

1 Prime editing enables high-throughput identification and characterization of 2 selective MutSβ inhibitors

3 Yangwode Jing¹, Jun Yan^{2,13}, Denny Yang^{3,16}, Yanjin Chen^{4,16}, Issam Senoussi^{5,6,16}, James Woodward³,
4 Katya Williams^{2,14}, Hahn Kim^{1,7,8}, Kangning He³, Herwin Daub⁹, Vinay Dahiya⁹, Maren Thomsen⁹, Jia Ling¹⁰,
5 Purnima Ravisankar^{10,15}, Yuri Pritykin^{4,10}, Gabriel Balmus^{3,11}, Petr Cejka⁵, Jeffrey A. Hussmann¹⁰, Longbin
6 Liu¹², Thomas F. Vogt¹², Michael Finley¹², Britt Adamson^{2,10,17,*}

7 ¹ Department of Chemistry, Princeton University, Princeton, NJ 08544, USA

8 ² Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

9 ³ UK Dementia Research Institute at University of Cambridge and Department of Clinical Neurosciences, University of
10 Cambridge, Cambridge Biomedical Campus, Cambridge CB2 0AH, UK

11 ⁴ Department of Computer Science, Princeton University, Princeton, NJ 08544, USA

12 ⁵ Faculty of Biomedical Sciences, Institute for Research in Biomedicine, Università della Svizzera italiana (USI),
13 Bellinzona 6500, Switzerland

14 ⁶ Department of Biology, Institute of Biochemistry, Eidgenössische Technische Hochschule (ETH), Zürich 8093,
15 Switzerland

16 ⁷ Princeton University Small Molecule Screening Center, Princeton University, Princeton, NJ 08544, USA

17 ⁸ Ludwig Institute for Cancer Research, Princeton University, Princeton, NJ 08544, USA

18 ⁹ Proteros biostructures GmbH, Bunsenstr 7a, Martinsried 82152, Germany

19 ¹⁰ Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA

20 ¹¹ Department of Molecular Neuroscience, Transylvanian Institute of Neuroscience, Cluj-Napoca 400191, Romania

21 ¹² CHDI Management/CHDI Foundation, Princeton, NJ 08540, USA

22 ¹³ Current address: Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA 94158, USA

23 ¹⁴ Current address: Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA

24 ¹⁵ Current address: Immunology and Microbial Pathogenesis Program, Weill Cornell Graduate School of Medical
25 Sciences, New York, NY 10065, USA

26 ¹⁶ These authors contributed equally

27 ¹⁷ Lead contact

28

29 * Correspondence: badamson@princeton.edu (B.A.)

30

31

32 ABSTRACT

33 DNA mismatch repair (MMR) maintains genome stability but paradoxically drives expansion of pathogenic
 34 repeats responsible for trinucleotide repeat disorders. Attenuating repeat instability through targeted
 35 inhibition of MMR is therefore a therapeutic strategy of interest; however, progress is constrained by a lack
 36 of robust methods for defining and isolating MMR subpathways. We exploited features of prime editing to
 37 profile the effects of core MMR genes on thousands of programmed DNA mispairs in cells and resolved
 38 distinct MMR processes, including a noncanonical “loop-directed” mechanism. Guided by these results, we
 39 built a set of scalable reporters that detect loss of separable MMR processes and screened ~72,000 small
 40 molecules for activity against MutSβ-directed MMR. This effort identified a novel series of MutSβ inhibitors
 41 that reduced (CAG)_n expansion in a cell model. Altogether, this study establishes several methods for
 42 systematically interrogating MMR, with demonstrated capacity to identify inhibitors of therapeutic value.

43 INTRODUCTION

44 Huntington's disease (HD) is an incurable, neurodegenerative disorder characterized by progressive motor
45 dysfunction, cognitive decline, and psychiatric disturbances, leading to severe disability and death.¹ HD
46 exhibits a pattern of autosomal dominant inheritance linked to an expanded trinucleotide CAG repeat in
47 exon 1 of the gene *huntingtin* (*HTT*),² placing it within a broader class of disorders caused by unstable
48 repeats in the human genome.³ As defined in a recent position paper,⁴ individuals with ≥ 40 CAG repeats in
49 *HTT* have highly penetrant HD ($>98\%$), with $>60\%$ probability of clinical motor diagnosis by age 65 (at a
50 diagnostic confidence level of 99%), whereas individuals born with more moderately expanded *HTT* alleles
51 (36-39 CAGs) show reduced penetrance and delayed progression of the disease. Of significance,
52 disease-causing CAG tracts undergo additional lengthening in select cell types of people with HD, and
53 somatic expansion of these alleles in striatal medium spiny neurons is thought to contribute to
54 neurodegeneration and disease progression.^{5,6} Halting or slowing somatic lengthening of repeats has thus
55 emerged as a therapeutic strategy for HD and other repeat expansion disorders of unmet medical need.

56 Two complementary lines of evidence have established a tractable path forward for developing trinucleotide
57 repeat-stabilizing interventions: (1) genome-wide association studies have identified genetic modifiers of
58 somatic (CAG)_n expansion and HD clinical phenotypes in and around genes associated with DNA mismatch
59 repair (MMR),⁷⁻¹³ with some variants able to delay disease onset or progression, and (2) disruption of
60 several of those modifier genes has been shown to attenuate (CAG)_n expansion in experimental models of
61 HD, including cultured cells,^{14,15} patient-derived stem cells and differentiated neurons,^{16,17} and mice.¹⁸⁻²²
62 Together, these findings implicate mechanisms of MMR in driving somatic instability of HD-causing alleles
63 and suggest that targeted inhibition of select MMR components could provide therapeutic benefit.

64 Although implicated in somatic repeat instability, the primary role of MMR is to maintain genome stability by
65 correcting base-base mismatches and small loop-out structures that arise during normal cellular processes
66 (e.g., DNA replication and transcription).^{23,24} Exactly how MMR goes awry to expand unstable CAG repeats
67 remains unclear; however, converging lines of evidence implicate MutS β -directed mechanisms. For
68 context, small DNA mispairs can be detected by one of two heterodimeric MMR sensor complexes, MutS α
69 (MSH2-MSH6) or MutS β (MSH2-MSH3). These sensors bind overlapping but distinct classes of DNA
70 mispairs and recruit downstream factors to identify and remove error-containing DNA strands (Figure 1A).
71 Supporting a role for MutS β in somatic repeat expansion, the gene encoding the unique subunit of MutS β
72 (*MSH3*) is a consistent modifier of HD across human genetic studies and promotes (CAG)_n expansion in
73 HD models. Moreover, MutS β binds with high affinity to (CAG)_n loop-out structures,²⁵⁻²⁷ which can arise
74 within tracts of CAG repeats due to strand slippage and may initiate repeat expansion processes. MutS α ,
75 by contrast, has been reported to have a minimal¹⁹ or even opposite role¹⁸ in somatic repeat instability.

76 MutS β thus represents a compelling drug target; however, given the central role of MMR in genome
77 maintenance, concomitant risk of elevated cancer susceptibility is a concern. Reinforcing this concern,
78 inherited mutations in several MMR genes (e.g., *MSH6* and *MSH2*) predispose individuals to cancer^{28,29}
79 and mutations in *MSH3* have shown associations with cancer and adenomatous polyposis.^{30–32} Evidence
80 supporting a causal role for MutS β dysfunction in development of cancer though is lacking, and multiple
81 studies have shown that loss of *MSH3* does not elevate tumor incidence or affect survival in mice, unlike
82 disruption of other MMR genes.^{33,34} Additionally, knocking down *MSH3* with antisense oligonucleotides or
83 siRNAs does not alter DNA damage response, cancer-related pathways, or tumor-associated microsatellite
84 instability.^{16,21} MutS β inhibition may therefore have a tolerable safety profile, and current evidence supports
85 investigating that possibility, whilst also underscoring a need for pathway-selective inhibition strategies.

86 Despite the therapeutic potential of MutS β inhibition, two major barriers exist: (1) we lack small molecules
87 with selective activity against MutS β confirmed in cells, and (2) we lack robust cellular assays for identifying
88 and characterizing such compounds. The latter challenge represents a particularly significant bottleneck
89 and motivates the development of new methods for measuring MutS β deficiency in cells, to accelerate
90 inhibitor discovery and improve characterization of the few inhibitors already identified.^{35,36} Notably, a
91 common limitation of existing methods for assaying MutS β function is reliance on repeat expansions as
92 endpoint readouts.^{37–41} This design limits the dynamic range of detection relative to background, imposes a
93 need for long periods of cell growth before data collection, and can make the assays difficult to implement.
94 Moreover, assays based on repeat expansion do not always fully resolve the individual contributions of
95 MutS α and MutS β to observed phenotypes. By contrast, *in vitro* biochemical assays provide a way to
96 rapidly measure MutS β activity, often with well-resolved signals,^{42–45} however, these assays lack cellular
97 context and thus do not account for the influence of other cellular pathways or capture indirect effects.

