 or < -1, and p-value < 0.05 (Welch t-test) at any point in the time course . We found a total of 2,582 proteins to be dynamically regulated. RAD51 and MSH6L3 were dynamic in each treatment, aside from HU. Even though both proteins are upregulated during HU treatment, they have a relatively steady state, and therefore did not surpass the stringent dynamicity threshold (Figure 2D). We found

significantly more overlap than expected whenever there were four or more treatments overlapping in nearly every instance (Fisher's exact test, Figure 2E). As for the dynamic transcripts, the overlap of the core dynamic regulated proteins across all treatments had overrepresentation of the GO terms 'DNA repair' and 'cellular response to damage'. These terms were also overrepresented in the CPT_IR_MMS_UV and CPT_UV overlaps. Within the CPT_IR_MMS_UV overlap, there was an overrepresentation of KEGG terms related to MMR, HR, NER, and BER, whereas in the CPT_UV overlap there was on an overrepresentation of the strong CPT- and UV-associated repair pathway NER. This indicates that we detected unique pathway crosstalk across treatments while also having expected enrichment of DDR.

1.3 Early and unique transcriptional response is critical to the DNA damage response. To cluster the 8,815 dynamic and differentially regulated transcripts, we used an unsupervised machine learning approach (Figure 3). The transcripts detected across all six treatments fell into 15 distinct clusters (T1-T15), 689 transcripts could not be assigned (Figure 3A). Each transcript was mapped to its respective homolog(s) in S. cerevisiae, and functional enrichment analysis was performed using Gene Ontology (GO) and for these S. cerevisiae homologs (Gene Ontology Consortium 2021) (Figure 3B). Cluster T10 uniquely showed overrepresentation of genes related to 'DNA repair', 'cellular response to DNA damage stimulus', 'DNA replication', and 'DNA metabolic process'. Within the average expression profile for CPT-treated cells in this cluster, there is an immediate strong and continual upregulation. This response could be due to the known long half-life of CPT (Evans et al. 1982). At H0, the average expression profile of IR- and UV-treated cells shows strong upregulation, followed by gradual down regulation throughout the remaining time points. This also reflects these specific treatments, as both UV and IR treatments had only one initial application and were not sustained in culture. There also is a moderate increase of average expression of HP-, MMS-, and HU-treated cells in this cluster. Additionally, 17 DNA damage response proteins were found in this cluster.

Further general responses to DNA damage were found in the other SOMs. All histones (T7), 20S ribosomal proteins (T2), dense core granules (T3), were clustered together, and showed primarily similar regulation trends for each treatment. The degree and kinetics of decline depended on the treatment, but all subunits within each gene family behaved similarly within each treatment. Three families of previously studied chromatin remodelers in Tetrahymena; the Poly-(ADP-ribose) polymerases (PARPs)/PARP-associated proteins, histone acetyl transferases (HATs), histone
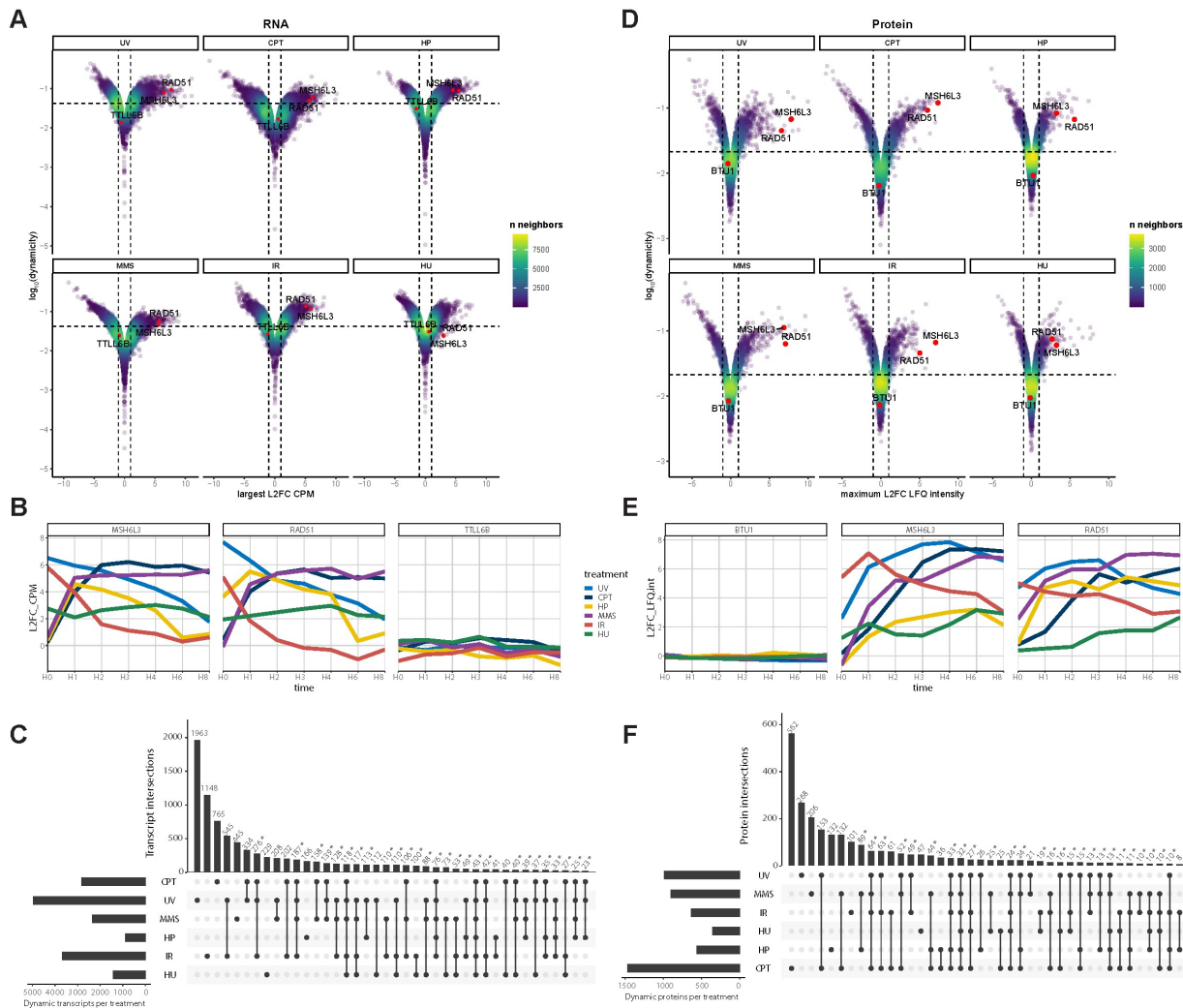
Figure 2. Mutagenic treatments cause global dynamic response. A) Volcano plots plotting dynamic transcripts for each treatment. The x-axis contains the maximal positive or negative fold change during the duration of the time course, and the y-axis contains a Gini score evaluating the dynamicity throughout the entire time course. B) Example line plots of the expression profiles dynamic (MSH6L3 and RAD51) and stable transcripts (TTLL6B). C) Upset plots of overlapping dynamic transcripts between treatments. D) Volcano plots plotting dynamic proteins for each treatment. The x-axis contains the maximal positive or negative fold change during the duration of the time course, and the y-axis contains a Gini score evaluating the dynamicity throughout the entire time course. E) Line plots of the expression profiles dynamic (MSH6L3 and RAD51) and stable proteins (BTU1). F) Upset plots of overlapping dynamic proteins between treatments.

deacetylases (HDACs) are differentially regulated (Wahab et al. 2020; Ashraf et al. 2019; Saettone et al. 2019; Chalker, Meyer, and Mochizuki 2013; Slade et al. 2011). The PARP and PARP-correlated proteins mediate DNA repair by chromatin modifications via ADP-ribosylation, as well as direct binding, modification, and recruitment of DNA repair proteins (Sousa et al. 2012; Morales et al. 2014). PARP7, PARP8, and PARP12 (T2), PCP1 (T7), PARP6 (T9), and PARP2 and PARP5 (T10) all showed unique responses to DNA damage. The histone acetylases and deacetylases are critical to changing chromatin architecture to facilitate DNA repair (Wahab et al. 2020). The histone acetylases (HATs)

HAT1 (T10) and MYST2 (T2) and histone deacetylases THD4, THD17, and THD18a (T2, T3, T6, respectively) also showed greatly differential regulation. This indicates that each of the PARPs, HATs, and HDACs in Tetrahymena has a particular role.

Within the 15 clusters, there are nine clusters enriched for 'protein phosphorylation' (T1, T2, T4, T5, T8, T10, T12-14) (Figure 3C). Eight of these nine clusters have at least a 0.5 log2 fold change at H0. All of these clusters have unique temporal and degrees of response. T1 and T2 have a variety of initial intensities but overall trends of downregulation. Whereas T4 and T5 show variable peaks of regulation
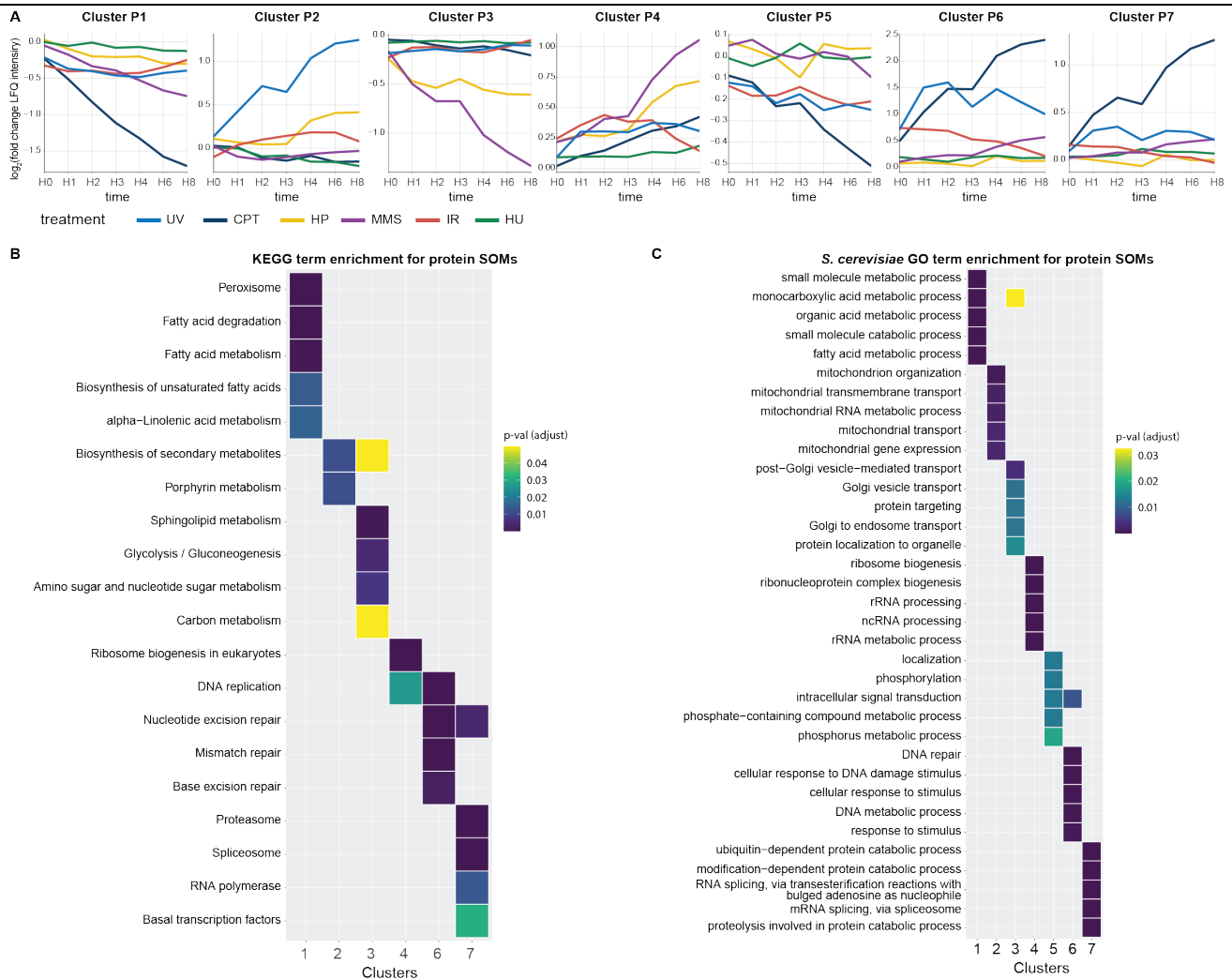
115
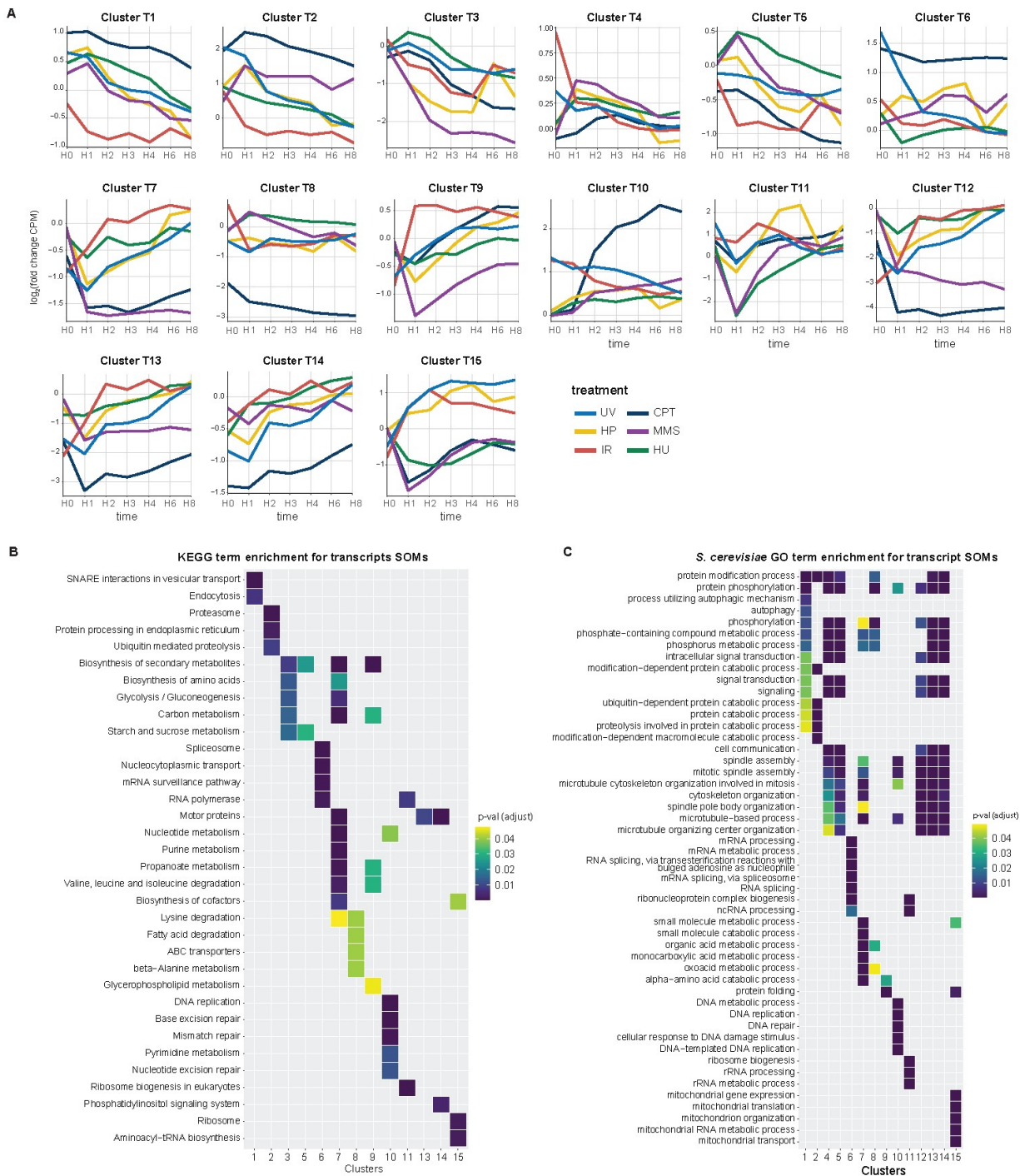
Figure 3. Transcript clusters reveal dynamic DNA damage response. A) Average expression profiles for clusters of dynamic transcripts. Using an unsupervised machine learning technique, we clustered dynamic transcripts based on their shared expression profiles into 15 different clusters. These line graphs are the average expression profiles for each treatment. B) Heat map of functional enrichment analysis using KEGG. Each row contains an over-represented KEGG term with a gradient representing the adjusted p-value. B) Heat map of functional enrichment analysis using GO analysis. Each gene was mapped to its respective homolog(s) in *S. cerevisiae*. Each row contains an over-represented GO term with a gradient representing the adjusted p-value.

