

H3K4me1 recruits DNA repair proteins in plants

Daniela Quiroz ,^{1,2} Satoyo Oya ,^{1,3} Diego Lopez-Mateos ,^{4,5} Kehan Zhao ,^{1,6} Alice Pierce ,^{1,6} Lissandro Ortega ,¹ Alissza Ali ,¹ Pablo Carbonell-Bejerano ,⁷ Vladimir Yarov-Yarovoy ,^{4,5} Sae Suzuki ,⁸ Gosuke Hayashi ,⁸ Akihisa Osakabe ,^{3,9} Grey Monroe ,^{1,2,6,*}

- 1 Department of Plant Sciences, University of California Davis, Davis, CA 95616, USA
- 2 Integrative Genetics and Genomics, University of California Davis, Davis, CA 95616, USA
- 3 Laboratory of Genetics, Department of Biological Sciences, The University of Tokyo, Tokyo 113-0033, Japan
- 4 Department of Physiology and Membrane Biology, University of California Davis, Davis, CA 95616, USA
- 5 Biophysics Graduate Group, University of California Davis, Davis, CA 95616, USA
- 6 Plant Biology Graduate Group, University of California Davis, Davis, CA 95616, USA
- 7 Instituto de Ciencias de la Vid y del Vino (ICVV, CSIC-CAR-UR), Logroño 26007, La Rioja, Spain
- 8 Department of Biomolecular Engineering, Graduate School of Engineering, Nagoya University, Nagoya 464-0814, Japan
- 9 PRESTO, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan

*Author for correspondence: gmonroe@ucdavis.edu

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (<https://academic.oup.com/plcell/pages/General-Instructions>) is J. Grey Monroe (gmonroe@ucdavis.edu).

Abstract

DNA repair proteins can be recruited by their histone reader domains to specific epigenomic features, with consequences on intragenomic mutation rate variation. Here, we investigated H3K4me1-associated hypomutation in plants. We first examined 2 proteins which, in plants, contain Tudor histone reader domains: PRECOCIOUS DISSOCIATION OF SISTERS 5 (PDS5C), involved in homology-directed repair, and MUTS HOMOLOG 6 (MSH6), a mismatch repair protein. The MSH6 Tudor domain of *Arabidopsis* (*Arabidopsis thaliana*) binds to H3K4me1 as previously demonstrated for PDS5C, which localizes to H3K4me1-rich gene bodies and essential genes. Mutations revealed by ultradeep sequencing of wild-type and *msh6* knockout lines in *Arabidopsis* show that functional MSH6 is critical for the reduced rate of single-base substitution (SBS) mutations in gene bodies and H3K4me1-rich regions. We explored the breadth of these mechanisms among plants by examining a large rice (*Oryza sativa*) mutation data set. H3K4me1-associated hypomutation is conserved in rice as are the H3K4me1-binding residues of MSH6 and PDS5C Tudor domains. Recruitment of DNA repair proteins by H3K4me1 in plants reveals convergent, but distinct, epigenome-recruited DNA repair mechanisms from those well described in humans. The emergent model of H3K4me1-recruited repair in plants is consistent with evolutionary theory regarding mutation modifier systems and offers mechanistic insight into intragenomic mutation rate variation in plants.

Introduction

DNA is damaged on the scale of 10,000 to 100,000 times per day in every cell (Alberts et al. 2002; Manova and Gruszka 2015; Nisa et al. 2019; Yousefzadeh et al. 2021). Preventing this widespread damage from becoming mutation requires the action of DNA repair proteins. Mechanisms that localize DNA repair proteins to certain epigenomic contexts, such as by recruitment to

certain histone modifications, can therefore reduce mutation rates in those regions of the genome. Such interactions between DNA repair and histone modifications are predicted to evolve if they promote repair in regions prone to deleterious mutations, such as gene bodies (Lynch 2010; Martincorena and Luscombe 2013; Lynch et al. 2016, 2023).

Associations between histone modifications and mutation rates have been established across diverse organisms

Received January 09, 2024. Accepted February 13, 2024. Advance access publication March 26, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of American Society of Plant Biologists. All rights reserved. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

(Schuster-Böckler and Lehner 2012; Makova and Hardison 2015; Habig et al. 2021; Yan et al. 2021; Yang et al. 2021; de la Peña et al. 2022; Monroe, Srikant, et al. 2022; Monroe et al. 2023; Staunton et al. 2023). Mechanisms responsible for such mutation biases are well characterized in humans (Foster et al. 2015; Supek and Lehner 2015, 2017, 2019; Katju et al. 2022). In humans, H3K36me3 recruits DNA repair via their PWWP domains in proteins contributing to homology-directed (zinc finger CW type and PWWP domain-containing ZCWPW1/2 and lens epithelium-derived growth factor LEDGF/p75) and mismatch repair (MMP) (MutS homolog MSH6), with H3K36me3 marking the gene bodies and exons of active genes (Li et al. 2013; Aymard et al. 2014; Huang et al. 2018; Sun et al. 2020; Fang et al. 2021; Xiao et al. 2021). As predicted, reduced mutation rates in H3K36me3-marked regions, active genes, and gene bodies have been observed in humans and other animals (Supek and Lehner 2017; Akdemir et al. 2020; Li et al. 2021; Moore et al. 2021; Chen et al. 2022; Katju et al. 2022).

Reduced mutation rates in gene bodies and active genes have also been observed in some algae and land plants, but the mechanisms have, to date, been unresolved (Krasovec et al. 2017; Belfield et al. 2018, 2021; López-Cortegano et al. 2021; Lu et al. 2021; Yan et al. 2021; Zhu et al. 2021; Monroe, Srikant, et al. 2022). In *Arabidopsis* (*Arabidopsis thaliana*), knockout lines of *msh2*, the MMP protein that dimerizes with MSH6 to form MutSa (Kolodner 1996; Adé et al. 1999), indicate that MMP preferentially repairs gene bodies in plants. Yet, how plants recruit DNA repair proteins to gene bodies has remained unclear (Belfield et al. 2018). While such mutation biases are well-established in humans, recent reports of these mutation biases in plants have been met with skepticism by some. Understanding the mechanisms responsible for biases like gene body hypomutation in plants is therefore critical for addressing ongoing debates and research on mutation rate heterogeneity (Charlesworth and Jensen 2022; Liu and Zhang 2022; Monroe, Murray, et al. 2022; Monroe, Srikant, et al. 2022; Monroe et al. 2023; Kenchanmane Raju et al. 2023; Majic and Payne 2023; Sutherland et al., 2023; Wang et al. 2023).

Mutation accumulation lines and natural variation in *A. thaliana* reveal H3K4me1 as an important mark associated with lower mutation rates (Monroe, Srikant, et al. 2022; Monroe et al. 2023; Kenchanmane Raju et al. 2023). Recent studies of CRISPR-mediated mutation efficiency further support a link between H3K4me1 and DNA repair, showing that H3K4me1 is associated with lower rates of insertion mutations (Schep et al. 2021; Zhu et al. 2021; Weiss et al., 2022). Still, this link between histone marks and mutation in plants remains contentious, necessitating further work to provide mechanistic clarity (Charlesworth and Jensen 2022; Liu and Zhang 2022; Monroe, Murray, et al. 2022; Monroe, Srikant, et al. 2022; Monroe et al. 2023; Majic and Payne 2023; Wang et al. 2023).

In plants, H3K4me1 marks gene bodies of active genes (Zhang et al. 2009; Liu et al. 2019; Lu et al. 2019; Niu et al. 2021; Oya et al., 2021). Recent work has demonstrated that H3K4me1 enrichment is largely mediated by a combination of transcription-

coupled (ARABIDOPSIS TRITHORAX-RELATED 7 [ATXR7]) and epigenome-guided (ARABIDOPSIS TRITHORAX 1 [ATX1] and ATX2) histone methyltransferases and demethylases (LYSINE-SPECIFIC DEMETHYLASE HOMOLOG 1 [LDL2], LDL3, and FLOWERING LOCUS D [FLD]) (Inagaki et al. 2017, 2021; Oya et al., 2021; Mori et al., 2023). Once established, H3K4me1 reading can occur by proteins containing histone reader domains such as “Royal family” Tudor domains, which bind methylated lysine residues on H3 histone tails (Maurer-Stroh et al. 2003; Kim et al. 2006; Lu and Wang 2013). Recently, the Tudor domain of PRECOCIOUS DISSOCIATION OF SISTERS 5 (PDS5C) was shown to specifically bind H3K4me1 in *A. thaliana* (Niu et al. 2021). This protein is a cohesion cofactor, important for homology-directed repair (HDR) of double-strand DNA breaks (Wang et al. 2002; Jin et al. 2009; Kusch 2015; Pradillo et al. 2015; Couturier et al. 2016; Morales et al. 2020; Zhang et al. 2021; Phipps and Dubrana 2022). The well-characterized H3K4me1-binding Tudor domain of PDS5C provides a starting point to identify other potential DNA repair proteins interacting with H3K4me1 in plants.

In this study, we develop a mechanistic model to elucidate reduced mutation rates linked with H3K4me1 in plant genomes. We demonstrate that in *A. thaliana*, the PDS5C protein, whose Tudor domain specifically binds to H3K4me1, is actively recruited to gene bodies of essential genes. Fluorescent polarization experiments further confirm that the Tudor domain of MSH6 has a specific affinity for H3K4me1. By examining *msh6* knockout lines, we identified changes in genomic mutation rates, demonstrating that MSH6 is important for lower mutation rates in gene bodies and areas enriched with H3K4me1. Additionally, in silico analysis reveals that Tudor domains in the MSH6 and PDS5C homologs from rice (*Oryza sativa*) exhibit a conserved binding preference for H3K4me1. These results imply a conserved role for H3K4me1 in recruiting DNA repair mechanisms, influencing mutation patterns in plant genomes. Further investigations into de novo SBS mutations in Kitaake rice, using data from whole-genome sequenced fast neutron mutation accumulation lines (Li et al. 2017), revealed H3K4me1-associated hypomutation akin to those observed in *A. thaliana*. These mechanisms are consistent with the evolution of mutation modifier systems, highlighting that epigenome-recruited repair processes can control mutation rate variation in plants.

Results and discussion

PDS5C targets H3K4me1 and is associated with regions of lower mutation rates

While in humans, reductions in mutation rates are known to be mediated by H3K36me3 recruitment of DNA repair genes (via PWWP histone reader domains), previous results suggest that H3K4me1 may be a target of DNA repair in plants (Monroe et al. 2023; Monroe, Srikant, et al. 2022). To examine this further, we considered proteins with known H3K4me1 targeting in *A. thaliana*.