98 Here, we present a suite of cell-based assays that enable reliable, rapid, and selective detection of MutS β
99 inhibition. To develop these assays, we profiled prime editing efficiencies across thousands of edits in cells
100 with targeted disruption of core MMR genes. These screens distinguished distinct MMR activities, including
101 those coordinated by the MutS α and MutS β sensor complexes, and resolved a poorly characterized,
102 noncanonical mechanism of “loop-directed” MMR. Prime edit types found to be selectively suppressed by
103 MutS α or MutS β were then repurposed to build a set of highly scalable, sensor-selective MMR reporters.
104 We used one of these reporters to screen ~72,000 small molecules for activity against MutS β and identified
105 a novel series of MutS β inhibitors. Characterization of these inhibitors revealed selectively for MutS β , with
106 no observed activity against MutS α in cells, and demonstrated reduction of (CAG)_n expansion in a cellular
107 model of somatic instability. Together, our results establish a set of versatile tools for studying MMR and for
108 advancing small molecule inhibitors of MutS β towards clinical applications in repeat expansion diseases.

109 RESULTS

110 Systematic evaluation of MutS and MutL activities in MMR using prime editing

111 Prime editing is a CRISPR-based genome editing technology that can write small insertions, deletions, and
 112 substitutions into genomes using reverse transcription.⁴⁶ We and others have shown that prime editing is
 113 suppressed by MMR, with inhibition occurring in accordance with known MMR substrate criteria.^{47–50} That
 114 is, MMR suppresses prime editing when programmed edits are expected to form MMR substrate-mimicking
 115 intermediates (e.g., short runs of base mispairs, extrusions of 1-10 nt^{51–55}) but has generally less impact
 116 when intermediary editing structures are expected to be poor MMR substrates (e.g., C-C mispairs
 117 introduced by G to C edits^{51,56,57} or longer extrusions of >10 nt^{53,54}). Collectively, these observations have
 118 established a model of prime editing wherein transient DNA mispairs are generated during edit installation
 119 and recognition of those mispairs by MMR leads to edit removal (Figure 1B).

120 The substrate recognition criteria of the two canonical MutS sensors in humans (MutS α and MutS β) have
 121 been well described. MutS α detects mispaired DNA bases and small DNA loop-outs (~1-2 nt),^{58,59} whereas
 122 MutS β recognizes larger loops (~1-15 nt) but not base-base mispairs (Figure 1A).^{42,59,60} However, how
 123 these preferences, which have been defined largely by biochemical analyses, stratify across mispairs in the
 124 cellular environment has been minimally described.⁶¹ Moreover, how three MutL effector complexes
 125 (MutL $\alpha/\beta/\gamma$) coordinate repair downstream of the sensors is incompletely understood, particularly with
 126 respect to repeat instability in neurons.^{62–64} We reasoned that, by profiling the effects of MutS and MutL
 127 genes across many prime edits, each designed to generate a different DNA mispair, we could elucidate
 128 separable MMR subpathways (i.e., those directed by different classes of MMR substrates) (Figure S1A).

129 To test this idea, we generated two prime editing libraries, each encoding a single target sequence (HEK3
 130 or *FANCF*) and 6000 prime editing guide RNAs (5898 and 5922 edit-specifying pegRNAs and 102 and 78
 131 non-editing control pegRNAs, respectively) and each designed to generate thousands of different edits,
 132 and thus thousands of distinct DNA mispair structures, in the corresponding target site (Figures 1C and
 133 S1B; Table S1). Sequencing these libraries detected >99% of the pegRNAs, 90% of which were within a
 134 range of < 4-fold (Figure S1C). After characterizing these libraries, we transduced them into K562 cells
 135 expressing a standard prime editor (PEmax)^{51,65} and several derivative knockout (KO) cell lines, each with
 136 genetic disruption of one of six MMR genes: *MSH6*, *MSH3*, *MLH1*⁵⁶, *PMS2*, *PMS1*, or *MLH3* (Figures
 137 S1D-S1N; Tables S2-S4). These cell lines were transduced at low multiplicities of infection (0.25-0.45) in
 138 one of two screening groups: (1) parental, *MSH6*-KO, *MSH3*-KO, and *MLH1*-KO cell lines for one screen
 139 with our HEK3-targeted library (HEK Screen 1) and one with our *FANCF*-targeted library (*FANCF* Screen),
 140 or (2) parental, *MLH1*-KO, *PMS2*-KO, *PMS1*-KO, and *MLH3*-KO cell lines for a second HEK3 screen

141 (HEK3 Screen 2) (Figures 1C and S1B). Cells were selected for library integration, and after seven days of
142 growth, we quantified intended editing efficiencies by sequencing pegRNA-target pairs and calculating the
143 percentage of corresponding target reads harboring the programmed prime edit (Tables S5 and S6).

144 Prior to any other analyses, we evaluated the reproducibility and validity of the intended editing efficiencies
145 measured in our screens. Indicating low technical variability and high reproducibility, replicate intended
146 editing efficiencies were well correlated (Figure S2A), even when measured from non-concurrent screens
147 (i.e., conditions repeated in HEK3 Screens 1 and 2). Moreover, when stratified by edit length and edit type,
148 intended editing efficiencies from MMR-proficient parental cells confirmed two observations from previous
149 studies (Figures 1D and S2B): (1) substitutions of ≥ 3 contiguous base pairs (bps) were installed more
150 efficiently than single substitutions or those spanning 2 bps,⁵¹ even when restricting comparisons to edits
151 that did not disrupt the PAM sequence (8.1-fold increase for 3-4 bp consecutive substitutions in HEK3, $p <$
152 0.0001, and 2.1-fold increase for those substitutions in *FANCF*, $p < 0.0001$), and (2) small insertions were
153 installed more efficiently than deletions of similar size.⁶⁶ The first of these observations has been attributed
154 to poor MutS recognition of contiguous mismatches,⁵¹ while the second has yet to be explained.

155 We next quantified the impact of *MSH6*, *MSH3*, *MLH1*, *PMS2*, *PMS1*, and *MLH3* on edits evaluated in our
156 screens by calculating log₂ fold changes in intended editing efficiencies between K562 PEmax parental
157 cells and the knockout cell lines. These measurements were also reproducible between replicates (Figures
158 1E and S2C) and captured expected features (Figures 1F and S2D-S2F). For example, disruption of *MLH1*,
159 which causes loss of all three MutL complexes and thus broadly impedes MMR (Figures S1M and S1N),
160 increased all types of small prime edits (~1-8 bps) but had generally less effect on longer edits (>8 bps).^{51,53}
161 Additionally, loss of each of the two MutS sensors in *MSH6*-KO and *MSH3*-KO cells impacted distinct
162 groups of edits (Figure 1G); that is, we observed an increase in small insertions, small deletions, and
163 substitutions (consecutive and non-consecutive) in *MSH6*-KO cells, which lack expression of MutS α but not
164 MutS β , and an increase in deletions across a wider range of lengths, some insertions, and generally not
165 substitutions in *MSH3*-KO cells, which lack expression of MutS β but not MutS α (Figures 1F, S1H, S1I, and
166 S2D-S2F). These results agree with prior examination of a limited number of prime edits in *MSH6* and
167 *MSH3* knockout cells⁵⁴ and, alongside expected mispair intermediates for each stratified edit type (Figures
168 1C and S1B), reinforce known MutS substrate recognition rules (Figure 1A).

169 We next considered results from our MutL knockout cell lines: *MLH1*-KO, *PMS2*-KO, *PMS1*-KO, and
170 *MLH3*-KO. Although eukaryotic cells express multiple MutL homologs, MutL α is considered the major
171 effector of canonical MMR,⁶⁷ which occurs in five steps: (1) mispair recognition by MutS α or MutS β , (2)
172 recruitment of the MutL α complex, (3) MutL α -catalyzed nicking of the error-containing DNA strand, which is
173 discriminated from the “correct” DNA strand by a preexisting strand discontinuity or nick somewhere

adjacent to the mispair, (4) excision of the error-containing DNA, and (5) DNA resynthesis.^{23,68} In our screens, PMS2-KO cells, which lack MutL α but not the other MutL complexes, demonstrated the strongest and most consistent promotion of prime editing among MutL-specific knockout cell lines, and generally phenocopied MLH1-KO cells, which lack all three MutL complexes (Figure 1H). These results confirm a major role for MutL α -catalyzed MMR in rejecting prime edited DNA, likely through excision of the inherently discontinuous, reverse-transcribed flaps or from interference with flap annealing⁵¹ (Figure 1B).

Of significance, our screen data also revealed several unanticipated results. For example, we found that the editing efficiencies of 1 bp insertions were decreased in MSH3-KO cells, relative to parental K562 PEmax cells, and editing efficiencies of 2 bp insertions were minimally altered in those cells (Figures 1F and S2D-S2F). According to the current model of prime editing (Figure 1B), which evokes only canonical MMR, those edits should be more efficiently installed without MutS β , similar to effects observed for slightly longer insertions (e.g., 3-6 bp) and deletions of the same size (1-2 bp). Additionally, we found that PMS1-KO and MLH3-KO cells, which lack MutL β and MutL γ , respectively, produced intriguing phenotypes; that is, we observed increased editing efficiencies for small deletions and, surprisingly, decreased editing efficiencies for similarly sized insertions (~1-8 bp) (Figure 1I).