followed by downregulation. In T8, T10, T12, T13, and T14, CPT-treated cells consistently show the most extreme differential regulation, while all other treatments showed variable degree of differential regulation. It has been previously reported that phosphorylation plays a critical role in the processing of interstrand crosslinks, as well as

preventing ICL proteins from conducting inappropriate repair (J. Huang et al. 2019; Lopez-Martinez et al. 2019; Deans and West 2011; Clingen et al. 2008). Overall, it is clear that protein phosphorylation is critical to immediate and sustained DNA damage response as a whole in Tetrahymena. Together, the transcriptome data indicates the importance of

Figure 4. Protein clusters reveal specific dynamic DNA damage response. A) Average expression profiles for clusters of dynamic proteins. Using an unsupervised machine learning technique, we clustered dynamic proteins based on their shared expression profiles into 7 different clusters. These line graphs are the average expression profiles for each treatment. B) Heat map of functional enrichment analysis using KEGG. Each row contains an over-represented KEGG term with a gradient representing the adjusted p-value. B) Heat map of functional enrichment analysis using GO analysis. Each gene was mapped to its respective homolog(s) in *S. cerevisiae*. Each row contains an over-represented GO term with a gradient representing the adjusted p-value.

the unique transcriptional regulation of the degree and kinetics of the DNA damage response to various genotoxic stressors.

1.4 Protein expression over time reveals specific trends involved in DNA damage response. As for the transcriptome, we used self-organizing maps clustering 2,582 dynamically expressed proteins into seven distinct clusters (P1-P7), 202 proteins could not be assigned (Figure 4A). Each protein included in these seven clusters was mapped to its respective homolog(s) in S. cerevisiae, and functional enrichment analysis was performed using GO. (Figure 4B). Cluster P6 uniquely showed overrepresentation of genes r elated to 'DNA repair', 'cellular response to DNA damage stimulus', 'DNA replication', and 'DNA metabolic process'.

In P6, there were 15 known DNA damage factors, which included ATR1, RAD53/Chk1, TKU80, RAD3, DNA2, three

members of the RFC complex, and TKU80. ATR1 and TKU80 are critical in both DNA damage response and conjugation in Tetrahymena (Loidl and Mochizuki 2009; Lin, Chao, and Yao 2012). Additionally, in this cluster, two MMR proteins were identified, MSH3L6 and TMLH1. Together with MSH2, MSH3 is part of the MMR MutSβ complex, which repairs larger insertions and deletions. Alternatively, if MSH6 and MSH2 form an MMR recognition heterodimer MutSα, one to two base pair mismatches and indels are repaired. Intriguingly, MSH6 is found in P3, indicating a differential regulation of these portions of the MMR recognition complex. TMHL1 and PMS2 interact with the recognition complex to initiate cleavage events. Our clustering data suggests that TMHL1 expression profiles are more comparable to MSH3 rather

than MSH6. Similarly to transcriptional regulation, the chromatin remodelers within the PARP family were differentially regulated across clusters. However, it is clear that this family of genes is not just being regulated at the transcriptional level. For example, PARP7, PARP8, and PARP12 all fell within T2, whereas now they are found within P6, P7, and P4, respectively.

Other gene families not directly responsible for DNA repair that had a similar time-dependent transcriptional regulation fell into different clusters and showed diverging expression patterns on protein level. For example, all dynamic dense core granules clustered together in T3, but were part of two different protein clusters (P2 and P3). All histones were transcriptionally regulated in a highly similar manner (T33), whereas the histone proteins had different protein expression profiles (P6 and P7). Also, the DNA repair-related PARP family was differentially regulated. While the dynamic transcripts also showed a differential regulation across all clusters, there was clustering of some of the PARP transcripts expression profiles in the same SOM. However, there were some instances of similar protein and transcriptional regulation, such as for the 20S proteasome (T2 and P7) and transcription related factors (T6 and P7). Generally, the differences between transcriptome and protein expression profiles indicate additional regulation at the protein level.

1.5 Protein and transcript correlation. Data analysis occurring currently.

1.6 Using novel knockdown system to characterize member of the PARP protein family. Utilizing a novel knockdown system created by J.J. Smith, we take advantage of the unique phenotypic characteristics of a mutant strain of Beta tubulin 1, BTU-1 in Tetrahymena (Gaertig et al. 1994). We targeted the btu-1 (K350M; pacs loci) allele of CU522. Previously characterized due to its unique effects on the macronuclear development, these mutants are sensitive to a number of selective pressures including sublethal treatments of Paclitaxel. This microtubule stabilizer is used for positive selection of successful transformants (Gaertig et al. 1999; Smith et al. 2004). The novel knockdown system contains a designed hairpin flanked by two Beta tubulin arms in the base pUC118 construct. After being integrated via biolistic bombardment and undergoing positive selection, cells were screened with whole cell PCR for successful integration. All targeted gene were confirmed via whole cell PCR. Now, proteome samples will be measured to verify the reduction of the targeted gene and examine what proteins become differentially regulated due to the knockdown. Subsequently, we will measure the influence the knockdowns have on resistance to DNA damage.

3. Discussion

This research is currently on going and firm conclusions will be drawn at the end of the study. More in depth discussion has been included in the body of the thesis. However, in short, in this study, we provide the first comprehensive view of DNA damage over time utilizing the ciliate *Tetrahymena thermophila.* Previously studied DNA damage proteins as well as homologs of known DDR proteins were upregulated in response to damage. There was an unanticipated group of DDR proteins which were core, intense responders. We are in the midst of characterizing two of these candidates now, RAD51 and MSH6L3. Treatments all induced a large dynamic response at both a transcriptome and proteome level. To explore these dynamic transcripts and proteins further we utilized unsupervised machine learning technique to create clusters of expression profiles. This revealed an unexpected differential regulation of PARP proteins over treatments. We are currently in the midst of characterizing a sub-set of these PARP-related proteins to see if the reduction of one of these proteins influences resistance to damage or causes particular differential regulation of associated PARP proteins.

First and foremost, we hope that this data set will serve as a resource to the ciliate community. Before submission, an online platform will be in place so that the data set can be easily navigated.

4. Methods

1.1 Cell culture.   The Tetrahymena thermophila wildtype strain SB210 (Tetrahymena Stock Center) was used throughout the study. Cultures were grown in a medium of 2% proteose peptone (BD Biosciences), 0.2% yeast extract (BD Biosciences), 12 µM FeCl, and 1x Penicillin/ Streptomycin/ Funizone (Hyclone) at 30 °C at 100-150 rotations per minute.

1.2 Collection of Tetrahymena *for mass spectrometry and RNA sequencing. Tetrahymena* were grown to a concentration between 1.5×10^5-3×10^5 cells/ml in 500 ml cultures. Samples were treated with six different conditions, grouping two treatments with one non-treated group (grouped as MMS, HP; CP, UV; and HU,IR). Cells were harvested samples at 0, 1, 2, 3, 4, 6, and 8 hours after the initial treatment. To collect samples for later quantitative mass spectrometry, 5×10^4 cells were centrifuged at 9,400 xg for 5 minutes. Supernatant was removed and cells were washed with 1 ml 10 mM Tris-HCl (pH=7.5), and centrifuged at 9,400 xg for 5 minutes. Supernatant was discarded leaving a total ~15 µl of cells and Tris, and 5 µl of 4x LDS (Thermo) and 2 µl of 1M DTT (Sigma) were added. Samples were heated to 90°C for 10 minutes. Samples were

stored at -20°C until mass spectrometry sample preparation. To collect samples for later RNA sequencing (RNA-seq), 5 ml of cells were collected and centrifuged at 1,400 xg for 3 minutes. The supernatant was decanted and cells were washed with 5 ml 10 mM Tris-HCl (pH=7.5). Cells were centrifuged at 1,400 xg for 3 minutes and the supernatant was removed. Cell pellet was resuspended in 600 ul Buffer RLT (Qiagen, RNeasy mini kit), flash frozen in liquid nitrogen, and stored at -80°C until RNA sequencing sample preparation.

1.3 Mass spectrometry sample preparation. LDS sample was loaded on a 4-12% NuPage NOVEX Bis-Tris gel (Thermo) and run for 10 min at 180V, in 1x MES buffer (Thermo Fisher Scientific). Samples were processed as previously described (Scherer et al. 2020). In short, the gel was stained and fixed with Coomassie Brilliant Blue G250 (Sigma Aldrich), initial destaining of the gels was done overnight with water. Gel pieces were cut, further destained with 50% EtOH / 50 mM ammonium bicarbonate (ABC) and dehydrated with acetonitrile (VWR), reduced with 10 mM DTT (Sigma) and alkylated using iodoacetamide (Sigma) and subsequently again dehydrated with acetonitrile (VWR) and digested with 1 µg of MS-grade trypsin (Sigma) at 37 °C overnight. The peptides were eluted from the gel pieces and loaded onto activated C18 material (Empore) StageTips (Rappsilber, Mann, and Ishihama 2007) and stored at 4 °C until elution and measurement.at 4 °C until measurement.