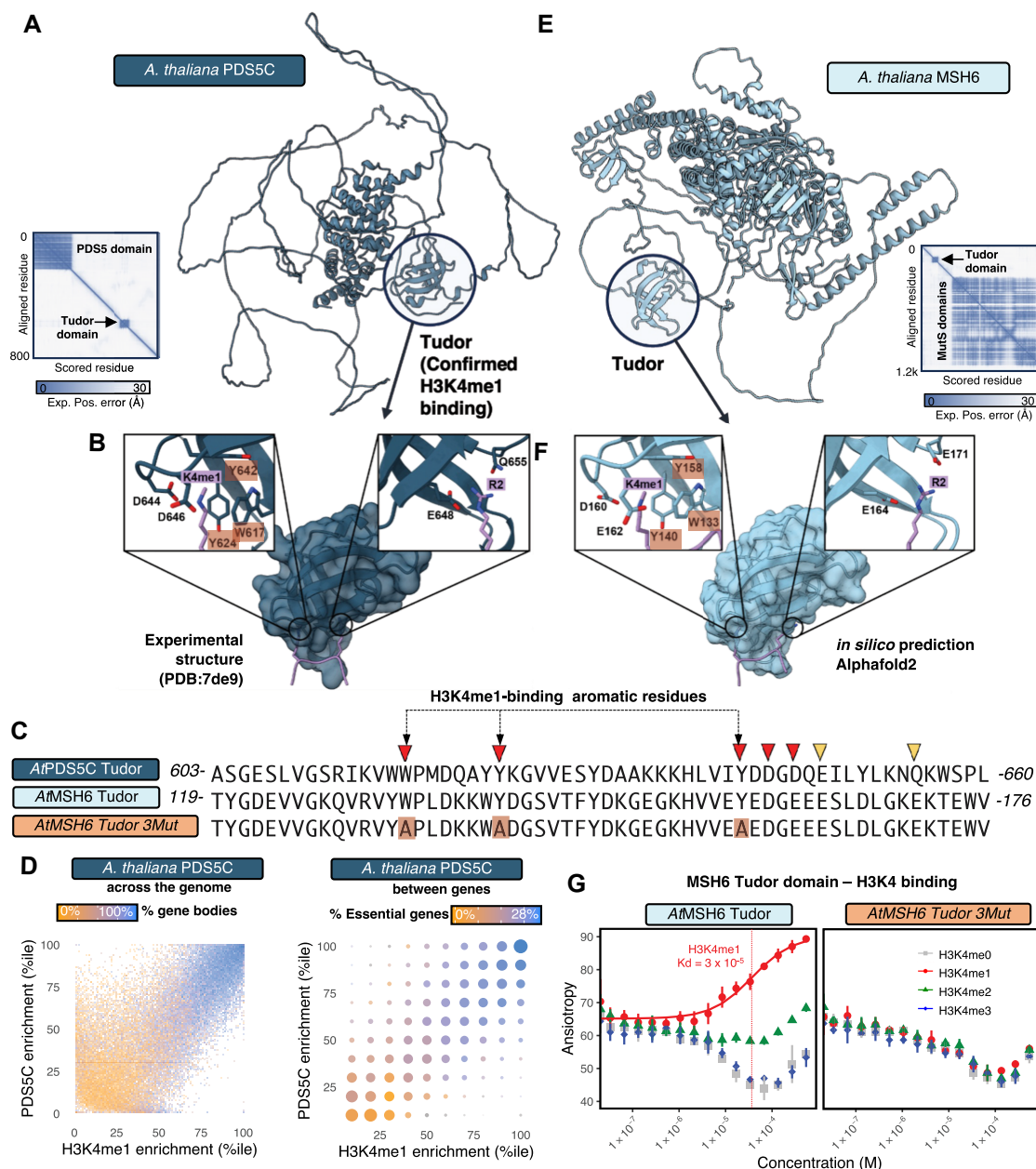


Figure 1. *A. thaliana* PDS5C and MSH6 Tudor domains have affinity to H3K4 methylation. **A)** AlphaFold2 structure of PDS5C protein from *A. thaliana* indicates Tudor domain is tethered to active domain **B)** Aromatic cage and H3K4me1 in experimental crystal structure of At_PDS5C (PDB: 7DE9) (Niu et al. 2021). **C)** ClustalW alignment of Tudor domains. Highlighted arrows mark residues binding H3K4me1 specifically, and yellow arrows mark residues interacting with H3R2 (Niu et al. 2021). Artificial triple mutant (AtMSH6 Tudor 3Mut) synthesized for fluorescent polarization assay sequence is also shown. **D)** Enrichment of PDS5C and H3K4me1 ChIP-seq data across the genome (200 bp windows, ranked) and between genes (ranked). Points in the right panel are colored by the fraction of regions overlapping gene bodies. The size of points in the left panel is scaled to the number of genes, colored by the fraction of genes that are classified as essential (Lloyd et al. 2015). **E)** AlphaFold2 structure of MSH6 and **F)** the best scored model of FlexPepDock in silico prediction of MSH6 Tudor domain-binding H3K4me1. **G)** Fluorescent polarization results from MSH6 Tudor domain with H3K4, H3K4me1, H3K4me2, and H3K4me3 peptides (left plot). Fluorescent polarization of MSH6 Tudor triple mutant (Aromatic cage forming residues: W133A, Y140A, and Y158A). Error bars show standard deviation for $n = 16$.

PDS5C, belonging to a family of cohesion cofactors that facilitate HDR, contains a Tudor domain that was recently described to specifically bind H3K4me1 (Niu et al. 2021) (Fig. 1, A to C). Analyses of ChIP-seq data of PDS5C-Flag from *A. thaliana* confirm PDS5C localizes to regions enriched for H3K4me1: gene

bodies and essential genes (Fig. 1D, Supplementary Fig. S1). We also find that PDS5C localized to regions of lower germline mutation rates in *A. thaliana*, which include only high-quality variants, and thus is unlikely to be due to sequencing errors (Monroe et al. 2023; Wang et al. 2023) (Supplementary Fig. S1).

Evolutionary models predict that epigenome-recruited repair mechanisms should evolve if they promote repair in sequences prone to purifying selection (Martincorena and Luscombe 2013; Lynch et al. 2016). As predicted by this theory, we find that PDS5C is recruited to coding sequences (CDSs), especially in essential genes, genes broadly expressed across tissues, and genes subject to stronger purifying selection in natural populations of *A. thaliana* (Supplementary Fig. S1).

The predicted *A. thaliana* PDS5C structure reveals that the PDS5C active domain is separated from the Tudor domain by a disordered and potentially flexible segment (Fig. 1A). PDS5 proteins are central in cohesion between sister chromatids required for HDR of double-strand breaks (Ren et al. 2005; Schubert et al. 2009; Bolaños-Villegas et al. 2013; Hill et al. 2016; Morales et al. 2020; Phipps and Dubrana 2022). The importance of PDS5 proteins for HDR has been confirmed experimentally in *A. thaliana*, with PDS5 mutants being hypersensitive to DNA damage (Pradillo et al. 2015). The observation that mutation rates are reduced at PDS5C and H3K4me1 peak regions supports the hypothesis that Tudor domain-mediated targeting in PDS5C, its putative paralogs (PDS5A, B, D, and E), or other repair-related proteins contributes to targeted hypomutation in the functionally important regions of the genome. Still, additional experiments are needed to quantify the effect of PDS5 proteins on mutation rate heterogeneity, necessitating experiments that can capture InDel mutations and overcome potential redundancies among PDS5 putative paralogs. As discussed below, the PDS5C Tudor domain ultimately led us to experimental study of MSH6, which represents a unique opportunity to study a convergent mechanism of epigenome-recruited DNA repair.

MSH6 has a conserved Tudor domain that preferentially binds H3K4me1

The discovery of the PDS5C Tudor domain as an H3K4me1-recruited domain (Niu et al. 2021) provided an opportunity to identify other proteins with potential for H3K4me1 recruitment. We used *blastp* to search the *A. thaliana* proteome for other proteins containing Tudor domains similar to that of PDS5C. These 29 genes were enriched for DNA repair functions ($P = 1 \times 10^{-11}$). Five of these were PDS5C homologs (PDS5A-E). We also found that MSH6, a DNA MMP protein, contains a Tudor domain similar to that of PDS5C, which was a compelling candidate for further consideration (Supplementary Fig. S2). That the MSH6 homologous protein in vertebrates instead contains a PWWP that targets gene bodies via H3K36me3 binding suggests a remarkable example of convergent evolution between plants and vertebrates (Li et al. 2013).

Structural modeling prediction of MSH6 structure using AlphaFold (Jumper et al. 2021) indicates that its Tudor domain may, like that in PDS5C, function as an anchor, tethering it to H3K4me1 leading to local increases in DNA repair (Fig. 1E). The crystal structure of PDS5C Tudor domain revealed the sites for H3K4me1 binding (Niu et al. 2021),

comprising aromatic and negatively charged amino acids that form an aromatic cage. Notably, the amino acids constituting the aromatic cage in the PDS5C Tudor domain are conserved in the MSH6 Tudor domain (Fig. 1, B and F) which suggests a similar binding mechanism.

Using fluorescent polarization, we tested experimentally if the Tudor domain from MSH6 binds H3K4me1 preferentially. We found that, as expected from the conserved structure and amino acid conformation of the aromatic cage of the Tudor domain from PDS5C, the Tudor domain of MSH6 binds H3K4me1 with higher affinity compared with all other possible methylated states of lysine 4 in histone 3 (Fig. 1G). Similar to what was previously observed with experimental isothermal calorimetry (ITC) of PDS5C Tudor domain (Niu et al. 2021), we also found that H3K4me2 has some affinity to the domain.

To further test this mechanism, we evaluated the influence of the residues constituting the aromatic cage on the binding affinity of MSH6 Tudor domain by performing fluorescent polarization with a synthesized Tudor domain containing 3 mutations in the key residues of the aromatic cage (W133A, Y140A, Y158A) (Supplementary Fig. S3). As expected, no signal of interaction with any of the tested H3 peptides was found (Fig. 1G), confirming that the aromatic cage, conserved between the MSH6 and PDS5C Tudor domains, is essential for binding to H3K4me1.

De novo mutations in *msh6* knockout lines indicate repair of gene bodies and H3K4me1-marked regions

The preceding results indicated that MSH6 is recruited to H3K4me1 in plants via its Tudor domain. In humans, MSH6 is recruited to H3K36me3 via a different domain, PWWP, with MSH6 deficiencies being a cause of elevated mutation rates in those regions (Li et al. 2013; Frigola et al. 2017; Huang et al. 2018). To experimentally test the effect of MSH6 deficiency in mutation rates in plants, we performed deep sequencing ($>250\times$ depth coverage; Supplementary Table S1) of 3-wk-old rosettes of *A. thaliana* of 7 *msh6* knockout SALK_037557 lines (*msh6/msh6*), 1 *msh6* knockout SALK_089638 line (*msh6/msh6*), 1 heterozygous SALK_089638 line (WT/*msh6*), and 8 wild-type lines (WT/WT). The variant caller Strelka2 was used in TUMOR-NORMAL mode to call SBS mutations for each sample following a stringent filtering process to call true mutations (see Materials and methods; Fig. 2A). This filtering provided a high confidence set of 721 de novo mutations— $<1\%$ of variants called passed all filters (Fig. 2B, Supplementary Data Set 1). As expected, the total number of somatic mutations was higher in the *msh6* knockout lines than in wild-type and heterozygous lines (40.3 vs 14.1 mutations/plant, *t* test, $P < 1 \times 10^{-5}$; Fig. 2C, Supplementary Fig. S4). Mutations from the heterozygous plant (WT/*msh6*) were combined with wild-type samples for further comparisons because it exhibited the wild-type phenotypes (e.g. mutation spectra, rate, and distribution, elaborated below) (Supplementary Fig. S4).

As expected due to mismatches arising from the tendency for oxidized guanine (8-oxoG) to mispair with adenine (Fig. 2D, Supplementary Fig. S4) (Kino et al. 2017), C>A and T>G