Together, these unexpected results suggest that a noncanonical mechanism of “loop-directed” MMR, which was recently described *in vitro*,^{69,70} acts on some prime edit types. In this mechanism, MutS β -stimulated MutL γ cleaves DNA across from small loop-out structures to retain extruded DNA in repaired duplexes. This mechanism should, accordingly, suppress small deletion edits but have opposite effects on small insertions (Figure 1J). Isolated loss of this mechanism in MLH3-KO cells, which lack the relevant endonuclease, MutL γ , provides a compelling explanation for the deletion and insertion phenotypes observed in those cells (Figures 1I and S2D). Moreover, parallel activity of this mechanism alongside canonical MMR explains the idiosyncratic insertion phenotypes observed in cells lacking MutS β (i.e., MSH3-KO cells). To elaborate, if loop-directed MMR acts on loop-out structures formed during editing, those intermediates should be processed by one of two MutS β -driven MMR pathways: (1) MutL α -catalyzed removal of the reverse transcribed DNA (i.e., canonical MMR) or (2) MutL γ -catalyzed removal of the strand opposite the loop. These pathways should therefore both suppress small deletion edits (Figure 1J, left), but we expect them to exert opposing effects on small insertions, due to different positioning of the loop-out relative to the reverse transcribed flap (Figure 1J, right). Under this dual-pathway model, loss of MutS β would consistently increase small deletion editing, as observed, but insertion phenotypes would likely vary, depending on the relative activity of each subpathway at the initiating substrate. Our data thus suggest an expanded model of prime editing, in which installation of small insertions, but not deletions, is aided by an endogenous repair mechanism. This model explains why small insertion edits are more efficiently installed

than small deletions in MMR-proficient cells⁶⁶ (Figures 1D and S2B). Additionally, these results show that, despite complex interactions among MMR genes, prime editing can resolve distinct MMR pathways.

Development of MutS α / β -specific reporters using different prime edit types

Our genetic screens showed that prime edits can differentially engage distinct MMR activities, with several edit types able to distinguish MutS α - and MutS β -driven MMR. For example, most 1-5 bp substitutions were suppressed by *MSH6* but not *MSH3*, while 4-6 bp deletions had reciprocal specificities (Figures 1F and S2D-S2F). Further evaluation of 1 bp substitutions and 5 bp deletions across additional target site sequences in K562 cells confirmed that loss of *MSH6* but not *MSH3* consistently increased installation of the former edit type (~2-14 fold) whereas *MSH3* but not *MSH6* had an inhibitory effect on the later (~2-10 fold) (Figure 2A). Similar results from two additional cell lines (i.e., HAP1 and RPE-1 hTERT cells) suggest that these sensitivities reflect intrinsic, sensor-specific stimulation of MMR, rather than cell line-specific effects (Figure S3). Moreover, confirming that these phenotypes are not artifacts of clonal selection, ectopic expression of *MSH6* in K562 and HAP1 *MSH6*-KO cells reduced 1 bp substitution editing, effectively restoring suppression towards levels observed in parental cells, and ectopic expression of *MSH3* in K562 and HAP1 *MSH3*-KO cells showed similar effects on a 5 bp deletion (Figures 2B and 2C, Methods). These results establish that 1 bp substitutions and 5 bp deletions can robustly, specifically, and consistently stimulate the activity of only one MMR sensor complex, making their editing efficiencies reliable readouts of sensor-directed MMR.

We next leveraged these sensor-selective edit classes to build a series of reporters that detect disruption of either MutS α - or MutS β -directed MMR. We envisioned reporters in which installation of a MutS α - or MutS β -suppressed edit would install an in-frame start codon and thus turn on a detectable gene product (Figures 2D and 2E). Elevated editing upon inhibition or loss of the corresponding MMR sensor would, in turn, increase expression of the indicator gene (herein, luciferase) and signal loss of that sensor. Unlike many alternative MMR assays,⁷¹⁻⁷⁶ these reporters would be delivered with lentiviral transduction and would thus monitor MMR occurring directly in the genome. Notably, to detect delivery, our reporters would also constitutively express a marker of transduction from the same mRNA as the indicator gene, driven by an internal ribosome entry site (IRES),⁷⁷ similar to a system we used previously⁶⁵ (Methods).

To build these reporters, we chose eighteen target sequences previously shown to support prime editing^{56,78} and designed eighteen 1 bp substitution and 5 bp deletion edits to introduce an ATG codon or its reverse complement (CAT) into each of those sequences. We cloned pegRNA sequences specifying these edits and the corresponding targets into the same expression constructs, with the targets positioned between a promoter driving mRNA expression and the indicator gene such that the specified edits would generate an

in-frame ATG codon on the coding strand (Figures 2D and 2E; Table S7; Methods). We then evaluated each of the resulting “self-targeting” reporters in cells expressing PEmax, measuring induction of luciferase in parental cells and MSH3-KO, MSH6-KO, and MLH1-KO derivatives three days post reporter transduction (Figures 2F and 2G; Table S7). All six constructs designed with 1 bp substitutions were robustly activated in MSH6 knockout cells but not MSH3 knockouts. We refer to these as our MutS α -specific reporters. Conversely, all twelve constructs designed with 5 bp deletions showed reciprocal specificities. We refer to these as our MutS β -specific reporters. Notably, all 18 reporters were activated in cells lacking MLH1, consistent with loss of both MutS α - or MutS β -driven MMR. Collectively, these findings establish a prime editing-based platform for selectively detecting inhibition of MutS α - and MutS β -directed MMR.

Establishing MutS α / β -specific reporters as robust and scalable readouts of sensor-resolved MMR

We selected two reporters for further characterization: MutS α -specific #4 and MutS β -specific #T5 (Figures 2F and 2G; hereafter referred to without the #4 and #T5 identifiers). Testing these reporters in parental K562 PEmax cells and seven derivative knockout cell lines recapitulated phenotypes expected from our genetic screens; that is, disruption of MSH3, MSH2, MLH1, PMS2, PMS1, and MLH3 activated the MutS β -specific reporter, but disruption of MSH6 did not; whereas, disruption of MSH6, MSH2, MLH1, and PMS2 robustly activated the MutS α -specific reporter, but disruption of MSH3 and MLH3 did not (Figure 3A). These phenotypes were stable over consecutive days (Figure S4A), were highly correlated with changes in intended prime editing determined by sequencing (Figure S4B), and were reversible by re-expression of the disrupted genes, with the strength of rescue often scaling with the level of ectopic protein expression (Figures 3B, 3C, and S4C-S4H; Methods). We also found that reporter phenotypes were robust across experimental conditions, as strong effects reproduced with an alternative delivery vector (Figure S4I), an additional cell line (RPE-1 hTERT; Figure S4J), and a distinct method of perturbing MutS β (siRNA; Figure S4K). Moreover, in addition to pathway disruption, our MutS β -specific reporter was able to register increased MutS β -driven MMR, as evident in the reduction of reporter signal observed upon strong overexpression of MSH3 (Figures 3B, S4D, and S4L). These results establish our MutS α / β -specific reporters as robust and reliable indicators of sensor-resolved MMR, amenable to functional studies.

As proof-of-concept for chemical applications, we next used our reporters to evaluate two recently described MMR inhibitors: cmpd1,³⁵ which has been shown to covalently bind the DNA-binding domain of MSH3 *in vitro* but has unresolved cellular activity,³⁶ and a recently disclosed PMS2 inhibitor⁷⁹ (NP1867, hereafter called PMS2i). Indicative of selective MutS β inhibition, a single dose of cmpd1 (7.5 μ M) activated our MutS β -specific reporter in K562 PEmax cells (with increased signal relative to controls observed as early as 24 hours post treatment), but, under similar conditions, did not activate our MutS α -specific reporter (Figure 3D). Further testing revealed that the effect of cmpd1 on our MutS β -specific reporter was

dose-dependent, with maximal activation observed at concentrations near the IC_{50} for cytotoxicity of *cmpd1*, beyond which activation of the reporter was diminished, likely due to reduced cell viability (Figures S5A and S5B). These results show that *cmpd1* selectively inhibits MutS β -driven MMR in cultured cells. Moreover, we observed similar activity for *cmpd1* in three additional cell lines (HAP1, HeLa, U-2 OS; Figure S5C), supporting a general effect, but as expected, *cmpd1* did not activate our reporter in MMR-deficient cells (Lenti-X 293T, HCT 116; Figure S5D). PMS2i, by contrast, activated both of our primary reporters, consistent with MMR inhibition downstream of both MutS α and MutS β (Figures 3E and S5E).

Having established that our reporters robustly detect chemical inhibition of MMR, we next sought to apply our MutS β -specific reporter to screen for additional MutS β inhibitors. Towards that end, we developed a scalable, four-day reporter assay (Figure 3F) wherein K562 PEmax cells transduced with the MutS β -specific reporter are grown for two days, dispensed into microwell plates containing test and control compounds, and analyzed for luminescence with Nano-Glo luciferase assay reagents two days later. We rigorously optimized the performance of this assay by (1) determining transduction conditions that provide maximal activation (Figure S5F), (2) confirming that NanoLuc provided the largest dynamic range among commonly used luciferases, and establishing conditions for use in both 96- and 384-well plates (Figure S5G). Together, these efforts produced a scalable MutS β -specific reporter assay for high-throughput, array-based small molecule screening, designed to detect cell-active inhibitors of MutS β -driven MMR.