1.4 Mass spectrometry measurement. Peptides were eluted from the StageTips using 80% acetonitrile / 0.1% formic acid and concentrated prior to loading on an Easy-nLC-1200 system coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher). The peptides were loaded on a 50 cm column (75 µm inner diameter, New Objective) in-house packed with ReproSil-Pur 120 C18-AQ (Dr. Maisch GmbH). We used a 103-min gradient from 3% to 40% acetonitrile with 0.1% formic acid at a flow of 250 nl/min. The mass spectrometer was operated in positive ion mode with a top 20 MS/MS data-dependent acquisition strategy of one MS full scan (scan range 300 - 1,650 m/z; 60,000 resolution; normalized AGC target 300%; max IT 28 ms) and up to twenty MS/MS scans (15,000 resolution; AGC target 100%, max IT 40 ms; isolation window 1.4 m/z) with peptide match preferred using HCD fragmentation.

1.5 Mass spectrometry data analysis. Raw files were analyzed using MaxQuant (version 1.6.10.43). As a search space the *T. thermophila* protein database was used (June 2014, from the TGD). Oxidation and acetylation were set as variable modifications, Carbamidomethylation as fixed modification. Fast LFQ was used to calculate and normalize intensities. The minimum ratio count used was 2. Match between runs was used to match within each time point per treatment, and to the time points right before and after, with a match time window of 0.7 min, match ion mobility window of 0.05, an alignment time window of 20 min, and alignment ion mobility of 1. Matching of unidentified features was deactivated. For protein quantification label minimum ratio count 2 and unique + razor peptides were used.

1.6 RNA sample preparation and sequencing. Previously obtained samples were thawed on ice. RNA isolation was performed with RNeasy mini kit (Qiagen) in accordance with manufacturer instructions with the addition of the optional DNaseI on column digestion. This digestion was carried out with 3 units of DNaseI (Qiagen) per sample, and samples were digested 15 minutes at room temperature on the column. NGS library prep was performed with Lexogen`s QuantSeq 3´mRNA-Seq Library Prep Kit FWD following Lexogen`s standard protocol (015UG009V0252). Libraries were prepared with a starting amount of 300 ng and amplified in 14 PCR cycles. Libraries were profiled in a High Sensitivity DNA on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life technologies). All libraries from the two treatments and coordinating non-treatment were pooled together in equimolar ratio and sequenced on 1 NextSeq 500 high output
flow cell, SR for 1x84 cycles plus 7 cycles for the index read.

1.7 RNA-seq analysis. All demultiplexed, raw sequencing files of each treatment set were analyzed together. Initial analysis was done through a modified version of the NGS pipeline by the bioinformatics core facility of the IMB (available at https://gitlab.rlp.net/imbforge/NGSpipe2go). For reference, "subread2rnatypes", "genebodyCov2", "rMATS", and the GO enrichment analysis were removed from the pipeline. In short, the library quality was assessed with FastQC before alignment against the *T. thermophila* genome assembly SB210 and a custom built GTF file, which included gene annotations from *T. thermophila* (TGD, T_thermophila_June2014.gff3). Alignment was performed with STAR aligner version 2.7.3a. Reads mapping to annotated features in the custom GTF file were counted with featureCounts. Initial CPM counts were calculated with DESeq2 in R (R Core Team).

1.8 Further bioinformatics analysis. All further analysis was done with scripts developed in R (R Core Team 2022), incorporating ggplot2 for visualization (Wickham 2016) among other packages.
For proteome data, contaminants, reverse database hits, protein groups only identified by site, and protein groups with less than two peptides (at least one of them classified as

unique) were removed. Additionally, only protein groups present in at least 2 out of 4 technical replicates were kept. Missing values were imputed by shifting a compressed beta distribution obtained from the LFQ intensity values to the limit of quantitation (between 0.2 and 2.5 percentile of the measured intensity distribution per sample). LFQ intensities were log2 transformed, after which fold changes for individual comparisons of time points or strains could be calculated per protein, a Welch t-test was used to calculate p-values. The general protein enrichment threshold was set to a p-value lower than 0.05 and an absolute fold change higher than 1. All calculated values can be found in supplemental table XX.

For transcriptome data, transcripts which did not have any CPM value across the time points and treatments above the 25th quantile of all CPM values (CPM < 1.673028) were removed. All CPM values were log2 transformed. Differential regulation thresholds were set at L2FC > 1 or < -1, and adjusted p-value (FDR) < 0.05.

Dynamicity of transcripts or proteins was calculated using the Gini ratio, as described before (Casas-Vila et al. 2017; Damgaard and Weiner 2000). Statistical testing of overlaps of dynamic genes was done with the R package SuperExactTest (Wang, Zhao, and Zhang 2015). Functional enrichment analysis was performed using Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2023), Gene Ontology, and the ClusterProfiler R package (Wu et al. 2021; Yu et al. 2012) for statistical analysis. Terms for groups of enriched proteins were assessed for overrepresentation with a Fisher's exact test, against all terms found in our complete dataset as background. The enrichment threshold was set to an adjusted (FDR) p-value < 0.05. Self-organizing map (SOM) clustering was done with the help of the Kohonen package in R (Wehrens and Buydens 2007).

All data can be explored through a user-friendly web interface at https://butterlab.imb-mainz.de/Tt_DDR. This web interface was designed and built with the use of R Shiny. All data and code for the analysis in this study was written in R, and is freely available via the workflowr (ref) website https://vivienschoonenberg.gitlab.io/Tetddr_wflowr/ or https://gitlab.com/vivienschoonenberg/Tetddr_wflowr.

Acknowledgements

Funding

References

Agrawal, Rohit Kumar, Rakesh Kantilal Patel, Varsha shah, Lalit Nainiwal, and Bhadra Trivedi. 2014. "Hydroxyurea in Sickle Cell Disease: Drug Review." Indian Journal of Hematology & Blood Transfusion 30 (2): 91–96. https://doi.org/10.1007/s12288-013-0261-4.

Ashraf, Kanwal, Syed Nabeel-Shah, Jyoti Garg, Alejandro Saettone, Joanna Derynck, Anne-Claude Gingras, Jean-Philippe Lambert, Ronald E. Pearlman, and Jeffrey Fillingham. 2019. "Proteomic Analysis of Histones H2A/H2B and Variant Hv1 in Tetrahymena Thermophila Reveals an Ancient Network of Chaperones." Molecular Biology and Evolution 36 (5): 1037–55. https://doi.org/10.1093/molbev/msz039.

Beard, William A., Julie K. Horton, Rajendra Prasad, and Samuel H. Wilson. 2019. "Eukaryotic Base Excision Repair: New Approaches Shine Light on Mechanism." Annual Review of Biochemistry 88 (1): 137–62. https://doi.org/10.1146/annurev-biochem-013118-111315.

Bharati, Akhilendra Pratap, Sunita Kumari, and Md Sohail Akhtar. 2020. "Proteome Analysis of Saccharomyces Cerevisiae after Methyl Methane Sulfonate (MMS) Treatment." Biochemistry and Biophysics Reports 24 (October): 100820. https://doi.org/10.1016/j.bbrep.2020.100820.

Bowen, Nikki, Catherine E. Smith, Anjana Srivatsan, Smaranda Willcox, Jack D. Griffith, and Richard D. Kolodner. 2013. "Reconstitution of Long and Short Patch Mismatch Repair Reactions Using Saccharomyces Cerevisiae Proteins." Proceedings of the National Academy of Sciences of the United States of America 110 (46): 18472–77. https://doi.org/10.1073/pnas.1318971110.

Campbell, C, and D P Romero. 1998. "Identification and Characterization of the RAD51 Gene from the Ciliate Tetrahymena Thermophila." Nucleic Acids Research 26 (13): 3165–72.

Casas-Vila, Nuria, Alina Bluhm, Sergi Sayols, Nadja Dinges, Mario Dejung, Tina Altenhein, Dennis Kappei, Benjamin Altenhein, Jean-Yves Roignant, and Falk Butter. 2017. "The Developmental Proteome of Drosophila Melanogaster." Genome Research 27 (7): 1273–85. https://doi.org/10.1101/gr.213694.116.

Chakraborty, Ujani, and Eric Alani. 2016. "Understanding How Mismatch Repair Proteins Participate in the Repair/Anti-Recombination Decision." FEMS Yeast Research 16 (6): fow071. https://doi.org/10.1093/femsyr/fow071.

Chalker, Douglas L., Eric Meyer, and Kazufumi Mochizuki. 2013. "Epigenetics of Ciliates." Cold Spring Harbor Perspectives in Biology 5 (12): a017764. https://doi.org/10.1101/cshperspect.a017764.

Chapman, J. Ross, Martin R. G. Taylor, and Simon J. Boulton. 2012. "Playing the End Game: DNA Double-Strand Break Repair Pathway Choice." Molecular Cell 47 (4): 497–510. https://doi.org/10.1016/j.molcel.2012.07.029.

Ciccia, Alberto, and Stephen J. Elledge. 2010. "The DNA Damage Response: Making It Safe to Play with Knives." Molecular Cell 40 (2): 179–204. https://doi.org/10.1016/j.molcel.2010.09.019.

Clingen, P. H., J. Y.-H. Wu, J. Miller, N. Mistry, F. Chin, P. Wynne, K. M. Prise, and J. A. Hartley. 2008. "Histone H2AX Phosphorylation as a Molecular Pharmacological Marker for DNA Interstrand Crosslink Cancer Chemotherapy." Biochemical Pharmacology 76 (1): 19–27. https://doi.org/10.1016/j.bcp.2008.03.025.

Damgaard, Christian, and Jacob Weiner. 2000. "Describing Inequality in Plant Size or Fecundity." Ecology 81 (4): 1139–42. https://doi.org/10.1890/0012-9658(2000)081[1139:DIIPSO]2.0.CO;2.

Deans, Andrew J., and Stephen C. West. 2011. "DNA Interstrand Crosslink Repair and Cancer." Nature Reviews Cancer 11 (7): 467–80. https://doi.org/10.1038/nrc3088.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. "STAR: Ultrafast Universal RNA-Seq Aligner." Bioinformatics 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Duan, Mingrui, Jenna Ulibarri, Ke Jian Liu, and Peng Mao. 2020. "Role of Nucleotide Excision Repair in Cisplatin Resistance." International Journal of Molecular Sciences 21 (23): E9248. https://doi.org/10.3390/ijms21239248.

Evans, William E., Gary C. Yee, William R. Crom, Charles B. Pratt, and Alexander A. Green. 1982. "Clinical Pharmacology of Bleomycin and Cisplatin." Drug Intelligence & Clinical Pharmacy 16 (6): 448–58. https://doi.org/10.1177/106002808201600602.

Fayyad, Nour, Farah Kobaisi, David Beal, Walid Mahfouf, Cécile Ged, Fanny Morice-Picard, Mohammad Fayyad-Kazan, et al. 2020. "Xeroderma Pigmentosum C (XPC) Mutations in Primary Fibroblasts Impair Base Excision Repair Pathway and Increase Oxidative DNA Damage." Frontiers in Genetics 11. https://www.frontiersin.org/articles/10.3389/fgene.2020.561687

Gaertig, J., Y. Gao, T. Tishgarten, T. G. Clark, and H. W. Dickerson. 1999. "Surface Display of a Parasite Antigen in the Ciliate Tetrahymena Thermophila." Nature Biotechnology 17 (5): 462–65. https://doi.org/10.1038/8638.

Gaertig, J, T H Thatcher, L Gu, and M A Gorovsky. 1994. "Electroporation-Mediated Replacement of a Positively and Negatively Selectable Beta-Tubulin Gene in Tetrahymena Thermophila." Proceedings of the National Academy of Sciences of the United States of America 91 (10): 4549–53.

Gene Ontology Consortium. 2021. "The Gene Ontology Resource: Enriching a GOld Mine." Nucleic Acids Research 49 (D1): D325–34. https://doi.org/10.1093/nar/gkaa1113.

Hauser, Melinda, Paul E. Abraham, Lorenz Barcelona, and Jeffrey M. Becker. 2019. "UV Laser-Induced, Time-Resolved Transcriptome Responses of Saccharomyces Cerevisiae." G3:

Genes|Genomes|Genetics 9 (8): 2549–60. https://doi.org/10.1534/g3.119.400291.

Huang, Huifang, Leiming Zhu, Brian R. Reid, Gary P. Drobny, and Paul B. Hopkins. 1995. "Solution Structure of a Cisplatin-Induced DNA Interstrand Cross-Link." Science 270 (5243): 1842–45. https://doi.org/10.1126/science.270.5243.1842.