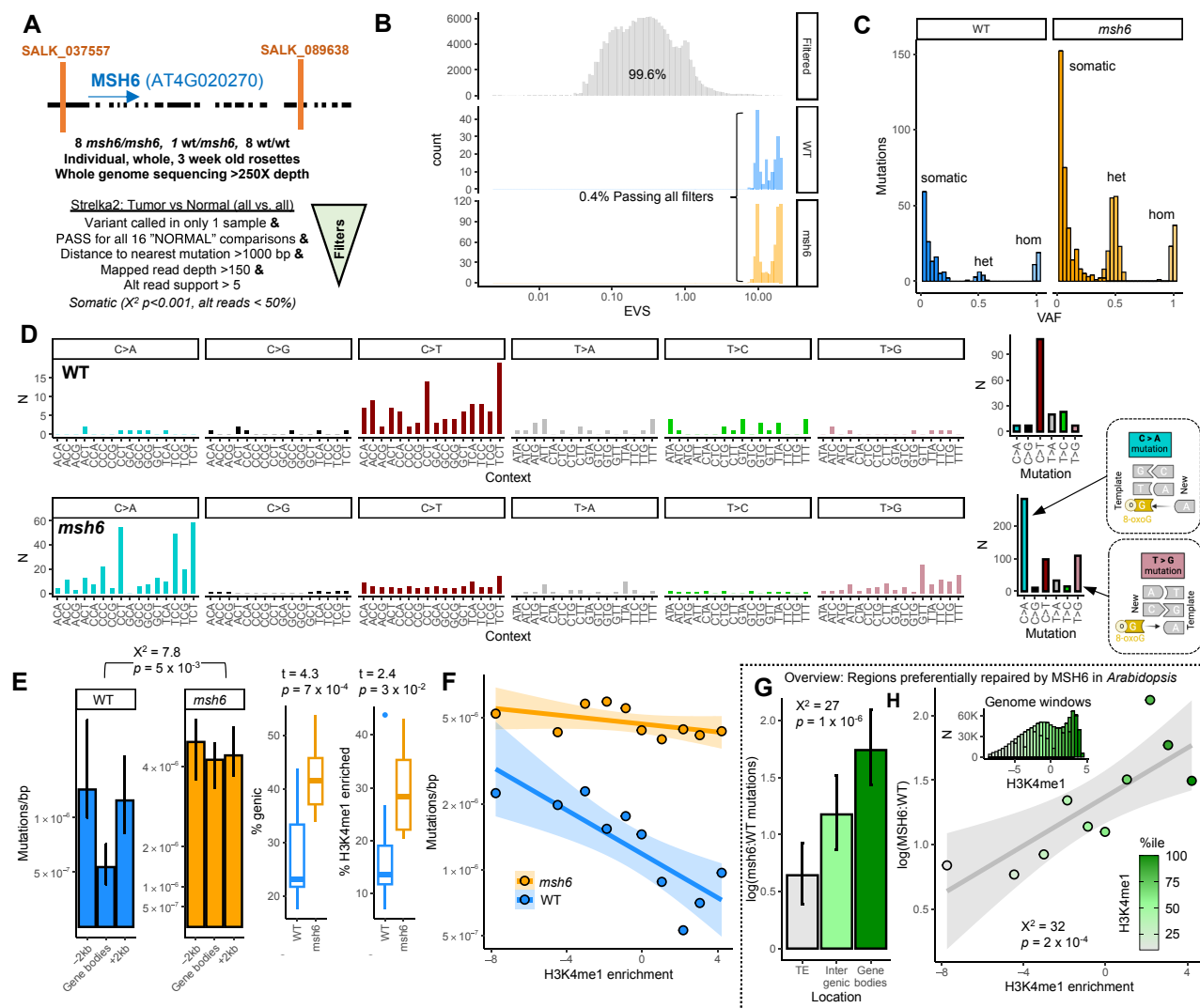


Figure 2. MSH6 contribution to repair of H3K4me1-rich regions in *A. thaliana*. **A**) Diagram representing MSH6 mutant SALK insertions used to phenotype mutation rates in MSH-deficient lines. General pipeline for somatic mutation experimental design and strict variant calling using Strelka2. **B**) Empirical variant scoring (EVS) for mutations called by Strelka2. Greater than 99% of variants were filtered. **C**) Alternative allele frequency of mutations found in *wt/wt* + *wt/msh6* and *msh6/msh6* lines. **D**) SBS mutation signature of *wt/wt* + *wt/msh6* and *msh6/msh6* showing expected effects of unrepaired mismatches between adenine and oxidized guanine. **E**) Left: proportion of mutation distribution in gene bodies in H3K4me1-enriched genes in *wt/wt* + *wt/msh6* and *msh6/msh6* lines. Error bars show 95% CI, χ^2 test genic vs genotype, $n = 591$ mutations. Right: t test comparing percent of mutations in genes and H3K4me1-enriched sites between lines. Boxplots show the median at the center, with box marking the first and third quartiles. Whiskers show 1.5 times the interquartile range. $n = 17$. **F**) mutation rates in relation to H3K4me1 enrichment in genome deciles. **G**) *msh6/msh6* to *wt/wt* + *wt/msh6* mutation ratio in different genomic features with error bars showing 95% bootstrapped CI (χ^2 test, $n = 721$) and **H**) in genome regions grouped by H3K4me1 enrichment (χ^2 test, $n = 721$).

substitutions were proportionally higher in *msh6* knockout lines (Fig. 2C, Supplementary Fig. S4), which is consistent with mutational signatures observed in other MMR-deficient organisms (Lujan et al. 2014; Sanders et al. 2021) and the affinity of MutS α to 8-oxoG:A mispairings (Mazurek et al. 2002). Considering the findings that MSH6 not only binds with MutY and PCNA but also plays a role in facilitating repair processes beyond post-DNA replication stages (Gu et al. 2002; Li 2008; Edelbrock et al. 2013; Hahm et al. 2022) and its synergistic interaction with 8-OXOGUANINE-DNA GLYCOSYLASE 1 (OGG1) (Ni et al. 1999; Pavlov et al. 2003)—a key protein in

8-oxoG:A base excision repair—it is plausible that MSH6 contributes to multiple repair processes. Because MSH6 also influences somatic recombination in *A. thaliana* (Li et al. 2006; Gonzalez and Spampinato 2020), future studies could also explore the potential effects of somatic recombination on local mutation rates.

To test whether MSH6 contributes to the reduction in mutation rates in gene bodies and H3K4me1-marked regions, we explored changes in the proportion of mutations in *msh6* knockout and wild-type lines. The observed hypomutation in gene bodies of wild-type plants was consistent with

what has been previously reported (Monroe, Srikant, et al. 2022). In contrast, in *msh6* knockout mutant lines, mutation rates in H3K4me1-rich gene bodies were significantly more increased than those in gene flanking regions (Fig. 2E, Supplementary Fig. S4). The increase in gene body mutation rates of *msh6* knockout lines was significantly greater than that in intergenic or transposable element (TE) regions (Fig. 2G). Comparing lines directly, the fraction of mutations in genic regions was significantly higher in *msh6* knockout lines (Fig. 2E). These findings support previous work showing that MSH2 (which functions as a dimer with MSH6)-deficient Arabidopsis lacks the gene body hypomutation observed in wild-type plants (Belfield et al. 2018).

We also estimated the mutation rate in relation to H3K4me1 ChIP-seq enrichment and found that *msh6* knockout lines had the greatest increase in mutation rates in H3K4me1-rich regions (Fig. 2H). Likewise, a comparison among lines confirmed a significant increase in the fraction of mutation at H3K4me1-rich sites in *msh6* knockout lines (Fig. 2E).

These experiments address previously expressed uncertainty about the existence of mutation bias in plants (Monroe, Murray, et al. 2022; Monroe, Srikant, et al. 2022). Wang et al. (2023) argued that observed mutation biases are due to sequencing errors. Here, we control for any effect of sequence error by subjecting wild-type and *msh6* mutant lines to the same sequencing and variant calling protocol. We find a clear mutation rate bias associated with H3K4me1 enrichment in wild-type plants, and such bias is diminished in *msh6* mutant lines (Fig. 2). Majic and Payne (2023) proposed that mutation biases were due to developmental selection. Here, we grew wild-type and *msh6* knockout lines in identical selective experimental conditions, with the results not supporting the hypothesis that observed mutation biases (genic hypomutation) are due to selection rather than DNA repair.

Using a strict variant calling pipeline and with the gene body mutation rate being especially low in wild-type plants, we lacked statistical power to test differences between classes of genes (e.g. essential vs nonessential) because 0 mutations were observed for whole gene classes in wild-type plants. Future work may benefit from increasing overall rates of mutation with mutagens (e.g. fast neutron irradiation) or somatic mutation screening methods that can confidently call single-molecule somatic mutations, for greater power to study targeted repair between genes in wild-type and knockout lines.

These experimental results mirror findings from MSH6 deficiency in humans and are consistent with H3K4me1-recruited MSH6 DNA repair in Arabidopsis, with its Tudor domain providing a plausible mechanistic basis. This hypothesis is also supported by the observation of PDS5C, which shares a conserved Tudor domain, physically localized to these regions where mutation rates increased the most (Fig. 1, Supplementary Fig. S1). It should be noted that these findings are not mutually exclusive of other mechanisms and proteins involved in targeted DNA repair, including transcription-coupled nucleotide excision repair, whose relationship with H3K4me1 (in plants) and MSH (in humans) has been

previously described, motivating further investigation (Mellon et al. 1996; Guillemette et al. 2014; Huang et al. 2018; Huang and Li 2018; Oztas et al. 2018; Georgakopoulos-Soares et al. 2020; Kaya et al. 2022). These results inspire future functional studies of this and other potential mechanisms influencing targeted DNA repair and emergent mutation biases.

Conservation of Tudor-binding domains in rice repair proteins

We expanded our research to assess the conservation of these mechanisms in plants. We identified putative orthologs of MSH6 and PDS5C proteins in rice and compared their sequences and structures. Like in Arabidopsis, both proteins contain Tudor domains separated from the functional domain by disordered segments (Fig. S2, B and C). The PDS5C and MSH6 Tudor domains in *O. sativa* share the critical amino acids constituting the aromatic cage, where H3K4me1 binding specificity is determined. This conservation over ~400 Myr of divergence (Wolfe et al. 1989) (Fig. 3A) is expected if selection acts to maintain H3K4me1 recruitment for repair.

Comparative analysis revealed a high degree of structural similarity between the modeled Tudor domains of *A. thaliana* and *O. sativa*, with backbone RMSD values below 1 Å (Supplementary Fig. S2). Furthermore, in silico docking of all methylation states of H3K4 to these Tudor domains showed that the monomethylated state appeared bound to the aromatic cage with the largest frequency (Fig. 3, B to D, Supplementary Fig. S5; see Materials and methods). Notably, these same in silico models accurately predict in vitro binding affinities (ITC for PDS5C from Niu et al. (2021) or FP for MSH6 Tudor in Arabidopsis, here). These results indicate rice Tudor domains from PDS5C and MSH6 have similar binding specificity for H3K4me1 as those experimentally established for Arabidopsis. Altogether, these findings suggest a conserved structural mechanism in rice for H3K4me1 targeting of MSH6 and PDS5C DNA repair.

Neutrality of SBS mutations in fast neutron-irradiated rice

To further investigate the potential for conservation of H3K4me1-recruited DNA repair in rice, we examined mutations generated from fast neutron mutagenesis in Kitaake rice (Li et al. 2017) (Fig. 4A, Supplementary Fig. S6 and Data Set 2). Of 10,000 mutagenized lines, only 7,333 were fertile, with 1,504 ultimately being sequenced for mutation analyses (Li et al. 2017). Structural variants and InDel mutations causing loss-of-function in essential genes indicated selection. To preclude selection on such variants from influencing analyses, we focused on SBS mutations (Li et al. 2017). After masking genomic regions without perfect mappability and removing mutations found in multiple lines, we retained 36,401 de novo SBS mutations for further study (Materials and methods).

A critical test prior to analyzing SBS distributions as a measure of intragenomic mutation rate heterogeneity is to first test whether they exhibit evidence of being affected by selection,