Application of a MutS β -specific reporter to high-throughput small molecule screening

Using our optimized MutS β -specific reporter assay (Figure 3F), we screened 71,623 small molecules for MutS β inhibitors. We screened two collections: one containing 5,923 bioactive molecules with annotated targets and known effects, and another comprising 65,700 molecules with drug-like structures but no annotated targets or effects (organized as 47,780 individual compounds and 4,480 mixtures of stereoisomers). We evaluated the smaller, bioactive collection on 20 microwell plates at two concentrations (10 μ M and 20 μ M) and the much larger, general collection on 164 plates at only a single concentration (10 μ M). On each plate, we included 32 wells of *cmpd1* (10 μ M) and 32 wells of solvent (DMSO) to serve as positive and negative controls, respectively. Using data from these controls, we calculated per-plate Z'-factors⁸⁰ and, to ensure high data quality, rescreened plates with low Z'-factors (< 0.2, 9 total plates). Indicative of good dynamic range and low variability, our final data set had a mean Z'-factor of 0.558 ± 0.095 (Figure S5H). To quantify inhibition of MutS β -driven MMR for each compound, we next calculated background-subtracted luminescence for each well and normalized those measurements to on-plate *cmpd1* controls, generating relative inhibition scores (Figures 3G and S5I; Methods). Hit molecules were determined as those for which this metric exceeded the DMSO control average by ≥ 3 standard deviations.

From the bioactive collection, we identified 55 hits. Providing confidence in these findings, results from this collection were reproducible across the two concentrations screened (Pearson's correlation $r = 0.86$; Figure S5J) and roughly a third of the hits (18) scored at both 10 μM and 20 μM (Table S8). Moreover, one hit from the bioactive collection, the FDA-approved antiviral Ribavirin, scored repeatedly from multiple collection entries (i.e., three times at 10 μM and twice at 20 μM). Additionally, retesting the strongest hits from this collection (i.e., 28 with > 0.1 normalized reporter activation) confirmed results for 11 molecules, albeit often with a weaker effect than in primary screening. From these 11 molecules, we resynthesized the 3 that maintained normalized reporter activation > 0.1 in retesting and found that 2 of them, Ribavirin and Liothyronine, elicited MutS β -specific reporter activation across a range of concentrations (Figures S5K and S5L). However, similar to most strong hits from this collection (Table S8), these 2 molecules also activated our primary MutS α -specific reporter and thus did not show specificity for MutS β -driven MMR.

Although much larger, the general collection produced only 5 hits (Figure 3G); however, among those hits, we identified an analog of cmpd1 and three novel molecules that share an identical chemical scaffold, differing at only two functional groups (R_1 and R_2 ; red inset in Figure 3G). Additionally, among compounds that scored just below our hit threshold, we identified a fourth compound with the same scaffold (normalized reporter activation = 0.07). We resynthesized these compounds and found that all four of them activated our MutS β -specific reporter in a dose-dependent manner and, indicative of sensor specificity, did not activate our MutS α -specific reporter (Figure 4A). We refer to these compounds as our SBi series of molecules.

Structure-activity relationship (SAR) analysis of candidate MutS β inhibitors

In principle, hits from our screen could act through several indirect effects (e.g., nonspecific stimulation of prime editing or luciferase expression); however, because the SBi molecules identified in our screen activated our MutS β -specific reporter but not our MutS α -specific reporter during counterscreening, those compounds are unlikely to operate through such effects. To confirm that selective reporter activation by these molecules was not limited to our primary reporter designs (e.g., did not represent selective suppression of only the exact DNA mispair specified in MutS β -specific #T5), we next evaluated two of the molecules, SBi-001 and SBi-004, with several additional reporters, each designed to generate a distinct but still sensor-selective DNA mispair (Figures 2D–2G). Consistent with broad pathway-level inhibition of MutS β -driven MMR, SBi-001 and SBi-004 activated four additional MutS β -specific reporters, with similar dose-dependencies, and failed to activate two additional MutS α -specific reporters (Figures 4B and S6A). We therefore next asked if activation of our primary MutS β -specific reporter by SBi-001 and SBi-004 was dependent on MutS β , as would be expected for on-target pathway inhibition, and we found that neither

compound activated that reporter in cells lacking MutS β (Figures 4C and S6B). Altogether, these results suggest that SBI molecules selectively inhibit MutS β -driven MMR.

Notably, at the highest concentration tested (20 μ M), SBI compounds from our screen decreased signals from our MutS α / β -specific reporters relative to DMSO-treated controls (Figures 4A, 4B, and S6A). These effects, which mirrored loss of signals observed with cmpd1 (Figure S5A), coincided with reductions in cell viability and were thus likely due to nonspecific cellular toxicity (Figure 4A). To identify less toxic SBI compounds, we tested 45 analogs, each with the same or highly similar scaffolds and different functional groups (R_1 and R_2) (Figure 4D; Table S9). Comparative analysis showed that MutS β -specific reporter activation was unaffected by small changes to R_1 but strongly influenced by the amide amine at R_2 (Figure 4E); that is, analogs with a 5-7 membered-ring tertiary amide at R_2 exerted a larger effect than those with a secondary benzylic amide group. Altogether, 16 analogs (36%) reproduced activation of our primary MutS β -specific reporter, and while none of these compounds produced a stronger effect than the most potent SBI molecule from our screen (SBI-001), one outperformed all other screen-nominated SBI molecules and showed no toxicity in K562 cells up to the highest concentration tested (20 μ M; SBI-005) (Figures 4D, 4F and S6C). Active SBI compounds with less toxicity can thus be identified—albeit with possible cell type differences, as both SBI-001 and SBI-005 reduced reporter activation at the highest concentrations in RPE-1 hTERT cells (Figures S6D and S6E).

We next investigated changes to the SBI scaffold region (1H-pyrido[2,3-e][1,3,4]thiadiazine 4,4-dioxide). Across 53 analogs with scaffold modifications, none activated our MutS β -specific reporter more strongly than SBI-001. However, several analogs demonstrated inverted reporter specificities, in that they elicited weak activation of our MutS α -specific reporter but showed no activation of the MutS β -specific reporter (Figures S6F and S6G; Table S9). Scaffolds in these molecules often consisted of a benzene ring fused to a six-membered thiazine ring (4H-benzo[b][1,4]thiazine 1,1-dioxide) and thus contained two less nitrogen atoms than the fused pyridine and thiadiazine rings in the SBI scaffolds. We refer to such analogs as SAi molecules (Figures S6H-S6J). Intriguingly, we also found a single compound (Si-001) that weakly activated our MutS α -specific reporter and showed moderate MutS β -specific reporter activity (Figures S6F and S6K). The scaffold in this molecule (1H-benzo[e][1,3,4]thiadiazine 4,4-dioxide) was similar to those in SBI and SAi compounds but had an intermediate number of nitrogen atoms: one less than in the SBI scaffold and one more than in the SAi scaffold. These results suggest that subtle changes to the bicyclic scaffold can alter MutS pathway sensitivity, possibly by differentially influencing interactions with MutS α and MutS β .

366 SBI compounds inhibit MutS β binding to DNA mispairs

367 We next sought to define the molecular mechanisms by which our SBI compounds inhibit MutS β -directed
368 MMR. We first considered two indirect mechanisms: (1) reduction of MutS β expression and (2) nonspecific
369 interference of MutS β binding to DNA from DNA intercalation. To determine whether SBI compounds affect
370 MutS β expression, we quantified MSH3 mRNA and protein levels in K562 PEmax cells with and without
371 SBI-001 and SBI-005 (Figures S7A and S7B). Neither compound diminished expression at active
372 concentrations. To evaluate the possibility of nonspecific DNA intercalation, we used surface plasmon
373 resonance (SPR) and quantified the interaction of SBI-005 with a small DNA mispair (i.e., two looped-out
374 CAGs within a 38 bp heteroduplex; Figure S7C). Unlike the known DNA intercalator mitoxantrone, which
375 bound the mispaired DNA starting at low concentrations ($>0.1 \mu\text{M}$), we observed no evidence of an
376 interaction with SBI-005 up to the highest concentration of compound tested ($100 \mu\text{M}$). These results
377 suggest that SBI compounds do not inhibit MutS β -directed MMR through these indirect mechanisms.

378 Given these results, we explored the possibility that SBI molecules act on MutS β directly. We first asked
379 whether SBI-005 alters binding of MutS β to the same 38 bp (CAG)₂ loop-out DNA using SPR and observed
380 concentration-dependent reductions in normalized SPR signal, consistent with inhibition of MutS β binding
381 to the substrate (Figures 5A, S7D, and S7E; Methods). In parallel, we evaluated the impact of SBI-005 on
382 MutS α using a heteroduplex harboring a canonical MutS α substrate (i.e., 38 bp DNA with a G-T mismatch),
383 and observed a comparably modest effect, suggesting selectivity for MutS β over MutS α (Figure 5B).

384 We next deployed an orthogonal MutS β binding assay, which uses Förster resonance energy transfer
385 (FRET) to measure heteroduplex bending upon interaction with MutS β (herein, a 70 bp substrate with one
386 CAG loop-out; Figure 5C). When performed with ATP, this assay reports an equilibrium FRET ratio that
387 reflects the distribution of three cycling reaction states: MutS β bound to the loop-out, ATP-dependent
388 sliding of MutS β on DNA, and unbound MutS β .⁸¹ Further supporting compound-dependent inhibition of
389 MutS β -DNA binding, SBI-005 caused dose-dependent reductions in this equilibrium ratio, with higher
390 SBI-005 concentrations producing ratios closer to controls performed without MutS β (Figure 5D). Notably,
391 in the absence of ATP, SBI-005 exhibited lower potency (Figure S7F), possibly due to a higher threshold for
392 inhibition in that context caused by formation of more stable MutS β -DNA complexes, as expected from
393 reduced release of MutS β from DNA without ATP, and as evident in stronger MutS β -induced FRET ratios.