Huang, Jing, Jing Zhang, Marina A. Bellani, Durga Pokharel, Julia Gichimu, Ryan C. James, Himabindu Gali, et al. 2019. "Remodeling of Interstrand Crosslink Proximal Replisomes Is Dependent on ATR, FANCM, and FANCD2." Cell Reports 27 (6): 1794-1808.e5. https://doi.org/10.1016/j.celrep.2019.04.032.

Jackson, Stephen P., and Jiri Bartek. 2009. "The DNA-Damage Response in Human Biology and Disease." Nature 461 (7267): 1071–78. https://doi.org/10.1038/nature08467.

Kanehisa, Minoru, Miho Furumichi, Yoko Sato, Masayuki Kawashima, and Mari Ishiguro-Watanabe. 2023. "KEGG for Taxonomy-Based Analysis of Pathways and Genomes." Nucleic Acids Research 51 (D1): D587–92. https://doi.org/10.1093/nar/gkac963.

Kelley, Mark R., Yoke W. Kow, and David M. Wilson III. 2003. "Disparity between DNA Base Excision Repair in Yeast and Mammals: Translational Implications1." Cancer Research 63 (3): 549–54.

Kim, Dong Ryoung, Rohan D Gidvani, Brian P Ingalls, Bernard P Duncker, and Brendan J McConkey. 2011. "Differential Chromatin Proteomics of the MMS-Induced DNA Damage Response in Yeast." Proteome Science 9 (October): 62. https://doi.org/10.1186/1477-5956-9-62.

Kovalchuk, I. 2016. "Chapter 38 - Conserved and Divergent Features of DNA Repair: Future Perspectives in Genome Instability Research." In Genome Stability, edited by Igor Kovalchuk and Olga Kovalchuk, 651–66. Boston: Academic Press. https://doi.org/10.1016/B978-0-12-803309-8.00038-0.

Kubota, Takashi, David A. Stead, Shin-ichiro Hiraga, Sara ten Have, and Anne D. Donaldson. 2012. "Quantitative Proteomic Analysis of Yeast DNA Replication Proteins." Methods (San Diego, Calif.) 57 (2): 196–202. https://doi.org/10.1016/j.ymeth.2012.03.012.

Kumar, Namrata, Arjan F. Theil, Vera Roginskaya, Yasmin Ali, Michael Calderon, Simon C. Watkins, Ryan P. Barnes, et al. 2022. "Global and Transcription-Coupled Repair of 8-OxoG Is Initiated by Nucleotide Excision Repair Proteins." Nature Communications 13 (1): 974. https://doi.org/10.1038/s41467-022-28642-9.

Kunkel, Thomas A., and Dorothy A. Erie. 2015. "Eukaryotic Mismatch Repair in Relation to DNA Replication." Annual Review of Genetics 49 (1): 291–313. https://doi.org/10.1146/annurev-genet-112414-054722.

Lehoczký, Peter, Peter J. McHugh, and Miroslav Chovanec. 2007. "DNA Interstrand Cross-Link Repair in Saccharomyces Cerevisiae." FEMS Microbiology Reviews 31 (2): 109–33. https://doi.org/10.1111/j.1574-6976.2006.00046.x.

Li, Xuan, and Wolf-Dietrich Heyer. 2008. "Homologous Recombination in DNA Repair and DNA Damage Tolerance." Cell Research 18 (1): 99–113. https://doi.org/10.1038/cr.2008.1.

Li, Zhongdao, Alexander H. Pearlman, and Peggy Hsieh. 2016. "DNA Mismatch Repair and the DNA Damage Response." DNA Repair 38 (February): 94–101. https://doi.org/10.1016/j.dnarep.2015.11.019.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. "FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features." Bioinformatics 30 (7): 923–30. https://doi.org/10.1093/bioinformatics/btt656.

Lin, I-Ting, Ju-Lan Chao, and Meng-Chao Yao. 2012. "An Essential Role for the DNA Breakage-Repair Protein Ku80 in Programmed DNA Rearrangements in Tetrahymena Thermophila." Edited by Orna Cohen-Fix. Molecular Biology of the Cell 23 (11): 2213–25. https://doi.org/10.1091/mbc.e11-11-0952.

Loidl, Josef, and Kazufumi Mochizuki. 2009. "Tetrahymena Meiotic Nuclear Reorganization Is Induced by a Checkpoint Kinase–Dependent Response to DNA Damage." Molecular Biology of the Cell 20 (9): 2428–37. https://doi.org/10.1091/mbc.E08-10-1058.

Lopez-Martinez, David, Marian Kupculak, Di Yang, Yasunaga Yoshikawa, Chih-Chao Liang, Ronghu Wu, Steven P. Gygi, and Martin A. Cohn. 2019. "Phosphorylation of FANCD2 Inhibits the FANCD2/FANCI Complex and Suppresses the Fanconi Anemia Pathway in the Absence of DNA Damage." Cell Reports 27 (10): 2990-3005.e5. https://doi.org/10.1016/j.celrep.2019.05.003.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." Genome Biology 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Marsh, Thomas C, Eric S Cole, and Daniel P Romero. 2001. "The Transition From Conjugal Development to the First Vegetative Cell Division Is Dependent on RAD51 Expression in the Ciliate Tetrahymena Thermophila." Genetics 157 (4): 1591–98. https://doi.org/10.1093/genetics/157.4.1591.

Marsh, Thomas C, Eric S Cole, Kathleen R Stuart, Colin Campbell, and Daniel P Romero. 2000. "RAD51 Is Required for Propagation of the Germinal Nucleus in Tetrahymena Thermophila." Genetics 154 (4): 1587–96. https://doi.org/10.1093/genetics/154.4.1587.

Mathiasen, David P., and Michael Lisby. 2014. "Cell Cycle Regulation of Homologous Recombination in Saccharomyces Cerevisiae." FEMS Microbiology Reviews 38 (2): 172–84. https://doi.org/10.1111/1574-6976.12066.

Morales, Julio, Longshan Li, Farjana J. Fattah, Ying Dong, Erik A. Bey, Malina Patel, Jinming Gao, and David A. Boothman. 2014. "Review of Poly (ADP-Ribose) Polymerase (PARP) Mechanisms of Action and Rationale for Targeting in Cancer and Other Diseases." Critical Reviews in Eukaryotic Gene Expression 24 (1): 15–28. https://doi.org/10.1615/critreveukaryotgeneexpr.2013006875.

Pannunzio, Nicholas R., Go Watanabe, and Michael R. Lieber. 2018. "Nonhomologous DNA End-Joining for Repair of DNA Double-Strand Breaks." The Journal of Biological Chemistry 293 (27): 10512–23. https://doi.org/10.1074/jbc.TM117.000374.

Petermann, Eva, Manuel Luís Orta, Natalia Issaeva, Niklas Schultz, and Thomas Helleday. 2010. "Hydroxyurea-Stalled Replication Forks Become Progressively Inactivated and Require Two Different RAD51-Mediated Pathways for Restart and Repair." Molecular Cell 37 (4): 492–502. https://doi.org/10.1016/j.molcel.2010.01.021.

Pizzul, Paolo, Erika Casari, Marco Gnugnoli, Carlo Rinaldi, Flavio Corallo, and Maria Pia Longhese. 2022. "The DNA Damage Checkpoint: A Tale from Budding Yeast." Frontiers in Genetics 13. https://www.frontiersin.org/articles/10.3389/fgene.2022.995163

R Core Team. 2022. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org.

Ransy, Céline, Clément Vaz, Anne Lombès, and Frédéric Bouillaud. 2020. "Use of H2O2 to Cause Oxidative Stress, the Catalase Issue." International Journal of Molecular Sciences 21 (23): 9149. https://doi.org/10.3390/ijms21239149.

Rappsilber, Juri, Matthias Mann, and Yasushi Ishihama. 2007. "Protocol for Micro-Purification, Enrichment, Pre-Fractionation and Storage of Peptides for Proteomics Using StageTips." Nature Protocols 2 (8): 1896–1906. https://doi.org/10.1038/nprot.2007.261.

Rodríguez-Lombardero, Silvia, Ángel Vizoso-Vázquez, Luis J. Lombardía, Manuel Becerra, M. Isabel González-Siso, and M. Esperanza Cerdán. 2014. "Sky1 Regulates the Expression of Sulfur Metabolism Genes in Response to Cisplatin." Microbiology 160 (Pt 7): 1357–68. https://doi.org/10.1099/mic.0.078402-0.

Saettone, Alejandro, Syed Nabeel-Shah, Jyoti Garg, Jean-Philippe Lambert, Ronald E. Pearlman, and Jeffrey Fillingham. 2019. "Functional Proteomics of Nuclear Proteins in Tetrahymena Thermophila: A Review." Genes 10 (5): 333. https://doi.org/10.3390/genes10050333.

Sandoval, Pamela Y., Po-Hsuen Lee, Xiangzhou Meng, and Geoffrey M. Kapler. 2015. "Checkpoint Activation of an Unconventional DNA Replication Program in Tetrahymena." PLoS Genetics 11 (7): e1005405. https://doi.org/10.1371/journal.pgen.1005405.

Schärer, Orlando D. 2013. "Nucleotide Excision Repair in Eukaryotes." Cold Spring Harbor Perspectives in Biology 5 (10): a012609. https://doi.org/10.1101/cshperspect.a012609.

Scherer, Marian, Michal Levin, Falk Butter, and Marion Scheibe. 2020. "Quantitative Proteomics to Identify Nuclear RNA-Binding Proteins of Malat1." International Journal of Molecular Sciences 21 (3): E1166. https://doi.org/10.3390/ijms21031166.

Scully, Ralph, Arvind Panday, Rajula Elango, and Nicholas A. Willis. 2019. "DNA Double-Strand Break Repair-Pathway Choice in Somatic Mammalian Cells." Nature Reviews Molecular Cell Biology 20 (11): 698–714. https://doi.org/10.1038/s41580-019-0152-0.

Slade, Kristin M., Sydney Freggiaro, Kyle A. Cottrell, Joshua J. Smith, and Emily A. Wiley. 2011. "Sirtuin-Mediated Nuclear Differentiation and Programmed Degradation in Tetrahymena." BMC Cell Biology 12 (1): 40. https://doi.org/10.1186/1471-2121-12-40.

Smith, Joshua J., J. Sebastian Yakisich, Geoffrey M. Kapler, Eric S. Cole, and Daniel P. Romero. 2004. "A β-Tubulin Mutation Selectively Uncouples Nuclear Division and Cytokinesis in Tetrahymena Thermophila." Eukaryotic Cell 3 (5): 1217–26. https://doi.org/10.1128/EC.3.5.1217-1226.2004.

Sousa, Fabricio G., Renata Matuo, Daniele G. Soares, Alexandre E. Escargueil, João A.P. Henriques, Annette K. Larsen, and Jenifer Saffi. 2012. "PARPs and the DNA Damage Response."

Carcinogenesis 33 (8): 1433–40.
https://doi.org/10.1093/carcin/bgs132.

Spivak, Graciela. 2015. "Nucleotide Excision Repair in Humans."
DNA Repair 36 (December): 13–18.
https://doi.org/10.1016/j.dnarep.2015.09.003.

Suter, Bernhard, Daniel Auerbach, and Igor Stagljar. 2006. "Yeast-
Based Functional Genomics and Proteomics Technologies: The
First 15 Years and Beyond." BioTechniques 40 (5): 625–44.
https://doi.org/10.2144/000112151.

Tatum, Danielle, Shisheng Li, Danielle Tatum, and Shisheng Li.
2011. Nucleotide Excision Repair in S. Cerevisiae. DNA Repair
- On the Pathways to Fixing DNA Damage and Errors.
IntechOpen. https://doi.org/10.5772/22129.

Wahab, Suzanne, Alejandro Saettone, Syed Nabeel-Shah, Nora
Dannah, and Jeffrey Fillingham. 2020. "Exploring the Histone
Acetylation Cycle in the Protozoan Model Tetrahymena
Thermophila." Frontiers in Cell and Developmental Biology 8:
509. https://doi.org/10.3389/fcell.2020.00509.

Wang, Minghui, Yongzhong Zhao, and Bin Zhang. 2015. "Efficient
Test and Visualization of Multi-Set Intersections." Scientific
Reports 5 (1): 16923. https://doi.org/10.1038/srep16923.

Wickham, Hadley. 2016. Ggplot2: Elegant Graphics for Data
Analysis. Springer-Verlag New York.
https://ggplot2.tidyverse.org.