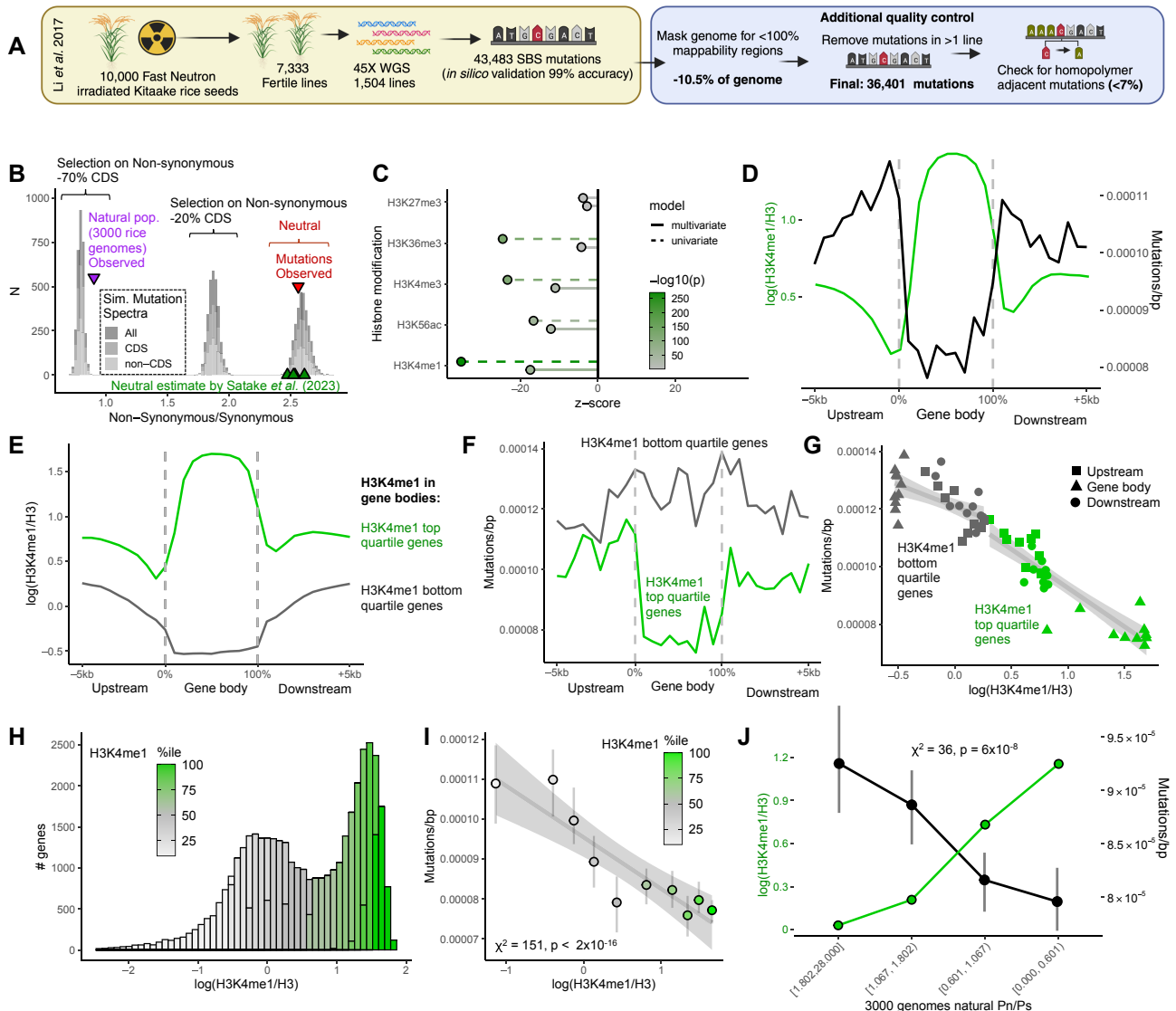


Figure 4. H3K4me1-associated hypomutation in rice. **A**) Experimental summary of Li et al. (2017) and additional quality control. **B**) Expected ratio of nonsynonymous to synonymous mutations as a function of mutation spectra and varying strengths of selection. Arrows highlight the neutral estimate by Satake et al. (2023), the observed value in the mutation data of rice, and ratio observed in natural populations. **C**) Association of histone modifications with mutation frequency from binomial regressions accounting for H3 and input controls. **D**) H3K4me1 distribution and observed mutation rates across gene regions (99.9th percentile shown). **E**) H3K4me1, **F**) mutation rates, and **G**) their relationship for genes in top and bottom quartiles of gene body H3K4me1 enrichment. **H**) Distribution of H3K4me1 enrichment among genes. **I**) Mutation rates in genes grouped by H3K4me1 deciles. **J**) Association of H3K4me1 and mutation rates with natural polymorphisms. Error bars show 95% bootstrap CI in **I**) and **J**) (χ^2 test, $n = 36,401$).

necessitated a loss of 60% of nonsynonymous mutations, which would have been readily detected as a deviation from neutrality (Fig. 4B). While silent sites could, in principle, also be selected against, in light of the neutrality of CDS mutations (Fig. 4B), it would require that silent sites are under substantially stronger selection than those impacting amino acid sequence. The hypomutation observed in rice is consistent with mutation rate heterogeneity in wild-type *Arabidopsis* but not DNA repair mutants (Fig. 2) (Belfield et al. 2018).

We then compared genes ranked by H3K4me1 enrichment levels (Fig. 4H). Genes with high levels of H3K4me1 experienced significantly lower mutation rates, even when we considered

only synonymous mutations and other non-CDSs within gene bodies (Fig. 4J, Supplementary Fig. S8). This result can also not be explained by mappability or HP occurrence. HPs are more frequent in genes with high H3K4me1 (Supplementary Fig. S7), which is the opposite to the hypothesis that HP sequencing errors are responsible for genome regions where low mutation rates are observed (Wang et al. 2023). Likewise, the nonsynonymous/synonymous mutation ratio in H3K4me1-rich genes was not lower, in contrast to what would be expected if the observed mutation reduction is the product of developmental selection (Majic and Payne 2023) (Supplementary Fig. S8).

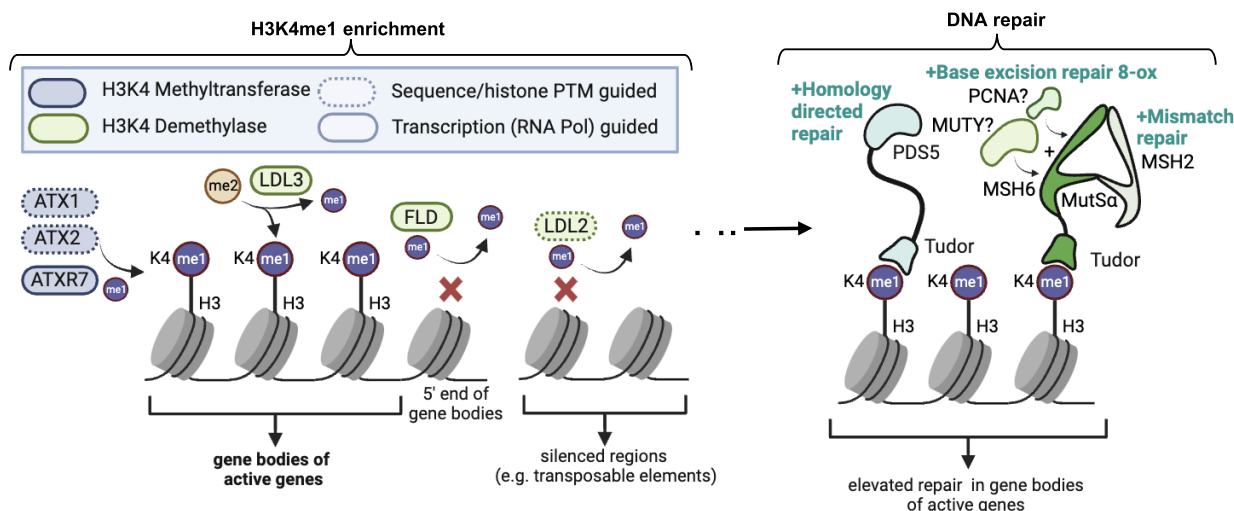


Figure 5. Emerging 2-part mechanistic model of H3K4me1-recruited DNA repair in plants. H3K4me1 enrichment of active gene bodies is achieved by the concerted action of multiple proteins including ATX1, ATX2, ATXR7, LDL3, FLD, and LDL2 (Inagaki et al. 2017; Inagaki et al. 2021; Oya et al. 2022; Mori et al. 2023). PDS5 and MSH6 proteins both contain Tudor domains that then bind H3K4me1. While PDS5 mutations are hypersensitive to DNA double-strand breaks (Pradillo et al. 2015), their consequence on repair at sequence level remains an area for future research. We confirm that MSH6 preferentially repairs H3K4me1-enriched regions and also find that it likely repairs ox-G damage, which could be explained by interactions with other DNA repair systems as has been previously reported (e.g. Edelbrock et al. 2013). Created with BioRender.com.

Because theory predicts mutation modifier systems that promote repair in regions where mutations are more likely to be deleterious, a critical test is whether mutation rates are lowered in regions subject to purifying selection in natural populations. We found that genes under stronger purifying selection in natural populations of rice (estimated by nonsynonymous polymorphism [Pn]/synonymous polymorphism [Ps]) have significantly higher levels of H3K4me1 (Fig. 4I), as seen in *A. thaliana* where H3K4me1 and PDS5C localization is enriched in genes under stronger selection (Supplementary Fig. S1). Correspondingly, we observed these genes subject to greater constraint have lower mutation rates (Fig. 4I). This reduction in mutation rate was reflected in noncoding gene body mutations as well (Supplementary Fig. S8). We did not see a lower nonsynonymous/synonymous mutation ratio in low Pn/Ps genes (Supplementary Fig. S8). As in preceding analyses, this result can also not be explained by HP occurrence; HPs are more common in genes with low Pn/Ps—which is the opposite of the alternative (sequencing error) theory proposed by Wang et al. (2023) regarding mutation biases in Arabidopsis (Monroe, Srikant, et al. 2022; Monroe et al. 2023) (Supplementary Fig. S7).

Thus, both when comparing around genes and between genes, we find that H3K4me1 enrichment is associated with a reduction in mutation rate. This is consistent with expectations resulting from H3K4me1-recruited repair mechanisms in plants (Figs. 1 to 3). Nevertheless, it is also important to note that these patterns observed in rice mutation data sets do not definitively demonstrate causality, requiring further functional work like the genetic experiments in Arabidopsis (Fig. 2).

Mutation biases in rice are consistent with mechanisms of H3K4me1-recruited DNA repair and match predictions from

evolutionary theory on mutation modifier systems. These theories predict that epigenome-recruited repair will evolve to reduce mutation rates in regions of the genome where mutations are most likely to be deleterious. While the drift barrier precludes the evolution of systems that target individual genes, by linking DNA repair with epigenome modifications, whole classes of genes and regions (e.g. bodies of essential genes) can be simultaneously targeted for elevated repair to reduce mutation rates (Lynch 2010; Martincorena and Luscombe 2013; Lynch et al. 2016).

Conclusions

We found evidence of mutation rate variation associated with H3K4me1-mediated DNA repair in *A. thaliana* and *O. sativa* and examined mechanisms conserved in plants (Fig. 5). Functional characterization of PDS5 proteins is still required to determine their influence on lower mutation rates in H3K4me1-associated regions. Because PDS5 proteins are involved in HDR of double-strand break damage, such future experiments should look for variation in InDel mutations. There are also 5 paralogs of PDS5 in *A. thaliana*, potentially necessitating stacked knockouts to overcome genetic redundancies. Genetic evidence confirms that MSH6 is important for reduced SBS mutation rate in gene bodies and H3K4me1-marked regions in *A. thaliana* (Fig. 2). These mechanisms shed light on a compelling case of convergent evolution, especially in MSH6, which in vertebrates is established to target gene bodies by H3K36me3 recruitment of MSH6 through its PWWP domain. Here, we find in plants that MSH6 targets gene bodies via its H3K4me1-binding Tudor domain that is distinct in sequence and structure from the PWWP domain in human

MSH6 (Supplementary Fig. S9). In both cases, the result is lower mutation rates in genic regions (Supek and Lehner 2015, 2017; Frigola et al. 2017; Huang et al. 2018) but appears to be achieved by different histone reader domains and recruitment by histone modifications. Such mechanisms are precisely aligned with evolutionary theories regarding expected evolution of mutation modifiers.

These findings shed light on a plant-specific and higher-resolution mechanistic model of epigenome-recruited DNA repair, with potential implications for evolution, breeding, and gene editing efforts in plants. This motivates investigations to further elucidate the extent, evolutionary origins, and consequences of epigenome-targeted DNA repair in plants and across the eukaryotic tree of life.

Materials and methods

Analysis of ChIP-seq data of AtPDS5C

To study the distribution of PDS5C and H3K4me1 in *Arabidopsis* (*A. thaliana*), we analyzed ChIP-seq data reported by Niu et al. (2021). Enrichment was calculated among regions as $\log_2[(1 + n_{\text{ChIP}})/N_{\text{ChIP}}] - \log_2[(1 + n_{\text{Input}})/N_{\text{Input}}]$, where n_{ChIP} and n_{Input} represent the total depth of mapped ChIP and input fragments in a region and N_{ChIP} and N_{Input} are the total depths of mapped unique fragments. We calculated PDS5C and H3K4me1 enrichment in gene bodies (TSS to TTS) across the TAIR10 *A. thaliana* genome (arabidopsis.org).

Relationship between AtPDS5C and functional constraint

We analyzed the enrichment of the PDS5C ChIP-seq peaks in *A. thaliana* in genetic features and estimated the relationships between those regions and mutation rates, H4K4me1, Pn/Ps, and tissue expression depth. Tissue expression data are from Mergner et al. (2020). Ps and Pn data are from the 1001 Genomes project (1001 Genomes Consortium 2016). Essential genes were defined on findings from Lloyd and Meinke (2012) and Lloyd et al. (2015). PDS5C enrichment was compared with germline mutation rates from *A. thaliana* mutation accumulation lines (Monroe, Srikant, et al. 2022; Weng et al. 2019).

Blastp of Tudor domain and protein structure prediction and visualization

We used *blastp* on Phytozome (Goodstein et al. 2012) to search the rice (*O. sativa*) proteome for PDS5C and MSH6 putative orthologs and to search the *A. thaliana* proteome for genes containing Tudor domains similar to that of PDS5C, which was validated experimentally to bind H3K4me1 (Niu et al. 2021). We submitted the resulting list of 29 genes with putative Tudor domains to gene ontology analysis with ShinyGO (bioinformatics.sdstate.edu/go/) (Ge et al. 2019).