394 Downstream of MutS β , MutL heterodimers (MutL α or MutL γ) nick the mispaired DNA to coordinate repair.
395 We therefore next asked whether SBI molecules also inhibit these events. In an established *in vitro* assay,⁶⁹
396 SBI-001 and SBI-005 compounds both compromised MutS β :MutL α -mediated cleavage of a plasmid
397 harboring an extrahelical (T)₄ loop-out, notably with stronger effect than cpmd1, which we included as a

positive control (Figure 5E). Additionally, in a similar assay, SBI-001 inhibited MutS β :MutL γ -mediated cleavage of the same substrate; although neither cpmd1 nor SBI-005 elicited a similar effect in this assay, possibly due to differences in the potencies required for inhibition (Figure 5F). Given these results, we next tested the effect of SBI-001 on a separate nuclease, FAN1, and observed no difference in cleavage with or without the compound, indicating that SBI compounds do not indiscriminately inhibit endonuclease activities (Figure S7G). Alongside our other findings, these results support a model wherein SBI compounds inhibit MutS β -driven MMR by interfering with MutS β substrate recognition, upstream of DNA cleavage.

SBI compounds attenuate CAG repeat expansion

Identifying selective inhibitors of MutS β -directed MMR is an important step; however, therapeutic potential requires that inhibition from candidate compounds translates into suppression of repeat expansion. We therefore asked whether our SBI molecules could attenuate repeat expansion in RPE-1 hTERT cells harboring a stably integrated (CAG)₁₂₉ mHTT exon 1. Cells were grown with SBI-001, SBI-005, and cpmd1 at 1 μ M or 2.5 μ M, depending on expected toxicities, collected for initial analysis after 2 days (designated day 0 samples), and then grown for 21 days with compound replenishment and intermittent sampling. CAG tracts were amplified from genomic DNA and analyzed by capillary electrophoresis (Figure 5G). Comparison of modal repeat lengths revealed significantly decreased (CAG)_n expansion in cells treated with SBI-001 and SBI-005, compared to DMSO controls (Figures 5H and 5I). Moreover, despite potential for attenuation of repeat expansion to arise from indirect effects on the cell cycle (i.e., reduced occupancy in S phase resulting in fewer replication-associated strand slippages), cell cycle analysis of similarly treated RPE-1 hTERT (CAG)₁₂₉ cells showed only mild, inconsistent changes, suggesting that observed effects were more likely caused by inhibition of MutS β (Figure S7H). These results solidify the utility of our prime editing-based reporter assays for identifying MutS β inhibitors relevant to cellular repeat expansion.

DISCUSSION

Previous studies have suggested that prime editing forms transient DNA mispairs at targeted genomic loci,^{51–55} and because prime editing is easily programmable, the technology allows flexible modeling of those editing intermediates directly in genomes (e.g., base-base mismatches and insertion/deletion loop-outs). We exploited the scalability of prime editing to profile the effects of core MMR genes across thousands of specified mispairs. Our results revealed information-rich “fingerprints” of MMR gene functions, resolved the substrate preferences of MutS α and MutS β at scale, and isolated poorly characterized MMR activities in cells. Among key observations, we found that *MLH3* suppresses small deletions and promotes similarly sized insertion edits. These phenotypes, alongside those observed in cells lacking *MSH3*, support a mechanism of loop-directed MMR, which has been defined biochemically^{69,70} but not resolved in cells.

430 This noncanonical, MutL γ -mediated mechanism favors retention of looped-out DNA in repaired duplexes
431 and is thus highly relevant to repeat expansion. We identified phenotypes consistent with activity of this
432 mechanism in cells and thus provide a straightforward way to investigate it further (e.g., characterize
433 sequence features).

434 While many questions remain about MutL γ -mediated MMR, less is known about the role of MutL β in DNA
435 repair. Like MutS β , MutL β contributes to (CAG) $_n$ expansion^{8,9,18,19} and is therefore an HD drug target of
436 interest⁸²; however, because MutL β lacks the DQHA(X) $_2$ E(X) $_4$ E endonuclease motif present in other MutL
437 complexes⁶⁷ and exhibits limited MMR function *in vitro*,⁸³ its role in MMR has been unclear. Addressing this
438 gap, we found that loss of MutL β in PMS1-KO cells produced phenotypes similar to those caused by loss
439 of MutL γ in MLH3-KO cells; that is, we observed nearly identical changes in deletion editing and more
440 subtle but concordant effects on insertions (Figures 1F, 1I, and S2D). These similarities suggest that, like
441 MutL γ , MutL β promotes loop-directed MMR. Supporting this idea, a recent study showed that MutL β
442 enhances MutS β -stimulated MutL γ endonuclease activity *in vitro* and blocks an inhibitory, potentially
443 antagonistic effect from MutS α .⁸⁴

444 Two additional, related observations from our genetic screens are that (1) MutL α deficiency in PMS2-KO
445 cells produced phenotypes similar to those from MLH1-KO cells, which lack all three MutL complexes, and
446 (2) phenotypes attributed to *MLH1* did not generally account for effects caused by targeted disruption of
447 MutL β or MutL γ in PMS1-KO and MLH3-KO cells, respectively. These observations suggest that epistatic
448 relationships exist among MutL complexes, and given that MutL β and MutL γ are promising HD drug
449 targets⁸⁵ but MutL α is not, resolving these relationships will be of high interest. Our genetic screening
450 framework, which should pair easily with technologies that disrupt multiple genes,^{86–94} will enable such
451 future studies.

452 In addition to providing insight into cellular mechanisms of MMR, our genetic screens resolved classes of
453 prime edits selectively suppressed by the two MMR sensors, MutS α and MutS β . Guided by these results,
454 we built a set of scalable, sensor-selective MMR reporters with several advantages over alternative
455 methods for evaluating MMR in cells. Unlike other methods, these reporters measure separable MMR
456 activities within only a few days, have high dynamic range, and, due to having a positive readout (i.e.,
457 luciferase expression), minimize false positives from cellular toxicity. In proof-of-principle studies, we used
458 these reporters to identify and characterize a series of selective MutS β inhibitors. Additional studies will be
459 required to determine the therapeutic potential of these and other SBi-type molecules; however, those
460 identified herein already provide valuable tools for investigating MutS β function and for exploring MutS β
461 targeting strategies (e.g., identifying druggable pockets). More broadly, the identification of these molecules
462 definitively establishes our reporters as useful tools for therapeutic discovery. Applied across larger

libraries, we expect these reporters to identify even more potent and less toxic selective inhibitors of MutS β , and because our MutS β -specific reporters register loss of multiple HD modifier genes (i.e., *MLH3* and *PMS1*), such screens may also uncover compounds against those targets.

Notably, our reporters can also be applied *post hoc* to characterize MMR inhibitors found through orthogonal approaches (e.g., from *in vitro* screens), as we showed for cmpd1 and one PMS2 inhibitor. Moreover, as drug development efforts advance, these reporters can be used to compare pharmacological inhibition of MMR to loss of MMR components or expression of inactive variants, each of which may block MMR via distinct mechanisms. Additionally, with straightforward modifications, our reporters could be adapted to enable massively parallel characterization of MMR genetic variants, including those from genome-wide association studies, or to catalog regulators of separable MMR pathways (e.g., loop-directed MMR) with large-scale genetic screens. Such applications will primarily require a simple design change; that is, the current luciferase-encoding reporter gene would need to be replaced by a gene more amenable to pooled screening (e.g., GFP).⁶⁵

Herein, we also establish new ways to improve prime editing. To date, multiple studies have confirmed that canonical MMR suppresses prime editing with small deletions, small insertions, and substitutions.^{51–57,65,95} To overcome this limitation, methods for transiently inhibiting MMR have been developed; however, these methods rely on expression of a dominant negative protein⁵¹ or an engineered peptide⁹⁵ and thus are limited by delivery challenges. Compared to these methods, the small molecule inhibitors identified and characterized in our study offer advantages, including easy and flexible delivery. For improvement across a wide range of edit types, we recommend the PMS2 inhibitor. Additionally, our findings establish an expanded model of prime editing, wherein installation of small insertions, but not deletions, is aided by an endogenous MMR subpathway. This finding changes our understanding of prime editing and explains why small insertions are generally preferable to deletions in applications of editing in MMR-proficient cells.

Altogether, our work establishes several versatile, cell-based assays for interrogating MMR. These systems provide powerful tools for studying mechanisms of genome maintenance and somatic instability in repeat expansion diseases, with demonstrated capacity to identify inhibitors of significant therapeutic value.

LIMITATIONS OF THE STUDY

Several limitations of our study should be considered. With respect to assay development, we highlight two limitations that complicate MMR pathway modeling: (1) While programmed heteroduplex structures generated by prime editing likely arise as expected, unprogrammed structures may also form (e.g., driven by alternative flap hybridizations). Given this potential, we focused on median phenotypes when analyzing results from our genetic screens (i.e., phenotypes observed across edit classes), rather than individual

effects; however, by accounting for heteroduplex ensembles, future studies may further resolve MMR subpathways. (2) Because prime editing-induced heteroduplex structures invariably form near a DNA strand discontinuity, programmed mispairs recognized by MutS α and MutS β will likely always stimulate canonical, MutL α -catalyzed repair. This feature unavoidably complicates isolation of other MMR processes, especially those that may operate on continuous, unbroken DNA. Moreover, because edit-associated strand discontinuities occur invariably on the 3' side of programmed mispairs, repair through canonical MMR may be biased towards 3' nick-directed mechanisms, rather than 5' nick-directed ones.⁹⁶ As highlighted above, future studies may address these challenges by performing experiments in defined genetic backgrounds, for example by knocking out *PMS2* to avoid MutL α activity or overexpressing DNA ligase to promote ligation of the 3' nick. Additionally, deploying our assays in non-dividing cells (e.g., cultured or iPSC-derived neurons) may further improve translational relevance for specific applications.