Wu, Tianzhi, Erqiang Hu, Shuangbin Xu, Meijun Chen, Pingfan
Guo, Zehan Dai, Tingze Feng, et al. 2021. "ClusterProfiler 4.0:
A Universal Enrichment Tool for Interpreting Omics Data." The
Innovation 2 (3): 100141.
https://doi.org/10.1016/j.xinn.2021.100141.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He.
2012. "ClusterProfiler: An R Package for Comparing Biological
Themes among Gene Clusters." Omics: A Journal of Integrative
Biology 16 (5): 284–87.
https://doi.org/10.1089/omi.2011.0118.Zhou, Chunshui,
Andrew E. H. Elia, Maria L. Naylor, Noah Dephoure, Bryan A.
Ballif, Gautam Goel, Qikai Xu, et al. 2016. "Profiling DNA
Damage-Induced Phosphorylation in Budding Yeast Reveals
Diverse Signaling Networks." Proceedings of the National
Academy of Sciences 113 (26): E3667–75.
https://doi.org/10.1093/pnas.1602827113.

Conclusions and future perspectives

Discovery and characterization of novel telomere-end binders using quantitative mass spectrometry

Previously, phylointeractomics have been used to discover novel telomere binders (Kappei et al. 2017). The original study conducted in *H. sapiens* revealed a number of previously characterized candidates such as shelterin and HOT1, as well as novel candidates, such as ZBTB10 and ZBTB48 (Kappei et al. 2013; de Lange 2018; Bluhm et al. 2019; Jahn et al. 2017). This same technique was applied to *C. elegans*, using label free and dimethyl labeling quantitative mass spectrometry, enriching for 12 and 8 proteins, respectively. Overlapping between these groups were all known single-stranded telomere end binders, indicating both approaches can recapitulate previous findings in the field. This also substantiates a likely strong candidates list. Two candidates with high sequence similarity, R06A4.2 and T12E12.3, were further characterized in both an *in vitro* and *in vivo* context.

When expressed recombinantly in *E. coli*, both proteins exhibited clear enrichment at the telomere. Further biochemical assays revealed not only general telomere binding but also specific double-stranded telomere affinity within the nanomolar range. As a result, R06A4.2 and T12E12.3 were named Telomere End Binding Protein 1 and 2 (TEBP-1 and TEBP-2), respectively. When tagged with a CRISPR-Cas9 system at the native locus in *C. elegans*, both proteins showed clear nuclear and telomere-specific localization. To characterize the influence of TEBP-1 and TEBP-2 on telomere homeostasis, knockout strains were generated. Deletion of *tebp-1* resulted in significant telomere lengthening. This indicates that TEBP-1 could play a role in either blocking telomerase activity, allowing unhindered telomere extension upon its removal or in actively suppressing recombination based alternative lengthening of the telomeres (ALT). Previous studies in *C. elegans* have shown POT-2 plays a critical role in suppressing ALT (Cheng et al. 2012). While we are unable to map interaction points between POT-2 and TEBP-1 due to limitations in recombinant protein expression conditions, our immunoprecipitation-mass spectrometry (IP-MS) results exhibited interaction between POT-2 and TEBP-1. It is possible that both POT-2 and TEBP-1 work together to facilitate ALT.

In contrast, the deletion of *tebp-2* resulted in significant telomere shortening. This could indicate that TEBP-2 plays a role in the active recruitment of telomerase. It would be interesting to investigate if TEBP-2 influences the efficiency of telomerase in *C. elegans*, possibly through the utilization of a TRAP

assay (Mender and Shay 2015). Alternatively, TEBP-2 could be crucial for the protection of the telomeres. It is well-established that the shelterin complex in vertebrates plays a critical role in protecting the chromosome ends from inappropriate resection or repair (de Lange 2018). Many of the shelterin proteins block very particular pathways. However, the interplay between DNA repair and telomeres in *C. elegans* remains largely unaddressed.

When a double deletion strain of *tebp-1;tebp-2* was created, it resulted in the surprising phenotype of synthetic sterility, indicating an unanticipated role in fertility. To ensure that this phenotype was specific to the knockout of TEBP-1 and TEBP-2, we crossed all possible combinations of telomere binder deletion strains resulting in short and long telomeres. This confirmed that the phenotype was indeed particular to TEBP-1 and TEBP-2, rather than being conditional on telomere length phenotype. In future investigations, it would be valuable to explore the exact influence of these interactions on fertility. One hypothesis could revolve around the possible role of TEBP-1 and TEBP-2 in the previously described nuclear tethering facilitated by POT-1 during embryogenesis in *C. elegans* (Ferreira et al. 2013). Alternatively, it could be possible that TEBP-1 and TEBP-2 are critical to the formation of a D-loop telomere cap. Eliminating one of these proteins might still allow particle formation of the D-loop, while upon elimination of both of these proteins could leave a loose unprotected chromosome end (Shay and Wright 2019). This of course would likely lead to tremendous amounts of genome instability through improper DNA repair, chromosome fusions, improper chromosome segregation, and potential gross chromosomal rearrangements (Aguilera and García-Muse 2013). Such effects could explain the ultimately catastrophic fertility consequences observed when both proteins are not present.

To further investigate whether TEBP-1 and TEBP-2 played a role in the previously characterized telomeric complex in *C. elegans*, we conducted both size exclusion chromatography and IP-MS. Size exclusion chromatography for both proteins revealed peak elution sizes far larger than expected, indicating complex formation. In order to identify potential interactors, we performed IP-MS with both proteins. This revealed enrichment of the reciprocal TEBP protein and all known single-stranded telomere binders. This strongly suggests that TEBP-1 and TEBP-2 are forming a complex with the previously known telomere binders in *C. elegans*.

We then proceeded to map the precise activity of the different portions of TEBP-1 and TEBP-2. Thus far our data had indicated three types of interaction: 1) between TEBP-1 and TEBP-2, 2) between DNA, and 3) between known telomere end binders. The N-terminus was found to be responsible for the heterodimer formation of TEBP-1 and TEBP-2. It would be fascinating to determine the criticality of specific interaction points and investigate how the removal of these interaction points would affect telomere

125

homeostasis in *C. elegans*. The DNA binding domain was located within the central region of the proteins, and both TEBP-1 and TEBP-2 interacted with POT-1 at their C-terminus. This suggests that the DNA binding domain serves as an anchoring point, allowing the N- and C-termini to interact with other proteins of interest. To further explore these findings, it would be interesting to map the precise DNA-binding point of each protein. By removing this interaction point, we could assess whether each protein was completely removed from the telomere, or if the heterodimerization formation of TEBP-1 and TEBP-2 would continue to anchor its partner.

An additional aspect of this study is the characterization of other novel telomere-end binders identified in the screening process. In unpublished results by M. Almeida, S. Dietz, and myself, there have been four proteins of particular interest: F57C9.4, Bed-1, Lin-40, and Dve-1. Although the data from these findings is not included in this thesis, I assisted in the further characterization of F57C9.4 and Bed-1. With the assistance of Nadezda Podvalnaya, I established tagged strains of F57C9.4, and I discovered interactors of F57C9.4 and Bed-1 in both embryos and young adults (YAs). I also assessed the telomere length phenotype associated with these proteins. While precise interactions with telomeres were not observed in these studies, further functional characterization would establish their role in *C. elegans*. Continued investigation of F57C9.4, Bed-1, Lin-40, and Dve-1 would contribute to a more comprehensive understanding of the telomeric architecture in *C. elegans*.

Discovery of novel DNA damage factors across the tree of life

Here, we utilized a similar mass spectrometry-based phylointeractomics workflow, but rather than using a telomere sequence of interest, we compared the binding capabilities of three well-established DNA damage lesions: 8-oxoG, abasic site, and ribonucleotide incorporated into DNA. To gain a broader appreciation for repair and recognition of these lesions, we included 11 different species in our study. We enriched 337 proteins, of which 99 were related to the 'DNA repair' GO term. In addition to known DNA repair genes, we identified both species-specific and non-DNA repair proteins. These 82 species-specific proteins had no orthologs in the 10 other included species, offering the opportunity to study potentially unique aspects of repair or damage response in their respective model organisms. The discovery of these proteins would not have been possible without our unbiased approach. We also enriched known DNA repair protein homologs, which is particularly valuable for the species in which these proteins have not been characterized.

An unexpected result from the data set was the enrichment of photolyases and MutY glycosylases, both highly associated with DNA repair; these lesions included in the study are considered non-canonical

targets. Photolyases are an ancient family of DNA repair proteins that utilize energy from light to repair pyrimidine-pyrimidine dimers. It is intriguing that despite their strong conservation, this activity is lost in placental mammals (Mei and Dvornyk 2015). However, we observed five instances of specific enrichment in the 8-oxoG and abasic lesion pulldowns, without enrichment at the RNA lesion pulldown, despite detection. This indicates that there is lesion specific enrichment that also spans many organisms. It could be possible that an unanticipated intermediate is being recognized although further studies are needed to confirm these findings. Also, at the surface level, the enrichment of the MutY BER-associated glycosylase was not an immediate surprise, as it is considered a canonical BER protein. However, MutY has a strong preference for 8-oxoG::A (Krokan and Bjørås 2013), while the oligos included in this study contained 8-oxoG::C. Despite this, there were no instances of detection of a MUTYH ortholog without enrichment at 8-oxoG, indicating highly specific binding that was independent of the species. Some data suggests that MUTYH can behave as a general DDR scaffolding protein (Raetz and David 2019). Further investigations are needed to explore the true potential role of MUTYH in this context.

An unanticipated finding was the crosstalk between all three lesions. We had anticipated a high degree of overlap between the 8-oxoG and abasic lesions since they both heavily rely on BER for repair. However, we also observed a high degree of overlap with proteins enriched at the uracil incorporated into the DNA. We believe this finding warrants further investigation. While there have been minimal studies relating some BER proteins to RER, more extensive research is needed (Kellner and Luke 2020; Sassa, Yasui, and Honma 2019). One protein family that showed a high degree of conservation between species and lesions was the AP endonucleases. This protein family has been extensively studied in an evolutionary context. As early as *E. coli*, a distinction could be made between the ExoIII-like AP exonucleases and EndoIV-like AP endonucleases. This divergence was detected within the OrthoMCL groups as well as within our Maximum likelihood tree. However, we did not observe any enrichment bias based on these two groups.

Next, we focused on the enriched proteins which were not associated with the 'DNA repair' GO term. Through network, domain, and phylogenetic analysis, we identified 44 additional proteins likely to have a role in the DNA damage response. Network analysis revealed that although they were not designated with the 'DNA repair' GO term, these proteins interacted with many known DNA repair proteins. For instance, within the *S. cerevisiae* network there was an especially intricate network of chromatin remodelers. While few had DNA repair designation, many interacted with those DNA repair-associated chromatin remodelers and amongst themselves. This leads to the conclusion that there may be a more extensive network of chromatin remodelers involved in DNA repair than previously thought. To perform domain analysis we used Pfam to find major domains in the non-DNA repair proteins. We then curated these into

categories utilizing the domain descriptions. Finally, the most significant and compelling finding in this study was the use of phylogenetic analysis to identify enriched proteins across species that had not been previously associated with repair. We found five instances of these orthology groups, providing strong evidence that our screen successfully discovered novel DNA repair proteins across species.

Our study systematically evaluates binding partners in both BER lesions and an RNA lesion in eleven model species across the tree of life. We enriched proteins which had previously been associated with the recognition and repair of these lesions. However, not all previously described proteins were identified in our screen. This is to be expected as the physiological conditions of cells are highly specific that could not always be replicated, such as pH, temperature, salt concentration, ect. Nonetheless, our ability to identify classical repair proteins reinforces the validity of our screen. Through network, domain, and phylogenetics analysis, we identified a subset of proteins classified as non-DNA repair that are likely involved in DNA repair. Overall, this study paves the way for further investigation of newly identified candidates to explore key factors in the crosstalk between BER and RER DNA damage pathways.

A systems view on DNA damage response kinetics in *Tetrahymena thermophila*

Here, we conducted one of the largest systems views of DDR to date in the model organism *Tetrahymena thermophila*. In this study, our objective was to gain a deeper understanding of the underlying DDR kinetics in *Tetrahymena*. We collected cells for both transcriptome and proteome measurements over an eight-hour time course following damage induction with six different treatments. There was clear enrichment of known DNA damage factors, and an overall robust dynamic response to damage.

When performing hierarchical clustering of DNA repair proteins, our analysis revealed a distinct cluster of ten different DNA damage proteins, including MSH6L3 and RAD51. MSH6L3 and RAD51 were especially robust responders. RAD51 and other mismatch repair proteins have been shown to be critical DNA repair proteins during sexual reproduction in Tetrahymena (Loidl 2021; Howard-Till, Lukaszewicz, and Loidl 2011; Wang et al. 2023). It's intriguing to speculate about the specific role these proteins play in repair, which might mirror their roles in conjugation. To further investigate this, we are conducting deeper characterization of MSH6L3 and RAD51 in a repair context.

After examining the known DNA repair proteins, we evaluated the extent of the fold change and dynamicity score of each transcript and protein. While there is variability in the dynamic transcripts and proteins across treatments, we were interested in identifying treatments that had overlapping dynamic responses. When we conducted Fisher's exact test, we discovered a significantly higher level of overlap

than expected whenever there were three or more treatments, indicating a specific shared response alongside the overall global response. Additionally, we identified a core overlap of eight proteins between all treatments, which included RAD51 and MSH6L3. However, there was no overlap between the 15 core dynamic transcripts and the core eight dynamic proteins, highlighting the differential regulation in transcription and protein expression.