Protein structure predictions were performed using AlphaFold2 (Jumper et al. 2021) in Google Colab (Mirdita et al. 2022) in no-template mode. All structures were

visualized, processed, and analyzed using UCSF ChimeraX (Goddard et al. 2018).

Fluorescence polarization assays of *A. thaliana* MSH6 Tudor domain

Peptide synthesis

All peptides (ARTK[me0, 1, 2, or 3]QTARKSTGGKAPRK[Alloc Q]) were synthesized via a standard Fmoc-SPPS protocol on Rink Amide-PEG Resin (Watanabe Chemical, 0.21 mmol/g, 25 μ mol scale for each peptides). Peptide chain elongation was performed by using automated synthesizers (Liberty Blue [CEM] or ResPep CF [Intavis]) except for methylated lysine residues. Reaction conditions are shown below.

For coupling, Liberty Blue: Fmoc-protected amino acids (4.0 eq.), N,N'-diisopropylcarbodiimide (DIC; 8.0 eq.), and ethyl (hydroxyimino)cyanoacetate (OxymaPure; 4.0 eq.) in N, N-dimethylformamide (DMF), 90 °C for 2 min. ResPep CF: Fmoc-protected amino acids (5.2 eq.), O-(1H-benzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HBTU; 5.0 eq.), and N,N'-diisopropylethylamine (DIEA; 10 eq.) in DMF, room temperature for 20 min twice. For Fmoc deprotection, Liberty Blue: 20% piperidine in DMF, 90 °C for 1 min. ResPep CF: 20% piperidine and 0.1 M 1-hydroxybenzotriazole (HOBt) in DMF, room temperature twice (5 and 6 min).

Fmoc-K(me1 • Boc)-OH, Fmoc-K(me2)-OH • HCl, and Fmoc-K(me3 • Cl)-OH (3.0 eq.) were manually coupled with 1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU; 2.9 eq.) and DIEA (6 eq. for Fmoc-K(me1 • Boc)-OH and Fmoc-K(me3 • Cl)-OH, 9 eq. for Fmoc-K(me2)-OH • HCl).

After elongation of the peptide chain, tetrakis(triphenylphosphine)palladium (0) (0.35 eq.) in degassed CH₂Cl₂ and phenylsilane (20 eq.) in degassed CH₂Cl₂ were added to the resin in order under Ar atmosphere for 30 min to remove the Alloc group. Then, fluorescein isothiocyanate isomer-I (FITC; 1.5 eq.) dissolved in DMF and DIEA (3.0 eq.) was added to the resin, and the mixture was stirred for 3 h at room temperature. A cleavage cocktail (92.5% [v/v] trifluoroacetic acid [TFA], 2.5% [v/v] triisopropylsilane [TIPS], and 5.0% [w/v] 1,3-dimethoxybenzene [DMB]) was added to the reaction column and shaken at room temperature for 3 h. Then, the solution was filtered to remove the resin, and 10 times the volume of cold diethyl ether was added to the filtered solution to precipitate the peptide. This suspension was vortexed and centrifuged at 10,000 × g at 3 °C for 5 min. Ether was removed by decantation. The precipitate was washed with diethyl ether twice. The crude peptide was dissolved in a water/acetonitrile mixture containing 0.1% (v/v) TFA, purified by HPLC (JASCO), and identified by MALDI-TOF mass spectrometry (AutoFlex Max [Bruker]). After lyophilization, peptides were obtained as orange solids (H3K4me0: 31.6 mg, 9.7 μ mol, 38%, H3K4me1: 12.7 mg, 3.9 μ mol, 16%, H3me2: 12.9 mg, 3.9 μ mol, 16%, H3K4me3: 26.9 mg, 8.2 μ mol, 33%). The isolated yields of each peptide were

estimated using the molecular weights of TFA salt at the N-terminal amine and side chains of Arg, Lys, and methylated Lys residues.

Expression system

To assess the binding affinity of the *A. thaliana* MSH6 (AtMSH6) Tudor domain (corresponds to aa 119 to 184 of MSH6) to histone H3 peptides with different methylation states at lysine 4 (H3K4me0, H3K4me1, H3K4me2, and H3K4me3), fluorescence polarization (FP) assays were conducted. The Tudor domain protein and its mutant version (AtMSH6 Tudor 3Mut), with aromatic residues involved in H3K4me1 binding replaced by alanine, were expressed in *E. coli* BL21 (DE3) cells containing minor tRNA expression vector (Codon [+]) RIL, Stratagene) with an N-terminal His tag and MBP tag for purification.

Purification

Cell lysates were prepared by sonication and clarified by centrifugation. The supernatant was subjected to affinity purification using a Ni-NTA agarose column, followed by an anion exchange chromatography on a RESOURCE Q column. Flow-through fractions were collected and further purified by gel filtration using a Superdex 200 column (Cytiva) in final buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% [v/v] glycerol, and 2 mM 2-mercaptoethanol). The purified proteins were then concentrated using an Amicon Ultra 30k device.

FP

The purified proteins, both wild-type and mutant, were incubated with FITC-labeled H3 peptides (20 nM) at various concentrations ranging from 1×10^{-8} to 1×10^{-3} M (0.033, 0.066, 0.132, 0.264, 0.527, 1.055, 2.109, 4.219, 8.438, 16.88, 33.75, 67.5, 135, 270, and 540 μ M) in a total volume of 20 μ L reaction solution, containing 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% [v/v] glycerol, and 2 mM 2-mercaptoethanol. The FP was measured to determine the binding affinity using Synergy H1M2F (BioTek), with the changes in FP indicating the degree of binding between the protein and the peptide.

Mutations in wild-type and *msh6* knockout lines of *A. thaliana*

Plant material

To study the effect of MSH6 in H3K4me1-recruited repair, we performed >250 $\times$ coverage sequencing analyses of 17 *A. thaliana* lines: 8 WT/WT lines, 8 *msh6/msh6* lines (7 SALK_037557 and 1 SALK_089638), and 1 *msh6*/WT line (1 heterozygous SALK_089638). Growth conditions were identical for mutant and wild-type plants. Plants were randomized and grown in standardized potting mix suited for Arabidopsis. The growth chamber conditions were maintained at a 16 h light/8 h dark cycle at a temperature of 25 $^{\circ}$ C. In brief, 3-wk-old plants were harvested, stems and roots were removed, and rosettes were stored at -80° C. Whole rosettes were ground using a

mortar and liquid nitrogen and Qiagen DNeasy Plant Mini Kit (Cat No: 69106) following the manufacturer's instructions. Illumina library preparation and whole-genome DNA sequencing with NovaSeq (paired-end 150 bp read length at high genomic coverage > 250 $\times$) was performed by the UC Davis Genome Center.

Variant calling

Reads were trimmed with trimmomatic (Bolger et al. 2014), and duplicates were marked with the samtools markup function (Li et al. 2009). Reads were mapped to the *A. thaliana* TAIR10 reference genome with bwa mem (Li and Durbin 2009). Strelka variant caller in TUMOR-NORMAL mode (each sample was compared with all other 16 samples as NORMALs) was used to call SBS in the sequenced samples (Kim et al. 2018).

Variant filtering

Variants were filtered by keeping those called in only 1 sample, given a PASS quality score in all 16 "NORMAL" comparisons. In addition to the built-in filters imposed by Strelka2 for variants to receive a PASS, which "include (1) the genotype probability computed by the core variant probability model, (2) root mean square mapping quality, (3) strand bias, (4) the fraction of reads consistent with locus haplotype model, and (5) the complexity of the reference context as measured by metrics such as HP length and compressibility" (Kim et al. 2018), we further filtered to only include variants in which the distance to the nearest mutation was >1,000 bp, total read depth > 150, and alt read support > 5.

Somatic vs germline mutations

To distinguish somatic mutations from de novo heterozygous variants, read depths between alternative and reference alleles were compared with a χ^2 test against the expected 50/50 ratio for heterozygous variants. Mutations were considered somatic if the χ^2 test statistic *P*-value was <0.001 (to account for multiple testing) and the alt read depth was <50%. Homozygous de novo germline mutations were identified when χ^2 test statistic *P*-value was <0.001 and the alt read depth was >50%.

Comparative mutation analyses of *msh6* and wild-type plants

Trimer mutation spectra were compared between *msh6* and wild-type plants across all mutations and 1-mer mutation spectra between individual lines. Because the heterozygous SALK_089638 line showed wild-type mutation spectra and mutation rates, we merged them with wild-type plants for downstream analyses. We compared mutation rates around gene bodies in genes grouped by H3K4me1 (greater or less than 0.83 enrichment, the mean H3K4me1 enrichment for all genes) by counting mutations 2 kb upstream and downstream of gene bodies and used 100 bootstrapping iterations of mutation data sets to generate 95% confidence intervals (CIs) of mutation rates in each window. We compared mutation rates among genome regions (200 bp windows) grouped

into deciles by H3K4me1 enrichment. We then compared mutation rates in genic, intergenic, and TE regions of the genome, with a χ^2 test. We also compared mutation rates in H3K4me1 deciles by a χ^2 test. To further evaluate the significance of the difference in mutation between *msh6* and wild-type plants, we performed *t* tests between individual lines for the % genic mutation and % mutation in H3K4me1-rich sites (enrichment > 1).

Conservation of Tudor domains in rice

We compared rice MSH6 and PDS5C Tudor domains with those in *A. thaliana* by aligning the protein sequences with Muscle. We utilized AlphaFold to model the Tudor domains of *A. thaliana* MSH6 and *O. sativa* MSH6 and PDS5C, aiming to explore their structural conservation and binding affinities. These models were compared with the experimentally determined structure of the *A. thaliana* PDS5C Tudor domain (PDB:7DE9) to assess structural similarities, using backbone root mean square deviation (RMSD) values as a quantitative measure.

We then modified the K4me1 mark on the histone H3 tail peptide bound to the PDS5C Tudor domain using ChimeraX, creating variants H3K4, H3K4me2, and H3K4me3. H3 tail peptides were docked to the experimental structure of *A. thaliana* PDS5C Tudor domain and to the models of *A. thaliana* and *O. sativa* MSH6 and *O. sativa* PDS5C Tudor domains. This was achieved using the Rosetta FlexPepDock tool (Raveh et al. 2010) in refinement mode, generating 10,000 docked models for each case. The top 10% of these models, based on Rosetta total score, were selected for further analysis. Subsequent examination involved measuring the average distance between the nitrogen atom in H3K4 and the center of the aromatic rings of the residues constituting the aromatic cage of Tudor domains. This analysis was facilitated by custom scripts developed in PyRosetta (Chaudhury et al. 2010). A threshold distance of 5.5 Å was established to categorize the models into 2 groups: those with the H3K4 side chain positioned inside the aromatic cage and those with it positioned outside (see Supplementary Fig. S5). This methodological approach was benchmarked against experimental binding data of the PDS5C Tudor domain to H3 tail peptides with H3K4 in various methylation states (Niu et al. 2021).

Mutation data set in *O. sativa*

Rice epigenome data and natural variation

ChIP-seq data for H3K4me1, H3K4me3, H3K36me3, H3K56ac, and H3K27me3, along with input and H3 controls, were previously generated and mapped to the *O. sativa* MSU release 7 reference (Ouyang et al. 2007). For each gene, we calculated the ChIP depth of each histone modification along with input and H3 controls. We estimated H3K4me1 enrichment with $\log(\text{H3K4me1}/\text{H3})$ in windows for downstream analyses comparing genes. We assessed nonsynonymous (including stop gained mutations) and synonymous

polymorphisms (Pn and Ps, respectively) from genomes of 3,010 natural accessions (Wang et al. 2018).

Mutation data

Germline de novo mutations in 1,504 fast neutron mutagenesis lines were downloaded from KitBase at kitbase.ucdavis.edu. These were sequenced at 45X depth, and de novo germline mutations were independently called against the *O. sativa* MSU release 7 reference (Ouyang et al. 2007) and validated in silico as previously described (Li et al. 2017). We focused specifically on SBS ($n = 43,483$; Supplementary Data Set 2), which were reported to have a >99% accuracy by Li et al. (2017). Li et al. (2017) also reported that many lines were not fertile along with evidence of selection on structural variants and InDels in their data, further motivating the exclusion of these mutation classes from our analyses.