With respect to the early-stage MutS β inhibitors identified in this study, we note generally pervasive toxicity. Two lines of evidence argue that this toxicity is not driven by on-target MutS β inhibition: (1) our MSH3-KO cells, which lack MutS β , were comparably viable, with no obvious growth defects, and (2) one analog of these molecules (SBI-005) retained targeted activity but showed significantly less toxicity in K562 cells, suggesting a more tolerable chemical structure, at least for that cell type. Additional studies will be required to distinguish off-target liabilities and determine the precise mechanisms of action of these molecules, however, pending such studies, this consideration motivates continued search for additional small molecules suitable for therapeutic development—efforts that will be greatly accelerated by our advances.

514

Acknowledgements. We thank Aaron Lin and other Adamson Lab members for discussions and suggestions. We thank Wei Wang, Jennifer M. Miller, Jean Arly Volmar, Julia E. Good, and Alexander Harms from the Genomics Core Facility at Princeton University, and Christina DeCoste and Jailene Garcia from the Flow Cytometry Resource Facility at Princeton University (partially supported by the Rutgers Cancer Institute of New Jersey, NCI-CCSG P30CA072720-5921). We thank Brinda Prasad and Simon Noble at the CHDI Foundation for their advice and suggestions on the manuscript, and Kezhi Chen at the University of Chicago for chemistry discussions. This work was supported by the nonprofit CHDI Foundation Inc. CHDI Foundation is a nonprofit biomedical research organization exclusively dedicated to collaboratively enabling the development of therapeutics that substantially improve the lives of those affected by Huntington's disease. Research in the Adamson lab was also supported by the National Institutes of Health (NIH) under award numbers R35GM138167 and RM1HG009490, Ludwig Cancer Research, and Princeton University, in part through the Intellectual Property Accelerator Fund. Research in the Balmus Lab was also supported by the UK Dementia Research Institute, with funding from the UK MRC, and the Romanian Ministry of Research, Innovation, and Digitization (grant #PNRR-III-C9-2022-I8-66, contract 760114). Yangwode Jing was supported by the George B. Rathmann *51 Graduate Fellowship and the James Mi *91 and Teresa Mi S *91 Graduate Fellowship from Princeton University. Jun Yan was supported by a fellowship provided by the China Scholarship Council (CSC), based on the April 2015 Memorandum of Understanding between the CSC and Princeton University. Yangwode Jing dedicates this work to his grandparents and parents, by whom he was raised up.

Author contributions. Conceptualization, Y.J., J.Y., and B.A. Generation and characterization of cell lines, Y.J. with J.Y. and P.R. Self-targeting prime editing screens, Y.J. with J.Y. and J.L. Formal analysis of prime editing screens, Y.C. with Y.J. and J.Y. Interpretation of prime editing screens, Y.J., J.A.H., and B.A. Construction of MutS α - and MutS β -specific reporters, Y.J. with J.Y. and K.W. Small molecule screen, Y.J. with J.Y. and H.K. SAR studies, Y.J., H.K., L.L., and M.F. SPR and FRET experiments and analysis, H.D., V.D., and M.T. *In vitro* MMR assays, I.S. Repeat expansion assays, D.Y., J.W., and K.H. All other experiments and data analysis: Y.J. Project administration, M.F., T.F.V., and B.A. Supervision, Y.P., G.B., P.C., M.F., and B.A. Writing, Y.J. and B.A. with the input from all authors.

Competing interests. B.A. has/had financial interest in Arbor Biotechnologies and Tessera Therapeutics. Y.J., J.Y., H.K., and B.A. are listed as inventors on patent application(s) related to this work and/or other prime editing technologies. H.K. is a member of the Rutgers Cancer Institute of New Jersey (RCINJ); a co-founder, director, stockholder, and consultant for Crescenta Biosciences Inc. and Farber Partners LLC; a co-founder and shareholder of Chiromics LLC; a shareholder and scientific advisor to Colorado Research Partners LLC and Integrated Biosciences Inc.; and an inventor of other patents held by Princeton University. M.T., H.D., and V.D. are employees of Proteros biostructures GmbH. G.B. is a co-founder, director, stockholder, and board member for FunctionRx Ltd. D.Y. is a consultant for FunctionRx Ltd. The remaining authors declare no competing interests.

551 MAIN FIGURES AND LEGENDS

FIGURE 1

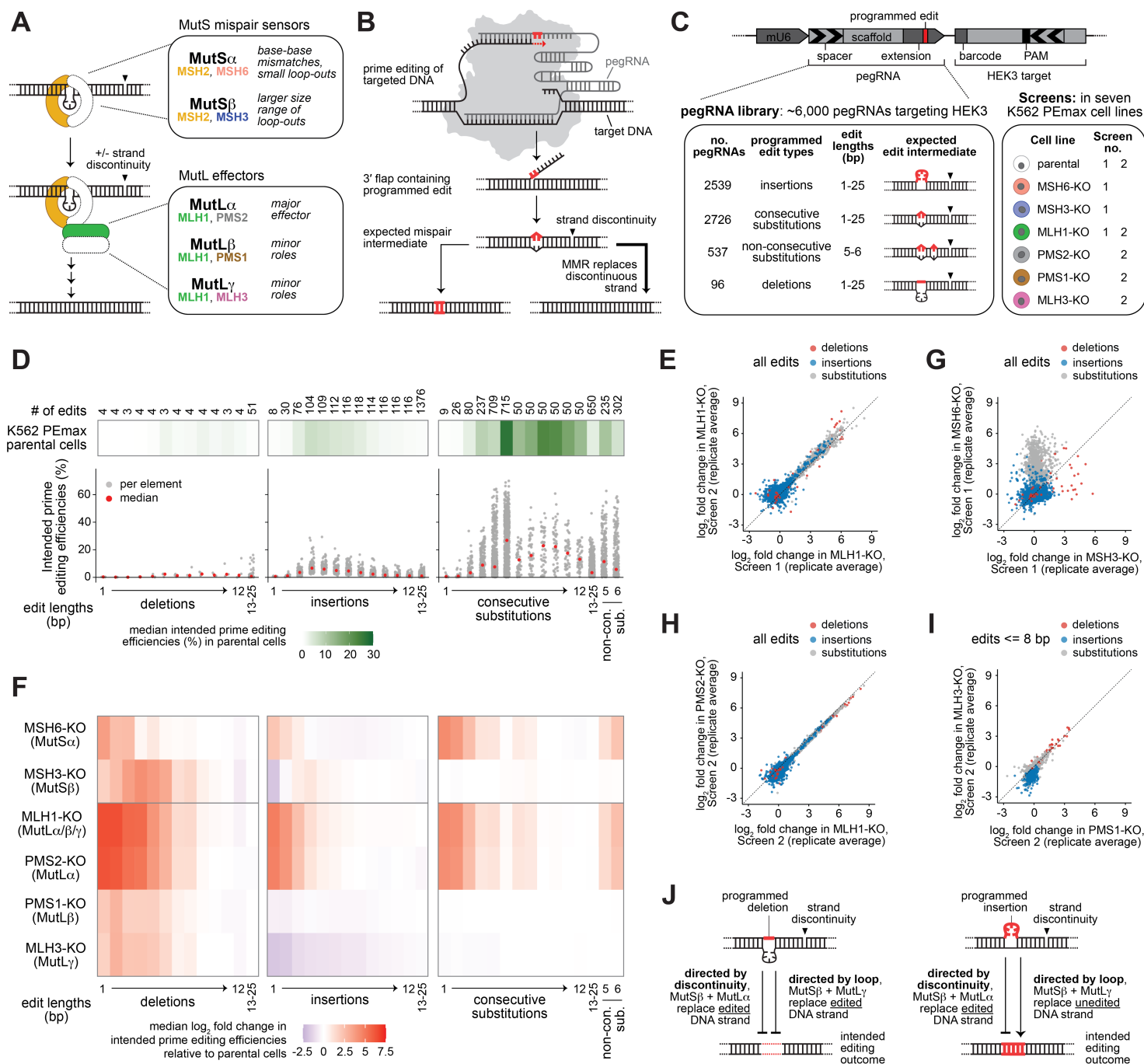


Figure 1. Pooled prime editing screens map mismatch repair gene functions across different prime edit types

(A) Schematic of DNA mismatch repair. Mismatches are sensed by MutS α / β sensor complexes, which recruit MutL α / β / γ and other effectors to complete downstream repair processes. Black triangles indicate a strand discontinuity in 1A-1C.

(B) Simple model of prime editing and programmed mispair rejection by DNA mismatch repair.

(C) Schematic depicting prime editing library using the HEK3 target site and related screens. Self-targeting lentiviral constructs (top) were used for delivery of the target sequence and pegRNAs programming many DNA mispairs to that site (bottom left). Screens were performed in K562 PEmax parental cells and derivative cell lines with core MMR genes knocked out (bottom right). Arrows within the depicted construct indicate directionality of the pegRNA spacer and matched protospacer within the target. Details on the other library and screen are included in Figure S1B.