To have a deeper understanding of these dynamic proteins, we used self-organizing maps (SOMs), an unsupervised machine learning approach, to cluster the expression profiles of all six treatments. This results in 15 and 7 transcript and protein expression profile clusters, respectively. Within both types of clusters, we observed clear complexes grouping together within the same cluster, such as 20S proteasome and transcription-related factors. This indicates that the clustering method successfully identifies similarly regulated complexes. We also investigated the PARP and PARP-correlated proteins. While this protein family is absent in other well-established complex singular cellular eukaryotes, like *S. cerevisiae* and *S. pombe, Tetrahymena* has a family of 16 proteins containing a Poly(ADP-ribose) polymerase and DNA-Ligase Zn-finger domains (Citarelli, Teotia, and Lamb 2010). This provides a unique opportunity to study PARP proteins within a unicellular organism. In *Tetrahymena*, PARP proteins influence cellular growth and reproduction and have established interactions with H2A/H2B and variant Hv1 histones (Ashraf et al. 2019; Xiong et al. 2013). In human cells, among the 17 PARP related proteins, PARP-1 is considered the primary DDR contributor, while PARP-2 and PARP-3 play more minor roles in repair facilitation (Sanderson and Cohen 2020; Ray Chaudhuri and Nussenzweig 2017; Sousa et al. 2012). Based on this, we hypothesized that there might be one particular PARP protein or a small subset of PARP proteins that are primarily responsible for DNA repair in *Tetrahymena*. However, we observed specific up- or downregulation to unique treatments. Additionally, some of these PARP proteins exhibited variable transcriptional and protein responses. The global data analysis for the correlation or anticorrelation of transcript and protein expression profiles over the time course is still ongoing. Once developed, we intend to investigate the PARP protein family.

We are currently in the midst of implementing a system for experimental validation through a novel knockdown system, which will ultimately be published for the first time in this work. Once we successfully verify the integration of this construct, our next step is to examine the effects of reducing these PARP proteins on global protein expression changes. It is likely that there will be some form of compensatory DDR response, although it remains unknown whether this response will originate from other members of the PARP family or other DNA damage repair proteins. We will also assess the effects on cell survivability when these PARP proteins are reduced. Some knockdowns may exhibit sensitivity to

particular DNA damaging agents. Additionally, to further investigate the core DNA damage responders, we will specifically knockdown RAD51 and MSH6L3. We are intrigued to observe the effects of these knockdowns on repair processes and overall cellular health.

Another major aim of this work is to make this dataset highly accessible to the DNA damage and ciliate communities. Thus, before this work is submitted for publication, we will have made an online database. We hope that this propels ongoing research questions forward and opens up new areas of inquiry.

References

Ackerson, Stephanie M., Carlan Romney, P. Logan Schuck, and Jason A. Stewart. 2021. "To Join or Not to Join: Decision Points Along the Pathway to Double-Strand Break Repair vs. Chromosome End Protection." *Frontiers in Cell and Developmental Biology* 9. https://www.frontiersin.org/articles/10.3389/fcell.2021.708763.

Aguilera, Andrés, and Tatiana García-Muse. 2013. "Causes of Genome Instability." *Annual Review of Genetics* 47 (1): 1–32. https://doi.org/10.1146/annurev-genet-111212-133232.

Armanios, Mary. 2022. "The Role of Telomeres in Human Disease." *Annual Review of Genomics and Human Genetics* 23 (1): 363–81. https://doi.org/10.1146/annurev-genom-010422-091101.

Ashraf, Kanwal, Syed Nabeel-Shah, Jyoti Garg, Alejandro Saettone, Joanna Derynck, Anne-Claude Gingras, Jean-Philippe Lambert, Ronald E. Pearlman, and Jeffrey Fillingham. 2019. "Proteomic Analysis of Histones H2A/H2B and Variant Hv1 in Tetrahymena Thermophila Reveals an Ancient Network of Chaperones." *Molecular Biology and Evolution* 36 (5): 1037–55. https://doi.org/10.1093/molbev/msz039.

Bai, Miaomiao, Dongdong Ti, Qian Mei, Jiejie Liu, Xin Yan, Deyun Chen, Xiang Li, Zhiqiang Wu, and Weidong Han. 2020. "The Role of Posttranslational Modifications in DNA Repair." *BioMed Research International* 2020 (November): e7493902. https://doi.org/10.1155/2020/7493902.

Bantscheff, Marcus, Markus Schirle, Gavain Sweetman, Jens Rick, and Bernhard Kuster. 2007. "Quantitative Mass Spectrometry in Proteomics: A Critical Review." *Analytical and Bioanalytical Chemistry* 389 (4): 1017–31. https://doi.org/10.1007/s00216-007-1486-6.

Baumann, P., and T. R. Cech. 2001. "Pot1, the Putative Telomere End-Binding Protein in Fission Yeast and Humans." *Science (New York, N.Y.)* 292 (5519): 1171–75. https://doi.org/10.1126/science.1060036.

Beard, William A., Julie K. Horton, Rajendra Prasad, and Samuel H. Wilson. 2019. "Eukaryotic Base Excision Repair: New Approaches Shine Light on Mechanism." *Annual Review of Biochemistry* 88 (1): 137–62. https://doi.org/10.1146/annurev-biochem-013118-111315.

Bhardwaj, Chhavi, and Luke Hanley. 2014. "Ion Sources for Mass Spectrometric Identification and Imaging of Molecular Species." *Nat. Prod. Rep.* 31 (6): 756–67. https://doi.org/10.1039/C3NP70094A.

Bhargava, Ragini, David O. Onyango, and Jeremy M. Stark. 2016. "Regulation of Single Strand Annealing and Its Role in Genome Maintenance." *Trends in Genetics : TIG* 32 (9): 566–75. https://doi.org/10.1016/j.tig.2016.06.007.

Bhole, R. P., S. R. Jagtap, K. B. Chadar, and Y. B. Zambare. 2020. "Liquid Chromatography-Mass Spectrometry Technique-A Review." *Research Journal of Pharmacy and Technology* 13 (1): 505–16. https://doi.org/10.5958/0974-360X.2020.00097.9.

Bi, Xin. 2015. "Mechanism of DNA Damage Tolerance." *World Journal of Biological Chemistry* 6 (3): 48. https://doi.org/10.4331/wjbc.v6.i3.48.

Bluhm, Alina, Nikenza Viceconte, Fudong Li, Grishma Rane, Sandra Ritz, Suman Wang, Michal Levin, Yunyu Shi, Dennis Kappei, and Falk Butter. 2019. "ZBTB10 Binds the Telomeric Variant Repeat TTGGGG and Interacts with TRF2." *Nucleic Acids Research* 47 (4): 1896–1907. https://doi.org/10.1093/nar/gky1289.

Boersema, Paul J., Reinout Raijmakers, Simone Lemeer, Shabaz Mohammed, and Albert J. R. Heck. 2009. "Multiplex Peptide Stable Isotope Dimethyl Labeling for Quantitative Proteomics." *Nature Protocols* 4 (4): 484–94. https://doi.org/10.1038/nprot.2009.21.

Boesl, Ulrich. 2017. "Time-of-Flight Mass Spectrometry: Introduction to the Basics." *Mass Spectrometry Reviews* 36 (1): 86–109. https://doi.org/10.1002/mas.21520.

Broccoli, D., A. Smogorzewska, L. Chong, and T. de Lange. 1997. "Human Telomeres Contain Two Distinct Myb-Related Proteins, TRF1 and TRF2." *Nature Genetics* 17 (2): 231–35. https://doi.org/10.1038/ng1097-231.

Ceccaldi, Raphael, Beatrice Rondinelli, and Alan D. D'Andrea. 2016. "Repair Pathway Choices and Consequences at the Double-Strand Break." *Trends in Cell Biology* 26 (1): 52–64. https://doi.org/10.1016/j.tcb.2015.07.009.

Chang, Howard H. Y., Nicholas R. Pannunzio, Noritaka Adachi, and Michael R. Lieber. 2017. "Non-Homologous DNA End Joining and Alternative Pathways to Double-Strand Break Repair." *Nature Reviews Molecular Cell Biology* 18 (8): 495–506. https://doi.org/10.1038/nrm.2017.48.

Chen, Zhen, and Junjie Chen. 2021. "Mass Spectrometry-Based Protein–protein Interaction Techniques and Their Applications in Studies of DNA Damage Repair." *Journal of Zhejiang University. Science. B* 22 (1): 1–20. https://doi.org/10.1631/jzus.B2000356.

Cheng, Chen, Ludmila Shtessel, Megan M. Brady, and Shawn Ahmed. 2012. "Caenorhabditis Elegans POT-2 Telomere Protein Represses a Mode of Alternative Lengthening of Telomeres with Normal Telomere Lengths." *Proceedings of the National Academy of Sciences of the United States of America* 109 (20): 7805–10. https://doi.org/10.1073/pnas.1119191109.

Ciccia, Alberto, and Stephen J. Elledge. 2010. "The DNA Damage Response: Making It Safe to Play with Knives." *Molecular Cell* 40 (2): 179–204. https://doi.org/10.1016/j.molcel.2010.09.019.

Citarelli, Matteo, Sachin Teotia, and Rebecca S. Lamb. 2010. "Evolutionary History of the Poly(ADP-Ribose) Polymerase Gene Family in Eukaryotes." *BMC Evolutionary Biology* 10 (1): 308. https://doi.org/10.1186/1471-2148-10-308.

Coster, Gideon, and Michal Goldberg. 2010. "The Cellular Response to DNA Damage: A Focus on MDC1 and Its Interacting Proteins." *Nucleus* 1 (2): 166–78. https://doi.org/10.4161/nucl.1.2.11176.

Cox, Jürgen, Marco Y. Hein, Christian A. Luber, Igor Paron, Nagarjuna Nagaraj, and Matthias Mann. 2014. "Accurate Proteome-Wide Label-Free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ." *Molecular & Cellular Proteomics : MCP* 13 (9): 2513–26. https://doi.org/10.1074/mcp.M113.031591.

Cox, Jürgen, and Matthias Mann. 2008. "MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification." *Nature Biotechnology* 26 (12): 1367–72. https://doi.org/10.1038/nbt.1511.

Cox, Jürgen, Nadin Neuhauser, Annette Michalski, Richard A. Scheltema, Jesper V. Olsen, and Matthias Mann. 2011. "Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment." *Journal of Proteome Research* 10 (4): 1794–1805. https://doi.org/10.1021/pr101065j.

Dietz, Sabrina, Miguel Vasconcelos Almeida, Emily Nischwitz, Jan Schreier, Nikenza Viceconte, Albert Fradera-Sola, Christian Renz, et al. 2021. "The Double-Stranded DNA-Binding Proteins TEBP-1 and TEBP-2 Form a Telomeric Complex with POT-1." *Nature Communications* 12 (1): 2668. https://doi.org/10.1038/s41467-021-22861-2.

Elias, Joshua E., and Steven P. Gygi. 2010. "Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics." *Methods in Molecular Biology (Clifton, N.J.)* 604: 55–71. https://doi.org/10.1007/978-1-60761-444-9_5.

Eliuk, Shannon, and Alexander Makarov. 2015. "Evolution of Orbitrap Mass Spectrometry Instrumentation." *Annual Review of Analytical Chemistry* 8 (1): 61–80. https://doi.org/10.1146/annurev-anchem-071114-040325.

Fagbemi, Adebanke F., Barbara Orelli, and Orlando D. Schärer. 2011. "Regulation of Endonuclease Activity in Human Nucleotide Excision Repair." *DNA Repair* 10 (7): 722–29. https://doi.org/10.1016/j.dnarep.2011.04.022.

Ferreira, H.C., B.D. Towbin, T. Jegou, and S.M. Gasser. 2013. "The Shelterin Protein POT-1 Anchors

Caenorhabditis Elegans Telomeres through SUN-1 at the Nuclear Periphery." *The Journal of Cell Biology* 203 (5). https://doi.org/10.1083/jcb.201307181.

Gerber, Scott A., John Rush, Olaf Stemman, Marc W. Kirschner, and Steven P. Gygi. 2003. "Absolute Quantification of Proteins and Phosphoproteins from Cell Lysates by Tandem MS." *Proceedings of the National Academy of Sciences* 100 (12): 6940–45. https://doi.org/10.1073/pnas.0832254100.

Giglia-Mari, Giuseppina, Angelika Zotter, and Wim Vermeulen. 2011. "DNA Damage Response." *Cold Spring Harbor Perspectives in Biology* 3 (1): a000745. https://doi.org/10.1101/cshperspect.a000745.