Additional quality control

Because Li et al. (2017) previously reported fewer variants were called in regions of low-mappability regions due to poor read alignment, we calculated mappability at every site in the genome with GenMap K = 100 and E = 0. We confirmed the previously reported low variant calling in low-mappability regions and subsequently masked regions and mutations with mappability < 100% from all downstream analyses. We also removed any mutation that was identified in more than 1 line, which can be enriched for sequencing errors or shared inherited mutations. This additional filtering left 36,401 SBS mutations for downstream analyses. For comparisons among genes, we excluded all genes with any site with mappability < 100% and genes lacking evidence of expression, with nonfinite Pn/Ps and nonfinite $\log(\text{H3K4me1}/\text{H3})$ ChIP.

Checking for effects of HP bleed errors

These mutations were called high-depth reads and subject to stringent filtering (10 reads and at least 30% supporting each ALT), likely removing sequencing errors (Li et al. 2017). Nevertheless, given the challenges they present (Monroe et al. 2023; Wang et al. 2023), we tested SBS mutations for potential HP bleed errors (HP-adjacent mutations: cases where the ALT is the same base pair as neighboring HP run of >2 bp). We examined the number of HP sequences/bp of each gene and across genic windows. Because we observed that HP sequences were more common outside of gene bodies, presenting a risk of inflating mutation counts, we confirmed that this was not driving the observed mutation variation by examining the proportion of mutations that were HP-adjacent around genes and also removing HP-adjacent mutations and repeating analyses.

Testing for selection on rice SBS mutations

We annotated each SBS mutation in coding regions as being a synonymous or nonsynonymous (including stop gained) mutation based on the effect on the amino acid sequence.

To test for selection, we generated a matrix representing all possible mutations at every position in the CDS of the rice

genome. We then classified each possible mutation as nonsynonymous and synonymous, maintaining codon position. We then converted observed 1-mer mutation spectra into probabilities and projected across possible mutation space in CDSs. We repeated this with different contexts: (i) spectra of mutation from the whole genome, (ii) from coding mutations (CDS) only, and (iii) non-CDS mutations only. From these probabilities, we drew random mutations equal to the number of observed CDS mutations in the rice data set and calculated the nonsynonymous: synonymous ratio. To estimate the expected ratio under different selection regimes, we removed nonsynonymous mutations to achieve 0% (neutral), 20%, and 70% reduction in variants observed in CDS. We repeated this 1,000 times for each mutation spectra to calculate expected mean and 95% CIs.

Mutation rate heterogeneity

We measured mutation rates in 500 bp windows within 5 kb of gene bodies and tested whether histone modifications predict probability of observing a mutation in windows with binomial regression in R, accounting for input and H3 controls. We performed this in a multivariate model with all histone modifications as well as univariate models with each histone modification. We visualized mutation rates and H3K4me1 enrichment around gene bodies (10 windows upstream, downstream, and within gene bodies). We then calculated the H3K4me1 enrichment of genes and repeated this analysis to compare the upper and lower quartiles of genes ranked by H3K4me1. We also split genes into deciles according to H3K4me1 enrichment and compared their mutation rates, as well as genes in quartiles by Pn/Ps. We used 1,000 bootstrapped iterations to calculate 95% CIs for mutation rates in each group of genes. To further confirm that the results of these analyses were not due to selection on nonsynonymous mutations, we repeated analyses with synonymous mutations/CDS bp and UTR + Intron mutations/bp. We then compared the observed nonsynonymous/synonymous mutations in these groups.

Statistical analysis

Statistical data are provided in [Supplementary Data Set 3](#).

Accession numbers

Sequence data from this article can be found in the NCBI Bioproject PRJNA1039302.

Acknowledgments

The research was conducted at the University of California Davis, which is located on the land which was the home of the Patwin people for thousands of years. We thank Tetsuji Kakutani and Soichi Inagaki for their feedback and insights on H3K4me1.

Author contributions

D.Q., S.O., D.L.-M., and J.G.M. designed the research. D.Q., S.O., D.L.-M., K.Z., L.O., A.A., S.S., G.H., A.O., and J.G.M. performed the research. D.Q., S.O., D.L.-M., K.Z., A.P., S.S., G.H., A.O., V.Y.-Y., P.C.-B., and J.G.M. contributed new analytic/computational/etc. tools. D.Q., S.O., D.L.-M., K.Z., S.S., G.H., A.O., V.Y.-Y., A.P., P.C.-B., and J.G.M. analyzed the data. D.Q., S.O., D.L.-M., K.Z., V.Y.-Y., P.C.-B., S.S., G.H., A.O., and J.G.M. wrote the paper.

Supplementary data

The following materials are available in the online version of this article.

Supplementary Figure S1. PDS5C targets H3K4me1 and localizes to regions with lower mutation rates.

Supplementary Figure S2. Multiple repair pathways with H3K4me1 targeting potential via Tudor domains.

Supplementary Figure S3. Experimental evaluation of *Arabidopsis* MSH6 Tudor domain affinity to H3K4 methylation.

Supplementary Figure S4. Supplementary analyses of de novo mutations in wild-type and *msh6* knockout *A. thaliana*.

Supplementary Figure S5. The average distance (in Å) distribution to the aromatic cage in FlexPepDock models of the different H3K4 marks tested.

Supplementary Figure S6. SBS trimer mutation spectrum in rice data set.

Supplementary Figure S7. Additional quality control of rice mutation data set.

Supplementary Figure S8. Additional analyses of rice mutations among genes.

Supplementary Figure S9. Comparison of *Arabidopsis* Tudor and human PWWP domains in MSH6 homologs.

Supplementary Table S1. Sequencing summary results from *A. thaliana* wild-type and *msh6* knockout lines.

Supplementary Data Set 1. Mutations identified in *A. thaliana* wild-type and *msh6* knockout lines.

Supplementary Data Set 2. Filtered mutations from Kitaake fast neutron-treated rice.

Supplementary Data Set 3. Statistical data.

Funding

The Monroe Lab is supported by Foundation for Food and Agriculture Research grant ICRC20-0000000014, United States Department of Agriculture National Institute of Food and Agriculture grant 108681-Z5327202, UC Davis Science Translation and Innovative Research grant, National Science Foundation grant 21-1-1112-2317191, and CPRB grant HG-2022-35. This research was partially supported by JSPS KAKENHI grant numbers JP22H05172 and JP22H05178 (to A.O.), Research Support Project for Life Science and Drug Discovery (Basis for Supporting Innovative Drug Discovery and Life Science Research [BINDS]) from Japan Agency for

Medical Research and Development under grant number JP22ama121009 (to G.H.), and Japanese Science and Technology Agency Precursory Research for Embryonic Science and Technology grant numbers JPMJPR20K3 (to A.O.) and JPMJPR19K6 (to G.H.).

Conflict of interest statement. None declared.

Data availability

Raw sequencing reads from wild-type and *msh6* knockout lines are available on NCBI BioProject PRJNA1039302. H3K4me1 and PDS5C ChIP-seq data are available from [Niu et al. \(2021\)](#) at NCBI GEO Series accession number GSE154302. The crystal structure of *A. thaliana* PDS5C bound to H3K4me1 is available at RCSB PDB (<https://www.rcsb.org/structure/7DE9>). Germline in *A. thaliana* mutations is available from [Weng et al. \(2019\)](#), [Monroe, Murray, et al. \(2022\)](#), and [Monroe, Srikant, et al. \(2022\)](#). Rice mutations are available at kitbase.ucdavis.edu. Rice ChIP-seq data from [Lu et al. \(2019\)](#) are available on NCBI GEO accession number GSE128434. Rice polymorphism data are available at <https://snp-seek.irri.org/>. The rice reference genome can be accessed at <https://phytozome-next.jgi.doe.gov/>. The *A. thaliana* TAIR10 genome can be accessed at Arabidopsis.org. Code for this project (including for previous versions of this manuscript), are maintained on: https://github.com/greymonroe/Quiroz_H3K4me1_mediated_repair.

References

- Adé J, Belzile F, Philippe H, Doutriaux MP.** Four mismatch repair paralogs coexist in *Arabidopsis thaliana*: AtMSH2, AtMSH3, AtMSH6-1 and AtMSH6-2. *Mol Gen Genet.* 1999;**262**(2):239–249. <https://doi.org/10.1007/pl00008640>
- Akdemir KC, Le VT, Kim JM, Killcoyne S, King DA, Lin Y-P, Tian Y, Inoue A, Amin SB, Robinson FS, et al.** Somatic mutation distributions in cancer genomes vary with three-dimensional chromatin structure. *Nat Genet.* 2020;**52**(11):1178–1188. <https://doi.org/10.1038/s41588-020-0708-0>
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P.** Molecular biology of the cell. New York: Garland Science; 2002.
- Aymard F, Bugler B, Schmidt CK, Guillou E, Caron P, Briois S, Iacovoni JS, Daburon V, Miller KM, Jackson SP, et al.** Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol.* 2014;**21**(4):366–374. <https://doi.org/10.1038/nsmb.2796>
- Belfield EJ, Brown C, Ding ZJ, Chapman L, Luo M, Hinde E, van Es SW, Johnson S, Ning Y, Zheng SJ, et al.** Thermal stress accelerates *Arabidopsis thaliana* mutation rate. *Genome Res.* 2021;**31**(1):40–50. <https://doi.org/10.1101/gr.259853.119>
- Belfield EJ, Ding ZJ, Jamieson FJC, Visscher AM, Zheng SJ, Mithani A, Harberd NP.** DNA mismatch repair preferentially protects genes from mutation. *Genome Res.* 2018;**28**(1):66–74. <https://doi.org/10.1101/gr.219303.116>
- Bolaños-Villegas P, Yang X, Wang H-J, Juan C-T, Chuang M-H, Makaroff CA, Jauh G-Y.** Arabidopsis CHROMOSOME TRANSMISSION FIDELITY 7 (AtCTF7/ECO1) is required for DNA repair, mitosis and meiosis. *Plant J.* 2013;**75**(6):927–940. <https://doi.org/10.1111/tpj.12261>
- Bolger AM, Lohse M, Usadel B.** Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;**30**(15):2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Charlesworth B, Jensen JD.** Population genetic considerations regarding evidence for biased mutation rates in *Arabidopsis thaliana*. *Mol Biol Evol.* 2022;**40**(2):msac275. <https://doi.org/10.1093/molbev/msac275>
- Chaudhury S, Lyskov S, Gray JJ.** Pyrosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics.* 2010;**26**(5):689–691. <https://doi.org/10.1093/bioinformatics/btq007>
- Chen E, Lin-Shiao E, Trinidad M, Doost MS, Colognori D, Doudna JA.** Decorating chromatin for enhanced genome editing using CRISPR-Cas9. *Proc Natl Acad Sci U S A.* 2022;**119**(49):e2204259119. <https://doi.org/10.1073/pnas.2204259119>
- Couturier AM, Fleury H, Patenaude A-M, Bentley VL, Rodrigue A, Coulombe Y, Niraj J, Pauty J, Berman JN, Dellaire G, et al.** Roles for APRIN (PDS5B) in homologous recombination and in ovarian cancer prediction. *Nucleic Acids Res.* 2016;**44**(22):10879–10897. <https://doi.org/10.1093/nar/gkw921>
- de la Peña, MV, Summanen, PAM, Liukkonen, M., Kronholm, I.** Chromatin structure influences rate and spectrum of spontaneous mutations in *Neurospora crassa*. *bioRxiv* 484164. <https://doi.org/10.1101/2022.03.13.484164>, 21 October 2022, preprint: not peer reviewed.
- Edelbrock MA, Kaliyaperumal S, Williams KJ.** Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities. *Mutat Res.* 2013;**743–744**:53–66. <https://doi.org/10.1016/j.mrfmmm.2012.12.008>
- Fang H, Zhu X, Yang H, Oh J, Barbour JA, Wong JWH.** Deficiency of replication-independent DNA mismatch repair drives a 5-methylcytosine deamination mutational signature in cancer. *Sci Adv.* 2021;**7**(45):eabg4398. <https://doi.org/10.1126/sciadv.abg4398>
- Foster PL, Lee H, Popodi E, Townes JP, Tang H.** Determinants of spontaneous mutation in the bacterium *Escherichia coli* as revealed by whole-genome sequencing. *Proc Natl Acad Sci U S A.* 2015;**112**(44):E5990–E5999. <https://doi.org/10.1073/pnas.1512136112>
- Frigola J, Sabarinathan R, Mularoni L, Muiños F, Gonzalez-Perez A, López-Bigas N.** Reduced mutation rate in exons due to differential mismatch repair. *Nat Genet.* 2017;**49**(12):1684–1692. <https://doi.org/10.1038/ng.3991>
- Ge SX, Jung D, Yao R.** ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics.* 2019;**36**(8):2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>
- Genomes Consortium.** 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell.* 2016;**166**(2):481–491. <https://doi.org/10.1016/j.cell.2016.05.063>
- Georgakopoulos-Soares I, Koh G, Momen SE, Jiricny J, Hemberg M, Nik-Zainal S.** Transcription-coupled repair and mismatch repair contribute towards preserving genome integrity at mononucleotide repeat tracts. *Nat Commun.* 2020;**11**(1):1980. <https://doi.org/10.1038/s41467-020-15901-w>
- Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE.** UCSF chimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* 2018;**27**(1):14–25. <https://doi.org/10.1002/pro.3235>
- Gonzalez V, Spampinato CP.** The mismatch repair protein MSH6 regulates somatic recombination in *Arabidopsis thaliana*. *DNA Repair (Amst).* 2020;**87**:102789. <https://doi.org/10.1016/j.dnarep.2020.102789>
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al.** Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 2012;**40**(D1):D1178–D1186. <https://doi.org/10.1093/nar/gkr944>
- Gu Y, Parker A, Wilson TM, Bai H, Chang D-Y, Lu A-L.** Human MutY homolog, a DNA glycosylase involved in base excision repair, physically and functionally interacts with mismatch repair proteins human MutS homolog 2/human MutS homolog 6*. *J Biol Chem.* 2002;**277**(13):11135–11142. <https://doi.org/10.1074/jbc.M108618200>
- Guillemette S, Branagan A, Peng M, Dhruva A, Schärer OD, Cantor SB.** FANCD1 localization by mismatch repair is vital to maintain genomic integrity after UV irradiation. *Cancer Res.* 2014;**74**(3):932–944. <https://doi.org/10.1158/0008-5472.CAN-13-2474>