(D) Heatmaps (top) show median intended prime editing efficiencies at the HEK3 target site in K562 PEmax parental cells grouped based on edit length (1-25 bp, with 13-25 bp edits collapsed into one group) and type (deletions, insertions, consecutive substitutions, and non-consecutive substitutions, which are abbreviated as non-con. sub.). The number of prime edits in each group is included. Dot plots (bottom) show individual (grey dots) and median (red dots) intended prime editing efficiencies within each group.

(E) Comparison of log₂ fold changes in intended prime editing efficiencies at the HEK3 target site in MLH1-KO cells between two different screens (HEK3 Screens 1 and 2). Each dot represents an element in the prime editing library, colored by edit type (red, deletions; blue, insertions; grey, substitutions). Black dashed line indicates y = x.

(F) Heatmaps show median log₂ fold changes in intended prime editing efficiencies for each knockout cell line at the HEK3 target site, stratified by prime edit length and type, as in D. Grey horizontal lines between heatmaps separate results from HEK3 Screens 1 and 2.

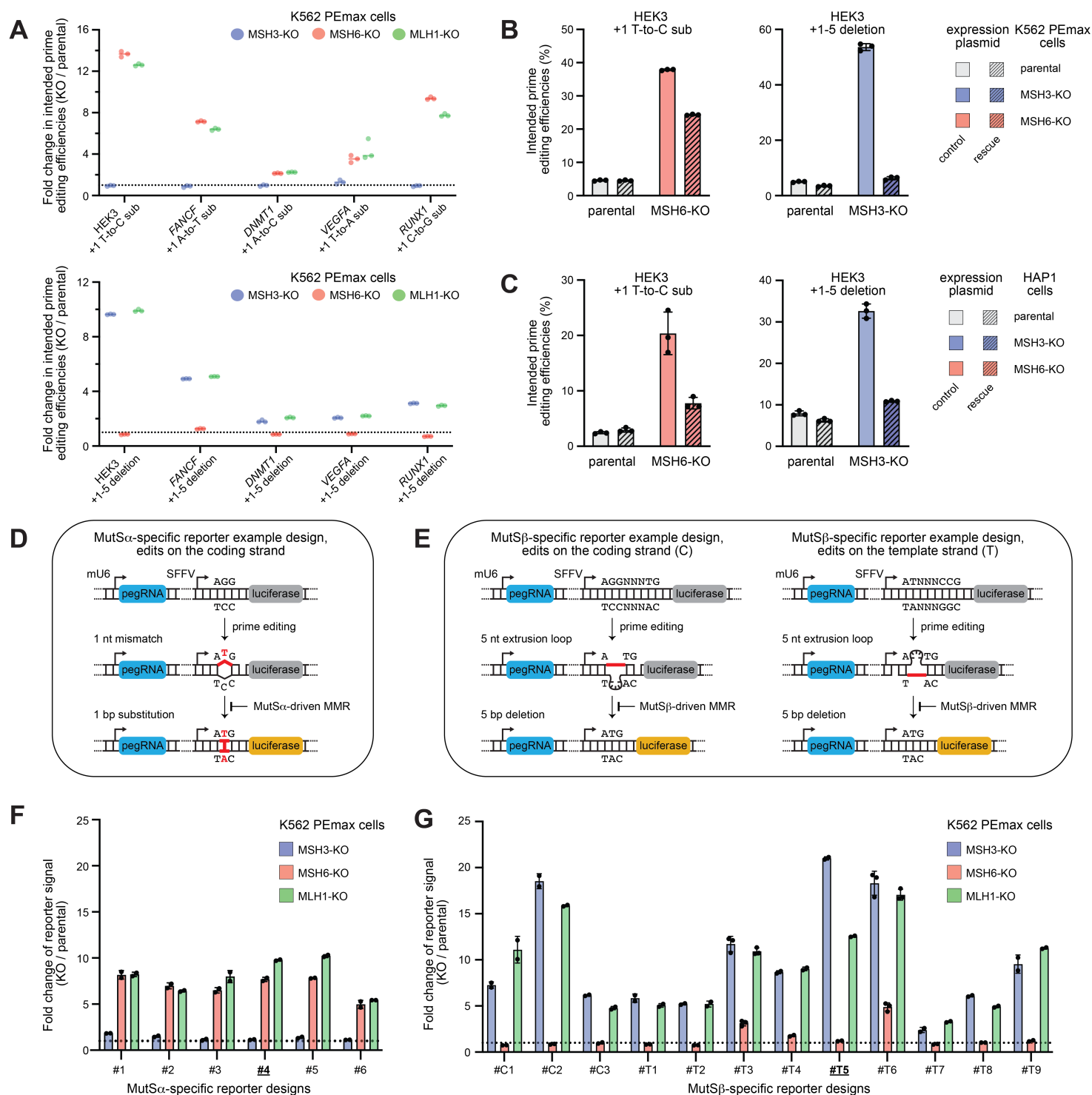
(G) Comparison of log₂ fold changes in intended prime editing efficiencies at the HEK3 target site in MSH3-KO cells (x-axis) to those from MSH6-KO cells (y-axis). Each dot represents an element in the prime editing library, colored by edit type, as in E. Black dashed line indicates y = x.

(H) Comparison of log₂ fold changes in intended prime editing efficiencies at the HEK3 target site in MLH1-KO cells (x-axis) to those from PMS2-KO cells (y-axis). Each dot represents an element in the prime editing library, colored by edit type, as in E. MLH1-KO data was from HEK3 Screen 2. Black dashed line indicates y = x.

(I) Comparison of log₂ fold changes in intended prime editing efficiencies at the HEK3 target site in PMS1-KO cells (x-axis) to those from MLH3-KO cells (y-axis). Each dot represents an element in the prime editing library, colored by edit type, as in E. Only edits $\leq$ 8 bp were included in this analysis. Black dashed line indicates y = x.

(J) Schematics of MMR pathways that can act on loop-out intermediates arising during deletion or insertion prime editing: discontinuity-directed MMR mediated by MutS β +MutL α , and loop-directed MMR mediated by MutS β +MutL γ . The former suppresses installation of both deletion and insertion prime edits, while the latter suppresses deletions and promotes insertions.

FIGURE 2



591 **Figure 2. MutS α / β -specific MMR reporters engineered with MutS resolving prime edits**

592 (A) Fold changes in intended prime editing efficiencies for 1 bp substitution edits (top) or 5 bp deletion edits
593 (bottom) at indicated genomic loci in indicated cell lines relative to parental cells. Groups of dots represent
594 data from n = 3 biological replicates. pegRNA plasmids were delivered by nucleofection.

595 (B) Intended prime editing efficiencies at the endogenous HEK3 target site in K562 PEmax parental cells or
596 indicated knockout cell lines, after nucleofection of pegRNA and rescue or control expression plasmids.
597 Bars represent means of n = 3 biological replicates and error bars indicate standard deviations.

598 (C) Intended prime editing efficiencies at the endogenous HEK3 target site in HAP1 parental cells or
599 indicated knockout cell lines, after nucleofection of pegRNA, PEmax editor, and rescue or control
600 expression plasmids. Bars represent means of n = 3 biological replicates and error bars indicate standard
601 deviations.

602 (D) Schematic of MutS α -specific reporters.

603 (E) Schematics of MutS β -specific reporters. Edits were designed to target either the coding strand (C, left)
604 or the template strand (T, right) of the luciferase indicator gene.

605 (F-G) Characterization of MutS α -specific reporters (F) and MutS β -specific reporters (G) in K562 PEmax
606 cells. Reporters were delivered to K562 PEmax parental cells or indicated knockout cell lines by lentiviral
607 transduction. Dots represent fold changes in luminescence, bars represent means of n = 2 or 3 technical
608 replicates, and error bars indicate standard deviations. Reporters selected for further characterization are
609 indicated **in bold and underlined** (#4 for MutS α - and #T5 for MutS β -specific reporters).