Gredilla, Ricardo, Christian Garm, and Tinna Stevnsner. 2012. "Nuclear and Mitochondrial DNA Repair in Selected Eukaryotic Aging Model Systems." *Oxidative Medicine and Cellular Longevity* 2012: 282438. https://doi.org/10.1155/2012/282438.

Hamilton, Eileen, Peter Bruns, Cindy Lin, Virginia Merriam, Eduardo Orias, Linh Vong, and Donna Cassidy-Hanley. 2005. "Genome-Wide Characterization of Tetrahymena Thermophila Chromosome Breakage Sites. I. Cloning and Identification of Functional Sites." *Genetics* 170 (4): 1611–21. https://doi.org/10.1534/genetics.104.031401.

Howard-Till, Rachel A., Agnieszka Lukaszewicz, and Josef Loidl. 2011. "The Recombinases Rad51 and Dmc1 Play Distinct Roles in DNA Break Repair and Recombination Partner Choice in the Meiosis of Tetrahymena." *PLOS Genetics* 7 (3): e1001359. https://doi.org/10.1371/journal.pgen.1001359.

Hsu, Jue-Liang, Sheng-Yu Huang, Nan-Haw Chow, and Shu-Hui Chen. 2003. "Stable-Isotope Dimethyl Labeling for Quantitative Proteomics." *Analytical Chemistry* 75 (24): 6843–52. https://doi.org/10.1021/ac0348625.

Hughes, Christopher S., Sophie Moggridge, Torsten Müller, Poul H. Sorensen, Gregg B. Morin, and Jeroen Krijgsveld. 2019. "Single-Pot, Solid-Phase-Enhanced Sample Preparation for Proteomics Experiments." *Nature Protocols* 14 (1): 68–85. https://doi.org/10.1038/s41596-018-0082-x.

Hustedt, Nicole, and Daniel Durocher. 2017. "The Control of DNA Repair by the Cell Cycle." *Nature Cell Biology* 19 (1): 1–9. https://doi.org/10.1038/ncb3452.

Jahn, Arne, Grishma Rane, Maciej Paszkowski-Rogacz, Sergi Sayols, Alina Bluhm, Chung-Ting Han, Irena Draškovič, et al. 2017. "ZBTB48 Is Both a Vertebrate Telomere-binding Protein and a Transcriptional Activator." *EMBO Reports* 18 (6): 929–46. https://doi.org/10.15252/embr.201744095.

Jasin, Maria, and Rodney Rothstein. 2013. "Repair of Strand Breaks by Homologous Recombination." *Cold Spring Harbor Perspectives in Biology* 5 (11): a012740. https://doi.org/10.1101/cshperspect.a012740.

Jiricny, Josef. 2013. "Postreplicative Mismatch Repair." *Cold Spring Harbor Perspectives in Biology* 5 (4): a012633. https://doi.org/10.1101/cshperspect.a012633.

Kappei, Dennis, Falk Butter, Christian Benda, Marion Scheibe, Irena Draškovič, Michelle Stevense, Clara Lopes Novo, et al. 2013. "HOT1 Is a Mammalian Direct Telomere Repeat-Binding Protein Contributing to Telomerase Recruitment." *The EMBO Journal* 32 (12): 1681–1701. https://doi.org/10.1038/emboj.2013.105.

Kappei, Dennis, Marion Scheibe, Maciej Paszkowski-Rogacz, Alina Bluhm, Toni Ingolf Gossmann, Sabrina Dietz, Mario Dejung, et al. 2017. "Phylointeractomics Reconstructs Functional Evolution of Protein Binding." *Nature Communications* 8 (February): 14334. https://doi.org/10.1038/ncomms14334.

Kellner, Vanessa, and Brian Luke. 2020. "Molecular and Physiological Consequences of Faulty Eukaryotic Ribonucleotide Excision Repair." *The EMBO Journal* 39 (3): e102309. https://doi.org/10.15252/embj.2019102309.

Kim, Sahn-ho, Christian Beausejour, Albert R. Davalos, Patrick Kaminker, Seok-Jin Heo, and Judith Campisi. 2004. "TIN2 Mediates Functions of TRF2 at Human Telomeres." *The Journal of Biological Chemistry* 279 (42): 43799–804. https://doi.org/10.1074/jbc.M408650200.

Kisker, Caroline, Jochen Kuper, and Bennett Van Houten. 2013. "Prokaryotic Nucleotide Excision Repair." *Cold Spring Harbor Perspectives in Biology* 5 (3): a012591. https://doi.org/10.1101/cshperspect.a012591.

Kovalchuk, I. 2016. "Chapter 38 - Conserved and Divergent Features of DNA Repair: Future Perspectives in Genome Instability Research." In *Genome Stability*, edited by Igor Kovalchuk and Olga Kovalchuk, 651–66. Boston: Academic Press. https://doi.org/10.1016/B978-0-12-803309-8.00038-0.

Krokan, Hans E., and Magnar Bjørås. 2013. "Base Excision Repair." *Cold Spring Harbor Perspectives in Biology* 5 (4): a012583. https://doi.org/10.1101/cshperspect.a012583.

Kunkel, Thomas A., and Dorothy A. Erie. 2015. "Eukaryotic Mismatch Repair in Relation to DNA Replication." *Annual Review of Genetics* 49 (1): 291–313. https://doi.org/10.1146/annurev-genet-112414-054722.

Kurthkoti, Krishna, Pradeep Kumar, Pau Biak Sang, and Umesh Varshney. 2020. "Base Excision Repair Pathways of Bacteria: New Promise for an Old Problem." *Future Medicinal Chemistry* 12 (4): 339–55. https://doi.org/10.4155/fmc-2019-0267.

Lackner, Daniel H., Marcela Raices, Hugo Maruyama, Candy Haggblom, and Jan Karlseder. 2012. "Organismal Propagation in the Absence of a Functional Telomerase Pathway in Caenorhabditis Elegans." *The EMBO Journal* 31 (8): 2024–33. https://doi.org/10.1038/emboj.2012.61.

Lange, Titia de. 2018. "Shelterin-Mediated Telomere Protection." *Annual Review of Genetics* 52 (November): 223–47. https://doi.org/10.1146/annurev-genet-032918-021921.

Lanz, Michael Charles, Diego Dibitetto, and Marcus Bustamante Smolka. 2019. "DNA Damage Kinase Signaling: Checkpoint and Repair at 30 Years." *The EMBO Journal* 38 (18): e101801. https://doi.org/10.15252/embj.2019101801.

Lee, Kyoo-Young, and Su Hyung Park. 2020. "Eukaryotic Clamp Loaders and Unloaders in the Maintenance of Genome Stability." *Experimental & Molecular Medicine* 52 (12): 1948–58. https://doi.org/10.1038/s12276-020-00533-3.

Li, B., S. Oestreich, and T. de Lange. 2000. "Identification of Human Rap1: Implications for Telomere Evolution." *Cell* 101 (5): 471–83. https://doi.org/10.1016/s0092-8674(00)80858-2.

Li, Guo-Min. 2008. "Mechanisms and Functions of DNA Mismatch Repair." *Cell Research* 18 (1): 85–98. https://doi.org/10.1038/cr.2007.115.

Linger, Benjamin R., and Carolyn M. Price. 2009. "Conservation of Telomere Protein Complexes: Shuffling through Evolution." *Critical Reviews in Biochemistry and Molecular Biology* 44 (6): 434–46. https://doi.org/10.3109/10409230903307329.

Loidl, Josef. 2021. "Tetrahymena Meiosis: Simple yet Ingenious." *PLoS Genetics* 17 (7): e1009627. https://doi.org/10.1371/journal.pgen.1009627.

Loidl, Josef, and Alexander Lorenz. 2016. "DNA Double-Strand Break Formation and Repair in Tetrahymena Meiosis." *Seminars in Cell & Developmental Biology* 54 (June): 126–34. https://doi.org/10.1016/j.semcdb.2016.02.021.

Long, Hongan, David J Winter, Allan Y.-C Chang, Way Sung, Steven H Wu, Mariel Balboa, Ricardo B R Azevedo, Reed A Cartwright, Michael Lynch, and Rebecca A Zufall. 2016. "Low Base-Substitution Mutation Rate in the Germline Genome of the Ciliate Tetrahymena Thermophila." *Genome Biology and Evolution* 8 (12): 3629–39. https://doi.org/10.1093/gbe/evw223.

Mah, L.-J., A. El-Osta, and T. C. Karagiannis. 2010. "TH2AX: A Sensitive Molecular Marker of DNA Damage and Repair." *Leukemia* 24 (4): 679–86. https://doi.org/10.1038/leu.2010.6.

March, Raymond E. 2009. "Quadrupole Ion Traps." *Mass Spectrometry Reviews* 28 (6): 961–89. https://doi.org/10.1002/mas.20250.

Maréchal, Alexandre, and Lee Zou. 2013. "DNA Damage Sensing by the ATM and ATR Kinases." *Cold Spring Harbor Perspectives in Biology* 5 (9): a012716. https://doi.org/10.1101/cshperspect.a012716.

Marteijn, Jurgen A., Hannes Lans, Wim Vermeulen, and Jan H. J. Hoeijmakers. 2014. "Understanding Nucleotide Excision Repair and Its Roles in Cancer and Ageing." *Nature Reviews Molecular Cell Biology* 15 (7): 465–81. https://doi.org/10.1038/nrm3822.

Medhe, Sharad. 2018. "Mass Spectrometry: Detectors Review." *Chemical and Biomolecular Engineering* 3 (4): 51–58. https://doi.org/10.11648/j.cbe.20180304.11.

Mei, Qiming, and Volodymyr Dvornyk. 2015. "Evolutionary History of the Photolyase/Cryptochrome Superfamily in Eukaryotes." *PLoS ONE* 10 (9): e0135940. https://doi.org/10.1371/journal.pone.0135940.

Meier, Bettina, Louise J. Barber, Yan Liu, Ludmila Shtessel, Simon J. Boulton, Anton Gartner, and Shawn Ahmed. 2009. "The MRT-1 Nuclease Is Required for DNA Crosslink Repair and Telomerase Activity in Vivo in Caenorhabditis Elegans." *The EMBO Journal* 28 (22): 3549–63. https://doi.org/10.1038/emboj.2009.278.

Mender, Ilgen, and Jerry W. Shay. 2015. "Telomerase Repeated Amplification Protocol (TRAP)." *Bio-Protocol* 5 (22): e1657.

Moldovan, George-Lucian, and Alan D. D'Andrea. 2009. "How the Fanconi Anemia Pathway Guards the Genome." *Annual Review of Genetics* 43: 223–49. https://doi.org/10.1146/annurev-genet-102108-134222.

Molinaro, Caroline, Alain Martoriati, and Katia Cailliau. 2021. "Proteins from the DNA Damage Response: Regulation, Dysfunction, and Anticancer Strategies." *Cancers* 13 (15): 3819. https://doi.org/10.3390/cancers13153819.

Nadler, Wiebke Maria, Dietmar Waidelich, Alexander Kerner, Sabrina Hanke, Regina Berg, Andreas Trumpp, and Christoph Rösli. 2017. "MALDI versus ESI: The Impact of the Ion Source on Peptide Identification." *Journal of Proteome Research* 16 (3): 1207–15. https://doi.org/10.1021/acs.jproteome.6b00805.

Nagornov, Konstantin O., Oleg Y. Tsybin, Edith Nicol, Anton N. Kozhinov, and Yury O. Tsybin. 2022. "Fourier Transform Ion Cyclotron Resonance Mass Spectrometry at the True Cyclotron Frequency." *Mass Spectrometry Reviews* 41 (2): 314–37. https://doi.org/10.1002/mas.21681.

Nesta, Alex V., Denisse Tafur, and Christine R. Beck. 2021. "Hotspots of Human Mutation." *Trends in Genetics* 37 (8): 717–29. https://doi.org/10.1016/j.tig.2020.10.003.

Nick McElhinny, Stephanie A., Brian E. Watts, Dinesh Kumar, Danielle L. Watt, Else-Britt Lundström, Peter M. J. Burgers, Erik Johansson, Andrei Chabes, and Thomas A. Kunkel. 2010. "Abundant Ribonucleotide Incorporation into DNA by Yeast Replicative Polymerases." *Proceedings of the National Academy of Sciences* 107 (11): 4949–54. https://doi.org/10.1073/pnas.0914857107.

Nischwitz, Emily, Vivien A.C. Schoonenberg, Albert Fradera-Sola, Mario Dejung, Olga Vydzhak, Michal Levin, Brian Luke, Falk Butter, and Marion Scheibe. 2023. "DNA Damage Repair Proteins across the Tree of Life." *IScience*, April, 106778. https://doi.org/10.1016/j.isci.2023.106778.

Orias, Eduardo, Marcella D. Cervantes, and Eileen P. Hamilton. 2011. "Tetrahymena Thermophila, a Unicellular Eukaryote with Separate Germline and Somatic Genomes." *Research in Microbiology* 162 (6): 578–86. https://doi.org/10.1016/j.resmic.2011.05.001.