- Habig M, Lorrain C, Feurtey A, Komlusi J, Stukenbrock EH. Epigenetic modifications affect the rate of spontaneous mutations in a pathogenic fungus. *Nat Commun*. 2021;**12**(1):5869. <https://doi.org/10.1038/s41467-021-26108-y>
- Hahm JY, Park J, Jang E-S, Chi SW. 8-Oxoguanine: from oxidative damage to epigenetic and epitranscriptional modification. *Exp Mol Med*. 2022;**54**(10):1626–1642. <https://doi.org/10.1038/s12276-022-00822-z>
- Hill VK, Kim J-S, Waldman T. Cohesin mutations in human cancer. *Biochim Biophys Acta*. 2016;**1866**(1):1–11. <https://doi.org/10.1016/j.bbcan.2016.05.002>
- Huang Y, Gu L, Li G-M. H3K36me3-mediated mismatch repair preferentially protects actively transcribed genes from mutation. *J Biol Chem*. 2018;**293**(20):7811–7823. <https://doi.org/10.1074/jbc.RA118.002839>
- Huang Y, Li G-M. DNA mismatch repair preferentially safeguards actively transcribed genes. *DNA Repair (Amst)*. 2018;**71**:82–86. <https://doi.org/10.1016/j.dnarep.2018.08.010>
- Inagaki S, Takahashi M, Hosaka A, Ito T, Toyoda A, Fujiyama A, Tarutani Y, Kakutani T. Gene-body chromatin modification dynamics mediate epigenome differentiation in Arabidopsis. *EMBO J*. 2017;**36**(8):970–980. <https://doi.org/10.15252/embj.201694983>
- Inagaki S, Takahashi M, Takashima K, Oya S, Kakutani T. Chromatin-based mechanisms to coordinate convergent overlapping transcription. *Nat Plants*. 2021;**7**(3):295–302. <https://doi.org/10.1038/s41477-021-00868-3>
- Jin H, Guacci V, Yu H-G. Pds5 is required for homologue pairing and inhibits synapsis of sister chromatids during yeast meiosis. *J Cell Biol*. 2009;**186**(5):713–725. <https://doi.org/10.1083/jcb.200810107>
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;**596**(7873):583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Katju V, Konrad A, Deiss TC, Bergthorsson U. Mutation rate and spectrum in obligately outcrossing *Caenorhabditis elegans* mutation accumulation lines subjected to RNAi-induced knockdown of the mismatch repair gene msh-2. *G3*. 2022;**12**(1):jkab364. <https://doi.org/10.1093/g3journal/jkab364>
- Kaya S, Adebali O, Oztas O, Sancar A. Genome-wide excision repair map of cyclobutane pyrimidine dimers in Arabidopsis and the roles of CSA1 and CSA2 proteins in transcription-coupled repair. *Photochem Photobiol*. 2022;**98**(3):707–712. <https://doi.org/10.1111/php.13519>
- Kenchanmane Raju SK, Lensink M, Kliebenstein DJ, Niederhuth C, Monroe G. Epigenomic divergence correlates with sequence polymorphism in Arabidopsis paralogs. *New Phytol*. 2023;**240**(3):1292–1304. <https://doi.org/10.1111/nph.19227>
- Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L, Zhang Y, Bedford MT. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep*. 2006;**7**(4):397–403. <https://doi.org/10.1038/sj.embor.7400625>
- Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, Chen X, Kim Y, Beyter D, Krusche P, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;**15**(8):591–594. <https://doi.org/10.1038/s41592-018-0051-x>
- Kino K, Hirao-Suzuki M, Morikawa M, Sakaga A, Miyazawa H. Generation, repair and replication of guanine oxidation products. *Genes Environ*. 2017;**39**(1):21. <https://doi.org/10.1186/s41021-017-0081-0>
- Kolodner R. Biochemistry and genetics of eukaryotic mismatch repair. *Genes Dev*. 1996;**10**(12):1433–1442. <https://doi.org/10.1101/gad.10.12.1433>
- Krasovec M, Eyre-Walker A, Sanchez-Ferandin S, Piganeau G. Spontaneous mutation rate in the smallest photosynthetic eukaryotes. *Mol Biol Evol*. 2017;**34**(7):1770–1779. <https://doi.org/10.1093/molbev/msx119>
- Kusch T. Brca2-Pds5 complexes mobilize persistent meiotic recombination sites to the nuclear envelope. *J Cell Sci*. 2015;**128**(4):717–727. <https://doi.org/10.1242/jcs.159988>
- Li G-M. Mechanisms and functions of DNA mismatch repair. *Cell Res*. 2008;**18**(1):85–98. <https://doi.org/10.1038/cr.2007.115>
- Li R, Di L, Li J, Fan W, Liu Y, Guo W, Liu W, Liu L, Li Q, Chen L, et al. A body map of somatic mutagenesis in morphologically normal human tissues. *Nature*. 2021;**597**(7876):398–403. <https://doi.org/10.1038/s41586-021-03836-1>
- Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;**25**(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;**25**(16):2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li G, Jain R, Chern M, Pham NT, Martin JA, Wei T, Schackwitz WS, Lipzen AM, Duong PQ, Jones KC, et al. The sequences of 1504 mutants in the model rice variety Kitaake facilitate rapid functional genomic studies. *Plant Cell*. 2017;**29**(6):1218–1231. <https://doi.org/10.1105/tpc.17.00154>
- Li L, Jean M, Belzile F. The impact of sequence divergence and DNA mismatch repair on homeologous recombination in Arabidopsis. *Plant J*. 2006;**45**(6):908–916. <https://doi.org/10.1111/j.1365-3113.2006.02657.x>
- Li F, Mao G, Tong D, Huang J, Gu L, Yang W, Li G-M. The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. *Cell*. 2013;**153**(3):590–600. <https://doi.org/10.1016/j.cell.2013.03.025>
- Liu Y, Liu K, Yin L, Yu Y, Qi J, Shen W-H, Zhu J, Zhang Y, Dong A. H3K4me2 functions as a repressive epigenetic mark in plants. *Epigenetics Chromatin*. 2019;**12**(1):40. <https://doi.org/10.1186/s13072-019-0285-6>
- Liu H, Zhang J. Is the mutation rate lower in genomic regions of stronger selective constraints? *Mol Biol Evol*. 2022;**39**(8):msac169. <https://doi.org/10.1093/molbev/msac169>
- Lloyd J, Meinke D. A comprehensive dataset of genes with a loss-of-function mutant phenotype in Arabidopsis. *Plant Physiol*. 2012;**158**(3):1115–1129. <https://doi.org/10.1104/pp.111.192393>
- Lloyd JP, Seddon AE, Moghe GD, Simenc MC, Shiu S-H. Characteristics of plant essential genes allow for within- and between-species prediction of lethal mutant phenotypes. *Plant Cell*. 2015;**27**(8):2133–2147. <https://doi.org/10.1105/tpc.15.00051>
- López-Cortegano E, Craig RJ, Chebib J, Samuels T, Morgan AD, Kraemer SA, Böndel KB, Ness RW, Colegrave N, Keightley PD. De novo mutation rate variation and its determinants in Chlamydomonas. *Mol Biol Evol*. 2021;**38**(9):3709–3723. <https://doi.org/10.1093/molbev/msab140>
- Lu Z, Cui J, Wang L, Teng N, Zhang S, Lam H-M, Zhu Y, Xiao S, Ke W, Lin J, et al. Genome-wide DNA mutations in Arabidopsis plants after multigenerational exposure to high temperatures. *Genome Biol*. 2021;**22**(1):160. <https://doi.org/10.1186/s13059-021-02381-4>
- Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat Plants*. 2019;**5**(12):1250–1259. <https://doi.org/10.1038/s41477-019-0548-z>
- Lu R, Wang GG. Tudor: a versatile family of histone methylation “readers”. *Trends Biochem Sci*. 2013;**38**(11):546–555. <https://doi.org/10.1016/j.tibs.2013.08.002>
- Lujan SA, Clausen AR, Clark AB, MacAlpine HK, MacAlpine DM, Malc EP, Mieczkowski PA, Burkholder AB, Fargo DC, Gordenin DA, et al. Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res*. 2014;**24**(11):1751. <https://doi.org/10.1101/gr.178335.114>
- Lynch M. Evolution of the mutation rate. *Trends Genet*. 2010;**26**(8):345–352. <https://doi.org/10.1016/j.tig.2010.05.003>
- Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet*. 2016;**17**(11):704–714. <https://doi.org/10.1038/nrg.2016.104>
- Lynch M, Ali F, Lin T, Wang Y, Ni J, Long H. The divergence of mutation rates and spectra across the Tree of Life. *EMBO Rep*. 2023;**24**(10):e57561. <https://doi.org/10.15252/embr.202357561>