Palm, Wilhelm, and Titia de Lange. 2008. "How Shelterin Protects Mammalian Telomeres." *Annual Review of Genetics* 42: 301–34. https://doi.org/10.1146/annurev.genet.41.110306.130350.

Pećina-Šlaus, Nives, Anja Kafka, Iva Salamon, and Anja Bukovac. 2020. "Mismatch Repair Pathway, Genome Stability and Cancer." *Frontiers in Molecular Biosciences* 7.

https://www.frontiersin.org/articles/10.3389/fmolb.2020.00122.

Podhorecka, Monika, Andrzej Skladanowski, and Przemyslaw Bozko. 2010. "H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy." *Journal of Nucleic Acids* 2010 (August): 920161. https://doi.org/10.4061/2010/920161.

Pomerantz, Richard T., and Mike O'Donnell. 2010. "What Happens When Replication and Transcription Complexes Collide?" *Cell Cycle* 9 (13): 2537–43. https://doi.org/10.4161/cc.9.13.12122.

Prorok, Paulina, Inga R. Grin, Bakhyt T. Matkarimov, Alexander A. Ishchenko, Jacques Laval, Dmitry O. Zharkov, and Murat Saparbaev. 2021. "Evolutionary Origins of DNA Repair Pathways: Role of Oxygen Catastrophe in the Emergence of DNA Glycosylases." *Cells* 10 (7): 1591. https://doi.org/10.3390/cells10071591.

QIU, Shan, and Jun HUANG. 2021. "MRN Complex Is an Essential Effector of DNA Damage Repair." *Journal of Zhejiang University. Science. B* 22 (1): 31–37. https://doi.org/10.1631/jzus.B2000289.

Raetz, Alan G., and Sheila S. David. 2019. "When You're Strange: Unusual Features of the MUTYH Glycosylase and Implications in Cancer." *DNA Repair* 80 (August): 16–25. https://doi.org/10.1016/j.dnarep.2019.05.005.

Raices, Marcela, Ramiro E. Verdun, Sarah A. Compton, Candy I. Haggblom, Jack D. Griffith, Andrew Dillin, and Jan Karlseder. 2008. "C. Elegans Telomeres Contain G-Strand and C-Strand Overhangs That Are Bound by Distinct Proteins." *Cell* 132 (5): 745–57. https://doi.org/10.1016/j.cell.2007.12.039.

Rappsilber, Juri, Matthias Mann, and Yasushi Ishihama. 2007. "Protocol for Micro-Purification, Enrichment, Pre-Fractionation and Storage of Peptides for Proteomics Using StageTips." *Nature Protocols* 2 (8): 1896–1906. https://doi.org/10.1038/nprot.2007.261.

Rauniyar, Navin, and John R. Yates. 2014. "Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics." *Journal of Proteome Research* 13 (12): 5293–5309. https://doi.org/10.1021/pr500880b.

Ray Chaudhuri, Arnab, and André Nussenzweig. 2017. "The Multifaceted Roles of PARP1 in DNA Repair and Chromatin Remodelling." *Nature Reviews. Molecular Cell Biology* 18 (10): 610–21. https://doi.org/10.1038/nrm.2017.53.

Robertson, A. B., A. Klungland, T. Rognes, and I. Leiros. 2009. "DNA Repair in Mammalian Cells: Base Excision Repair: The Long and Short of It." *Cellular and Molecular Life Sciences: CMLS* 66 (6): 981–93. https://doi.org/10.1007/s00018-009-8736-z.

Rogers, John C., and Ryan D. Bomgarden. 2016. "Sample Preparation for Mass Spectrometry-Based Proteomics; from Proteomes to Peptides." In *Modern Proteomics – Sample Preparation, Analysis and Practical Applications*, edited by Hamid Mirzaei and Martin Carrasco, 43–62. Advances in Experimental Medicine and Biology. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-41448-5_3.

Saldivar, Joshua C., David Cortez, and Karlene A. Cimprich. 2017. "The Essential Kinase ATR: Ensuring Faithful Duplication of a Challenging Genome." *Nature Reviews. Molecular Cell Biology* 18 (10): 622–36. https://doi.org/10.1038/nrm.2017.67.

Sanderson, Daniel J., and Michael S. Cohen. 2020. "Mechanisms Governing PARP Expression, Localization, and Activity in Cells." *Critical Reviews in Biochemistry and Molecular Biology* 55 (6): 541–54. https://doi.org/10.1080/10409238.2020.1818686.

Sassa, Akira, Manabu Yasui, and Masamitsu Honma. 2019. "Current Perspectives on Mechanisms of Ribonucleotide Incorporation and Processing in Mammalian DNA." *Genes and Environment* 41 (1): 3. https://doi.org/10.1186/s41021-019-0118-7.

Saxena, Sneha, and Lee Zou. 2022. "Hallmarks of DNA Replication Stress." *Molecular Cell* 82 (12): 2298–2314. https://doi.org/10.1016/j.molcel.2022.05.004.

Schärer, Orlando D. 2013. "Nucleotide Excision Repair in Eukaryotes." *Cold Spring Harbor Perspectives*

in Biology 5 (10): a012609. https://doi.org/10.1101/cshperspect.a012609.

Sfeir, Agnel, and Lorraine S. Symington. 2015. "Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway?" *Trends in Biochemical Sciences* 40 (11): 701–14. https://doi.org/10.1016/j.tibs.2015.08.006.

Shay, Jerry W., and Woodring E. Wright. 2019. "Telomeres and Telomerase: Three Decades of Progress." *Nature Reviews. Genetics* 20 (5): 299–309. https://doi.org/10.1038/s41576-019-0099-1.

Shevchenko, Andrej, Henrik Tomas, Jan Havlis, Jesper V. Olsen, and Matthias Mann. 2006. "In-Gel Digestion for Mass Spectrometric Characterization of Proteins and Proteomes." *Nature Protocols* 1 (6): 2856–60. https://doi.org/10.1038/nprot.2006.468.

Shtessel, Ludmila, Mia Rochelle Lowden, Chen Cheng, Matt Simon, Kyle Wang, and Shawn Ahmed. 2013. "Caenorhabditis Elegans POT-1 and POT-2 Repress Telomere Maintenance Pathways." *G3 (Bethesda, Md.)* 3 (2): 305–13. https://doi.org/10.1534/g3.112.004440.

Shyian, Maksym, and David Shore. 2021. "Approaching Protein Barriers: Emerging Mechanisms of Replication Pausing in Eukaryotes." *Frontiers in Cell and Developmental Biology* 9 (May): 672510. https://doi.org/10.3389/fcell.2021.672510.

Silva, André M. N., Rui Vitorino, M. Rosário M. Domingues, Corinne M. Spickett, and Pedro Domingues. 2013. "Post-Translational Modifications and Mass Spectrometry Detection." *Free Radical Biology and Medicine* 65 (December): 925–41. https://doi.org/10.1016/j.freeradbiomed.2013.08.184.

Sinha, Ankit, and Matthias Mann. 2020. "A Beginner's Guide to Mass Spectrometry–Based Proteomics." *The Biochemist* 42 (5): 64–69. https://doi.org/10.1042/BIO20200057.

Sousa, Fabricio G., Renata Matuo, Daniele G. Soares, Alexandre E. Escargueil, João A.P. Henriques, Annette K. Larsen, and Jenifer Saffi. 2012. "PARPs and the DNA Damage Response." *Carcinogenesis* 33 (8): 1433–40. https://doi.org/10.1093/carcin/bgs132.

Stadler, Jens, and Holger Richly. 2017. "Regulation of DNA Repair Mechanisms: How the Chromatin Environment Regulates the DNA Damage Response." *International Journal of Molecular Sciences* 18 (8): 1715. https://doi.org/10.3390/ijms18081715.

Stokes, Matthew P., Yiying Zhu, and Charles L. Farnsworth. 2018. "Mass Spectrometry-Based Proteomic Analysis of the DNA Damage Response." *Frontiers in Bioscience (Landmark Edition)* 23 (4): 597–613. https://doi.org/10.2741/4607.

Sugasawa, Kaoru, Jun-ichi Akagi, Ryotaro Nishi, Shigenori Iwai, and Fumio Hanaoka. 2009. "Two-Step Recognition of DNA Damage for Mammalian Nucleotide Excision Repair: Directional Binding of the XPC Complex and DNA Strand Scanning." *Molecular Cell* 36 (4): 642–53. https://doi.org/10.1016/j.molcel.2009.09.035.

Sung, Way, Abraham E. Tucker, Thomas G. Doak, Eunjin Choi, W. Kelley Thomas, and Michael Lynch. 2012. "Extraordinary Genome Stability in the Ciliate *Paramecium Tetraurelia*." *Proceedings of the National Academy of Sciences* 109 (47): 19339–44. https://doi.org/10.1073/pnas.1210663109.

Tatum, Danielle, Shisheng Li, Danielle Tatum, and Shisheng Li. 2011. *Nucleotide Excision Repair in S. Cerevisiae. DNA Repair - On the Pathways to Fixing DNA Damage and Errors*. IntechOpen. https://doi.org/10.5772/22129.

Thomas, Stefani N., Deborah French, Paul J. Jannetto, Brian A. Rappold, and William A. Clarke. 2022. "Liquid Chromatography–Tandem Mass Spectrometry for Clinical Diagnostics." *Nature Reviews Methods Primers* 2 (1): 1–14. https://doi.org/10.1038/s43586-022-00175-x.

Tyanova, Stefka, Tikira Temu, and Juergen Cox. 2016. "The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics." *Nature Protocols* 11 (12): 2301–19. https://doi.org/10.1038/nprot.2016.136.

Vivelo, Christina A., and Anthony K. L. Leung. 2015. "Proteomics Approaches to Identify Mono(ADP-Ribosyl)Ated and Poly(ADP-Ribosyl)Ated Proteins." *Proteomics* 15 (0): 203–17.

https://doi.org/10.1002/pmic.201400217.

Wang, Lin, Yuhuan Xue, Sitong Yang, Tao Bo, Jing Xu, and Wei Wang. 2023. "Mismatch Repair Protein Msh2 Is Necessary for Macronuclear Stability and Micronuclear Division in Tetrahymena Thermophila." *International Journal of Molecular Sciences* 24 (13): 10559. https://doi.org/10.3390/ijms241310559.

Williams, Ashley B., and Björn Schumacher. 2016. "P53 in the DNA-Damage-Repair Process." *Cold Spring Harbor Perspectives in Medicine* 6 (5): a026070. https://doi.org/10.1101/cshperspect.a026070.

Wiśniewski, Jacek R. 2018. "Filter-Aided Sample Preparation for Proteome Analysis." *Methods in Molecular Biology (Clifton, N.J.)* 1841: 3–10. https://doi.org/10.1007/978-1-4939-8695-8_1.

Wright, William Douglass, Shanaya Shital Shah, and Wolf-Dietrich Heyer. 2018. "Homologous Recombination and the Repair of DNA Double-Strand Breaks." *Journal of Biological Chemistry* 293 (27): 10524–35. https://doi.org/10.1074/jbc.TM118.000372.

Wu, R. Alex, Heather E. Upton, Jacob M. Vogan, and Kathleen Collins. 2017. "Telomerase Mechanism of Telomere Synthesis." *Annual Review of Biochemistry* 86 (June): 439–60. https://doi.org/10.1146/annurev-biochem-061516-045019.

Xin, Zhong-Tao, Adam D. Beauchamp, Rodrigo T. Calado, Jennifer W. Bradford, Joshua A. Regal, Aarthi Shenoy, Yuying Liang, Peter M. Lansdorp, Neal S. Young, and Hinh Ly. 2007. "Functional Characterization of Natural Telomerase Mutations Found in Patients with Hematologic Disorders." *Blood* 109 (2): 524–32. https://doi.org/10.1182/blood-2006-07-035089.

Xiong, Jie, Yuming Lu, Jinmei Feng, Dongxia Yuan, Miao Tian, Yue Chang, Chengjie Fu, Guangying Wang, Honghui Zeng, and Wei Miao. 2013. "Tetrahymena Functional Genomics Database (TetraFGD): An Integrated Resource for Tetrahymena Functional Genomics." *Database* 2013 (January): bat008. https://doi.org/10.1093/database/bat008.

Yue, Xiaoqiao, Chenjun Bai, Dafei Xie, Teng Ma, and Ping-Kun Zhou. 2020. "DNA-PKcs: A Multi-Faceted Player in DNA Damage Response." *Frontiers in Genetics* 11. https://www.frontiersin.org/articles/10.3389/fgene.2020.607428.

Zeman, Michelle K., and Karlene A. Cimprich. 2014. "Causes and Consequences of Replication Stress." *Nature Cell Biology* 16 (1): 2–9. https://doi.org/10.1038/ncb2897.

Acknowledgements