- Majic P, Payne JL.** Developmental selection and the perception of mutation bias. *Mol Biol Evol.* 2023;**40**(8):msad179. <https://doi.org/10.1093/molbev/msad179>
- Makova KD, Hardison RC.** The effects of chromatin organization on variation in mutation rates in the genome. *Nat Rev Genet.* 2015;**16**(4):213–223. <https://doi.org/10.1038/nrg3890>
- Manova V, Gruszka D.** DNA damage and repair in plants—from models to crops. *Front Plant Sci.* 2015;**6**:885. <https://doi.org/10.3389/fpls.2015.00885>
- Martincorena I, Luscombe NM.** Non-random mutation: the evolution of targeted hypermutation and hypomutation. *BioEssays.* 2013;**35**(2):123–130. <https://doi.org/10.1002/bies.201200150>
- Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, Wedge DC, Fullam A, Alexandrov LB, Tubio JM, et al.** Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science.* 2015;**348**(6237):880–886. <https://doi.org/10.1126/science.aaa6806>
- Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP.** The Tudor domain “Royal Family”: Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem Sci.* 2003;**28**(2):69–74. [https://doi.org/10.1016/S0968-0004\(03\)00004-5](https://doi.org/10.1016/S0968-0004(03)00004-5)
- Mazurek A, Berardini M, Fishel R.** Activation of human MutS homologs by 8-oxo-guanine DNA damage. *J Biol Chem.* 2002;**277**(10):8260–8266. <https://doi.org/10.1074/jbc.M111269200>
- Mellon I, Rajpal DK, Koi M, Boland CR, Champe GN.** Transcription-coupled repair deficiency and mutations in human mismatch repair genes. *Science.* 1996;**272**(5261):557–560. <https://doi.org/10.1126/science.272.5261.557>
- Mergner J, Frejno M, List M, Papacek M, Chen X, Chaudhary A, Samaras P, Richter S, Shikata H, Messerer M, et al.** Mass-spectrometry-based draft of the *Arabidopsis* proteome. *Nature.* 2020;**579**(7799):409–414. <https://doi.org/10.1038/s41586-020-2094-2>
- Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M.** ColabFold: making protein folding accessible to all. *Nat Methods.* 2022;**19**(6):679–682. <https://doi.org/10.1038/s41592-022-01488-1>
- Monroe JG, Murray KD, Xian W, Carbonell-Bejerano, P, Fenster, CB, Weigel, D.** Report of mutation biases mirroring selection in *Arabidopsis thaliana* unlikely to be entirely due to variant calling errors. *bioRxiv* 504682. <https://doi.org/10.1101/2022.08.21.504682>, 22 August 2022, preprint: not peer reviewed.
- Monroe JG, Murray KD, Xian W, Srikant T, Carbonell-Bejerano P, Becker C, Lensink M, Exposito-Alonso M, Klein M, Hildebrandt J, et al.** Reply to: re-evaluating evidence for adaptive mutation rate variation. 2023;**619**(7971):E57–E60. <https://doi.org/10.1038/s41586-023-06315-x>
- Monroe JG, Srikant T, Carbonell-Bejerano P, Becker C, Lensink M, Exposito-Alonso M, Klein M, Hildebrandt J, Neumann M, Kliebenstein D, et al.** Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature.* 2022;**602**(7895):101–105. <https://doi.org/10.1038/s41586-021-04269-6>
- Moore L, Cagan A, Coorens THH, Neville MDC, Sanghvi R, Sanders MA, Oliver TRW, Leongamornlert D, Ellis P, Noorani A, et al.** The mutational landscape of human somatic and germline cells. *Nature.* 2021;**597**(7876):381–386. <https://doi.org/10.1038/s41586-021-03822-7>
- Morales C, Ruiz-Torres M, Rodríguez-Acebes S, Lafarga V, Rodríguez-Corsino M, Megías D, Cisneros DA, Peters J-M, Méndez J, Losada A.** PDS5 proteins are required for proper cohesin dynamics and participate in replication fork protection. *J Biol Chem.* 2020;**295**(1):146–157. <https://doi.org/10.1074/jbc.RA119.011099>
- Mori, S, Oya, S, Takahashi, M, Takashima, K, Inagaki, S, Kakutani, T.** Cotranscriptional demethylation induces global loss of H3K4me2 from active genes in *Arabidopsis*. *bioRxiv* 528985. <https://doi.org/10.1101/2023.02.17.528985>, 18 February 2023, preprint: not peer reviewed.
- Ni TT, Marsischky GT, Kolodner RD.** MSH2 and MSH6 are required for removal of adenine misincorporated opposite 8-oxo-guanine in *S. cerevisiae*. *Mol Cell.* 1999;**4**(3):439–444. [https://doi.org/10.1016/S1097-2765\(00\)80346-9](https://doi.org/10.1016/S1097-2765(00)80346-9)
- Nisa M-U, Huang Y, Benhamed M, Raynaud C.** The plant DNA damage response: signaling pathways leading to growth inhibition and putative role in response to stress conditions. *Front Plant Sci.* 2019;**10**:653. <https://doi.org/10.3389/fpls.2019.00653>
- Niu Q, Song Z, Tang K, Chen L, Wang L, Ban T, Guo Z, Kim C, Zhang H, Duan C-G, et al.** A histone H3K4me1-specific binding protein is required for siRNA accumulation and DNA methylation at a subset of loci targeted by RNA-directed DNA methylation. *Nat Commun.* 2021;**12**(1):3367. <https://doi.org/10.1038/s41467-021-23637-4>
- Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, et al.** The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.* 2007;**35**(Database):D883–D887. <https://doi.org/10.1093/nar/gkl976>
- Oya S, Takahashi M, Takashima K, Kakutani T, Inagaki S.** Transcription-coupled and epigenome-encoded mechanisms direct H3K4 methylation. *Nat Commun.* 2022;**13**(1):4521. <https://doi.org/10.1038/s41467-022-32165-8>
- Oztas O, Selby CP, Sancar A, Adebali O.** Genome-wide excision repair in *Arabidopsis* is coupled to transcription and reflects circadian gene expression patterns. *Nat Commun.* 2018;**9**(1):1503. <https://doi.org/10.1038/s41467-018-03922-5>
- Pavlov YI, Mian IM, Kunkel TA.** Evidence for preferential mismatch repair of lagging strand DNA replication errors in yeast. *Curr Biol.* 2003;**13**(9):744–748. [https://doi.org/10.1016/S0960-9822\(03\)00284-7](https://doi.org/10.1016/S0960-9822(03)00284-7)
- Phipps, J, Dubrana, K.** DNA repair in space and time: safeguarding the genome with the cohesin complex. *Genes (Basel).* 2022;**13**(2):198. <https://doi.org/10.3390/genes13020198>
- Pradillo M, Knoll A, Oliver C, Varas J, Corredor E, Puchta H, Santos JL.** Involvement of the cohesin cofactor PDS5 (SPO76) during meiosis and DNA repair in *Arabidopsis thaliana*. *Front Plant Sci.* 2015;**6**:1034. <https://doi.org/10.3389/fpls.2015.01034>
- Raveh B, London N, Schueler-Furman O.** Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins* 2010;**78**(9):2029–2040. <https://doi.org/10.1002/prot.22716>
- Ren Q, Yang H, Rosinski M, Conrad MN, Dresser ME, Guacci V, Zhang Z.** Mutation of the cohesin related gene PDS5 causes cell death with predominant apoptotic features in *Saccharomyces cerevisiae* during early meiosis. *Mutat Res.* 2005;**570**(2):163–173. <https://doi.org/10.1016/j.mrfmmm.2004.11.014>
- Sanders, MA, Vöhringer, H, Forster, VJ, Moore, L, Campbell, BB, Hooks, Y, Edwards, M, Bianchi, V, Coorens, THH, Butler, TM, et al.** Life without mismatch repair. *bioRxiv* 437578. <https://doi.org/10.1101/2021.04.14.437578>, 15 April 2021 preprint: not peer reviewed.
- Satake A, Imai R, Fujino T, Tomimoto S, Ohta K, Na'iem M, Indrioko S, Widiyatno S, Purnomo S, Mollá-Morales A, et al.** Somatic mutation rates scale with time not growth rate in long-lived tropical trees. *eLife.* 2023;**12**:2023–01. <https://doi.org/10.7554/eLife.88456>
- Schep R, Brinkman EK, Leemans C, Vergara X, van der Weide RH, Morris B, van Schaik T, Manzo SG, Peric-Hupkes D, van den Berg J, et al.** Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol Cell.* 2021;**81**(10):2216–2230.e10. <https://doi.org/10.1016/j.molcel.2021.03.032>
- Schubert V, Weissleder A, Ali H, Fuchs J, Lermontova I, Meister A, Schubert I.** Cohesin gene defects may impair sister chromatid alignment and genome stability in *Arabidopsis thaliana*. *Chromosoma* 2009;**118**(5):591–605. <https://doi.org/10.1007/s00412-009-0220-x>
- Schuster-Böckler B, Lehner B.** Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature.* 2012;**488**(7412):504–507. <https://doi.org/10.1038/nature11273>
- Staunton PM, Peters AJ, Seoighe C.** Somatic mutations inferred from RNA-seq data highlight the contribution of replication timing to mutation rate variation in a model plant. *Genetics* 2023;**225**(2):iyad128. <https://doi.org/10.1093/genetics/iyad128>

- Sun Z, Zhang Y, Jia J, Fang Y, Tang Y, Wu H, Fang D. H3k36me3, message from chromatin to DNA damage repair. *Cell Biosci.* 2020;**10**(1):9. <https://doi.org/10.1186/s13578-020-0374-z>
- Supek F, Lehner B. Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature.* 2015;**521**(7550):81–84. <https://doi.org/10.1038/nature14173>
- Supek F, Lehner B. Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. *Cell.* 2017;**170**(3):534–547.e23. <https://doi.org/10.1016/j.cell.2017.07.003>
- Supek F, Lehner B. Scales and mechanisms of somatic mutation rate variation across the human genome. *DNA Repair (Amst).* 2019;**81**: 102647. <https://doi.org/10.1016/j.dnarep.2019.102647>
- Sutherland, CA, Prigozhin, DM, Monroe, JG, Krasileva, KV. High intraspecies allelic diversity in Arabidopsis NLR immune receptors is associated with higher transcription, gene body hypomethylation, and proximity to transposable elements. *bioRxiv* 523861. <https://doi.org/10.1101/2023.01.12.523861>, 13 January 2023, preprint: not peer reviewed.
- Wang L, Ho AT, Hurst LD, Yang S. Re-evaluating evidence for adaptive mutation rate variation [Review of *Re-evaluating evidence for adaptive mutation rate variation*]. *Nature.* 2023;**619**(7971):E52–E56. <https://doi.org/10.1038/s41586-023-06314-y>
- Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature.* 2018;**557**(7703):43–49. <https://doi.org/10.1038/s41586-018-0063-9>
- Wang S-W, Read RL, Norbury CJ. Fission yeast Pds5 is required for accurate chromosome segregation and for survival after DNA damage or metaphase arrest. *J Cell Sci.* 2002;**115**(3):587–598. <https://doi.org/10.1242/jcs.115.3.587>
- Weiss, T, Crisp, PA, Rai, KM, Song, M, Springer, NM, Zhang, F. Drastic differential CRISPR-Cas9 induced mutagenesis influenced by DNA methylation and chromatin features. *bioRxiv* 482333. <https://doi.org/10.1101/2022.02.28.482333>, 1 March 2022, preprint: not peer reviewed.
- Weng M-L, Becker C, Hildebrandt J, Neumann M, Rutter MT, Shaw RG, Weigel D, Fenster CB. Fine-grained analysis of spontaneous mutation spectrum and frequency in *Arabidopsis thaliana*. *Genetics* 2019;**211**(2):703–714. <https://doi.org/10.1534/genetics.118.301721>
- Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH. Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. *Proc Natl Acad Sci U S A.* 1989;**86**(16):6201–6205. <https://doi.org/10.1073/pnas.86.16.6201>
- Xiao C, Fan T, Tian H, Zheng Y, Zhou Z, Li S, Li C, He J. H3k36 trimethylation-mediated biological functions in cancer. *Clin Epigenetics.* 2021;**13**(1):199. <https://doi.org/10.1186/s13148-021-01187-2>
- Yan W, Deng XW, Yang C, Tang X. The genome-wide EMS mutagenesis bias correlates with sequence context and chromatin structure in rice. *Front Plant Sci.* 2021;**12**:579675. <https://doi.org/10.3389/fpls.2021.579675>
- Yang X, Breuss MW, Xu X, Antaki D, James KN, Stanley V, Ball LL, George RD, Wirth SA, Cao B, et al. Developmental and temporal characteristics of clonal sperm mosaicism. *Cell.* 2021;**184**(18):4772–4783.e15. <https://doi.org/10.1016/j.cell.2021.07.024>
- Yousefzadeh M, Henspit C, Vyas R, Soto-Palma C, Robbins P, Niedernhofer L. DNA damage—how and why we age? *eLife* 2021;**10**:e62852. <https://doi.org/10.7554/eLife.62852>
- Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE. Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol.* 2009;**10**(6):R62. <https://doi.org/10.1186/gb-2009-10-6-r62>
- Zhang N, Coutinho LE, Pati D. PDS5A and PDS5B in cohesin function and human disease. *Int J Mol Sci.* 2021;**22**(11):5868. <https://doi.org/10.3390/ijms22115868>
- Zhu X, Xie S, Tang K, Kalia RK, Liu N, Ma J, Bressan RA, Zhu J-K. Non-CG DNA methylation-deficiency mutations enhance mutagenesis rates during salt adaptation in cultured Arabidopsis cells. *Stress Biol.* 2021;**1**(1):12. <https://doi.org/10.1007/s44154-021-00013-2>