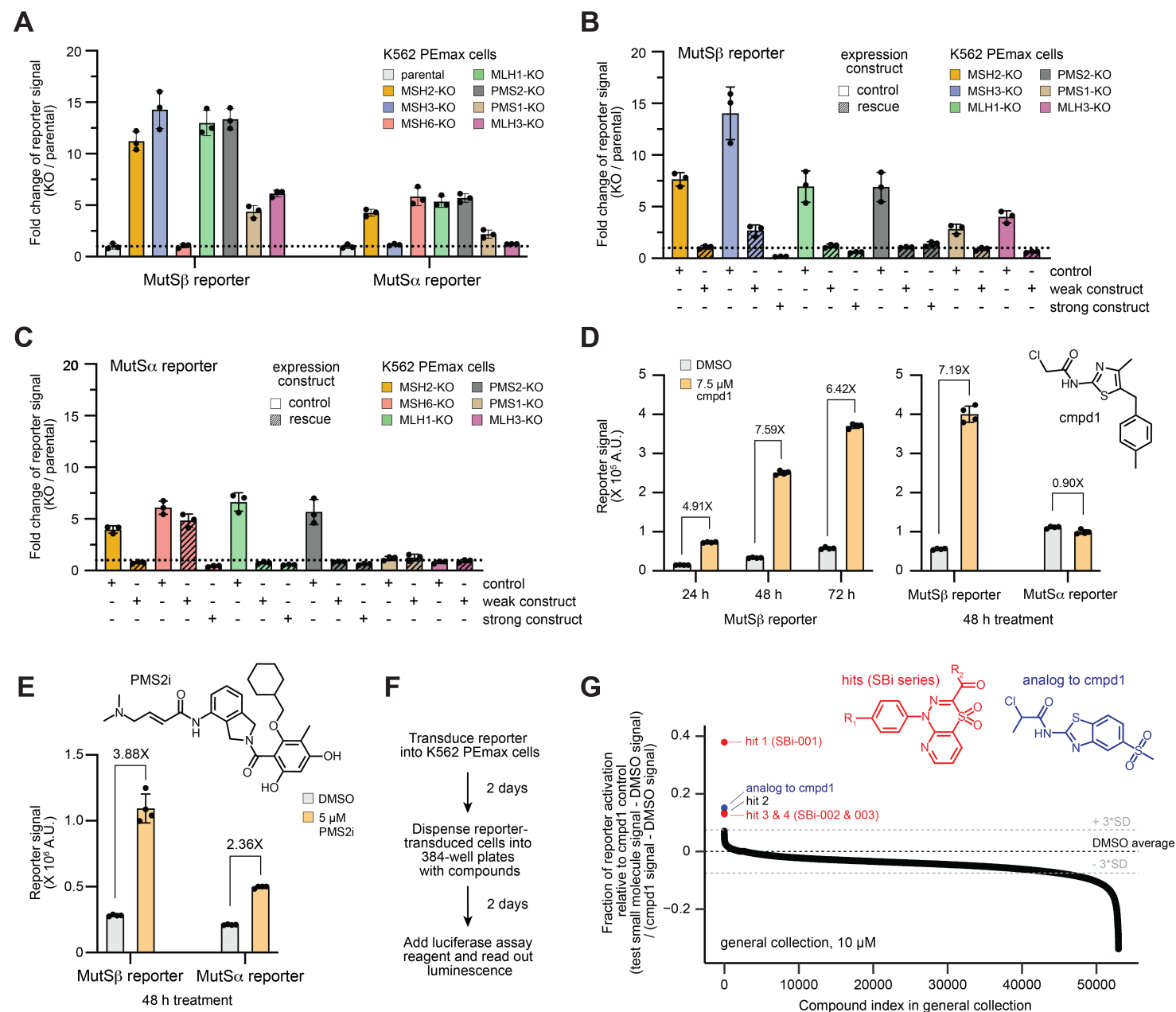


Figure 3. MutS α / β -specific reporters enable sensor-resolved measurement of MMR disruption and high-throughput small molecule screening

(A) Characterization of our primary MutS β - and MutS α -specific reporters in K562 PEmax parental cells and indicated knockout cell lines. Dots represent fold changes in luminescence, bars represent means of $n = 3$ biological replicates, and error bars indicate standard deviations.

(B) MMR rescue experiments with our primary MutS β -specific reporter. A weak or strong rescue construct or control (Methods) was delivered to the indicated K562 PEmax knockout cell line by lentivirus, with the rescue constructs matched to the knocked out gene. Dots represent fold changes in luminescence, bars represent means of $n = 3$ biological replicates, and error bars indicate standard deviations.

(C) As in B but for our primary MutS α -specific reporter.

(D) MutS β -specific reporter activation by cmpd1 in reporter-transduced K562 PEmax cells after drug treatments of varying durations (left). Comparison of the effect of cmpd1 on our primary MutS β - and MutS α -specific reporters in reporter-transduced K562 PEmax cells after 48 hours drug treatments (right). Dots represent raw luminescence signals, bars represent means of $n = 4$ biological replicates, and error bars indicate standard deviations. cmpd1 structure is depicted (inset).

(E) Comparison of the effect of PMS2i on our primary MutS β - and MutS α -specific reporters in reporter-transduced K562 PEmax cells after 48 hours drug treatments. Dots represent raw luminescence signals, bars represent means of $n = 4$ biological replicates, and error bars indicate standard deviations. PMS2i structure is depicted (inset).

(F) Description of our optimized MutS β -specific reporter assay used for high-throughput small molecule screening. Luminescence was read out using Nano-Glo luciferase assay.

(G) Results from screening the general collection of small molecules using our primary MutS β -specific reporter ($n = 1$). Dots represent fractions of reporter activation relative to cpmd1 controls, defined as background-subtracted luminescence from individual wells normalized to the mean of background-subtracted luminescence from on-plate cpmd1 controls, such that 0 represents average results from DMSO controls (black dashed line) and 1 represents average results from cpmd1 controls (Methods). Grey dashed lines indicate 3 standard deviations from the DMSO average ($\pm 3 \times \text{SD}$). Structures relevant to 4 out of 5 hit molecules (blue and red points, Table S8) are depicted (insets), with R_1 and R_2 representing different functional groups for SBI molecules (Figure 4A). We note that data for a few hits overlap on this plot, leaving only 3 such points visible.

FIGURE 4

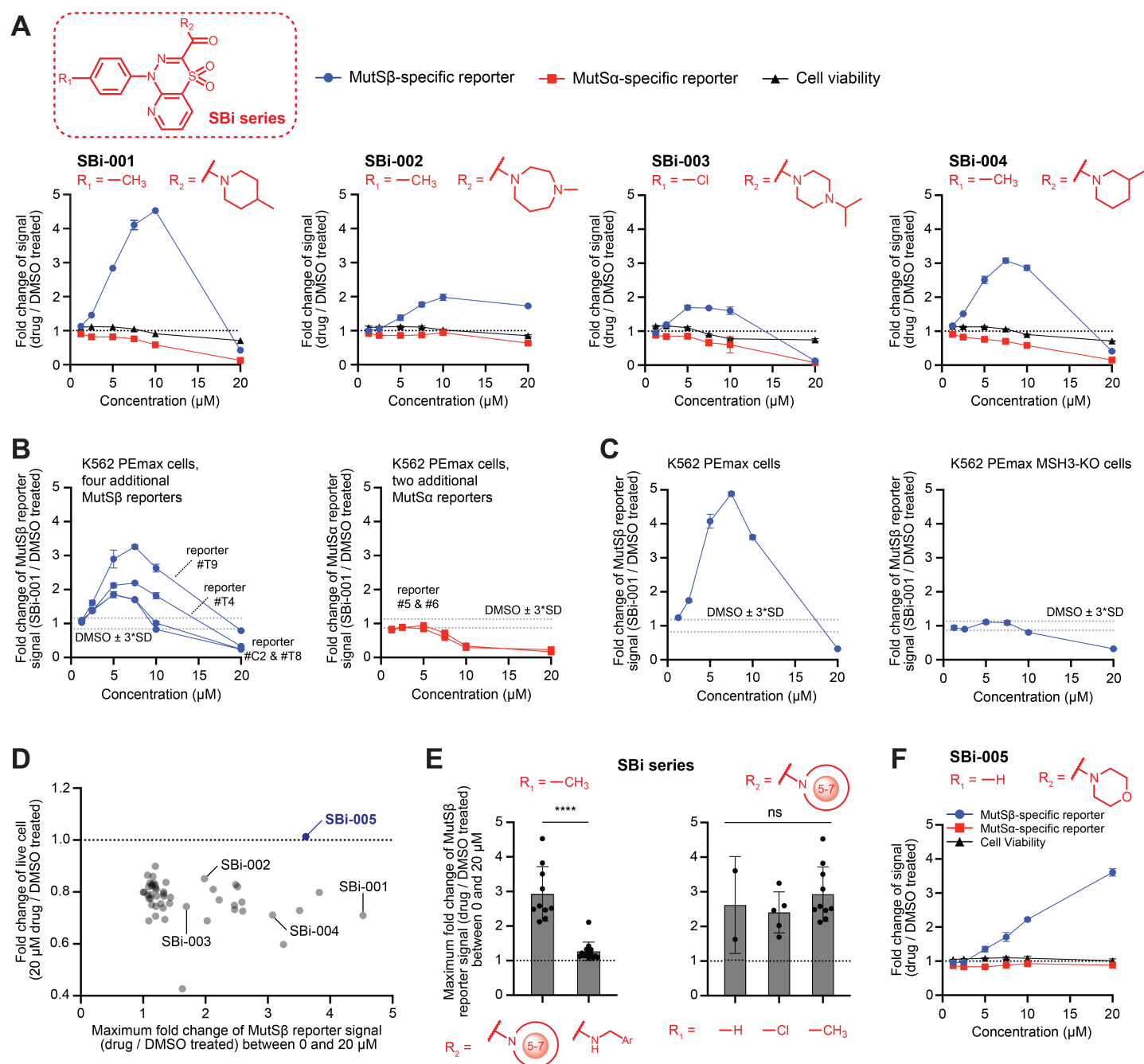


Figure 4. Cell-based validation and characterization of candidate MutSβ inhibitors

(A) Plots depict dose-dependent effects of SBI molecules on MutSβ-specific reporter activation (blue lines), MutSα-specific reporter activation (red lines), and cell viability (black lines, determined by Cell Titer-Glo) in K562 PEmax cells relative to the mean of DMSO controls. Dots represent fold changes in luminescence presented as means of $n = 3$ biological replicates and error bars indicate standard deviations. Core structure of SBI molecules included (top left). Relevant functional groups (R_1 and R_2) also included (plot insets).

(B) Plots depict dose-dependent effects of SBI-001 on activation of additional MutSβ-specific reporters (left, blue lines) and additional MutSα-specific reporters (right, red lines) in K562 PEmax cells relative to the mean of DMSO controls. Dots represent fold changes in luminescence presented as means of $n = 3$ biological replicates, error bars indicate standard deviations, and dashed lines indicate 3 standard deviations from the mean of DMSO controls.

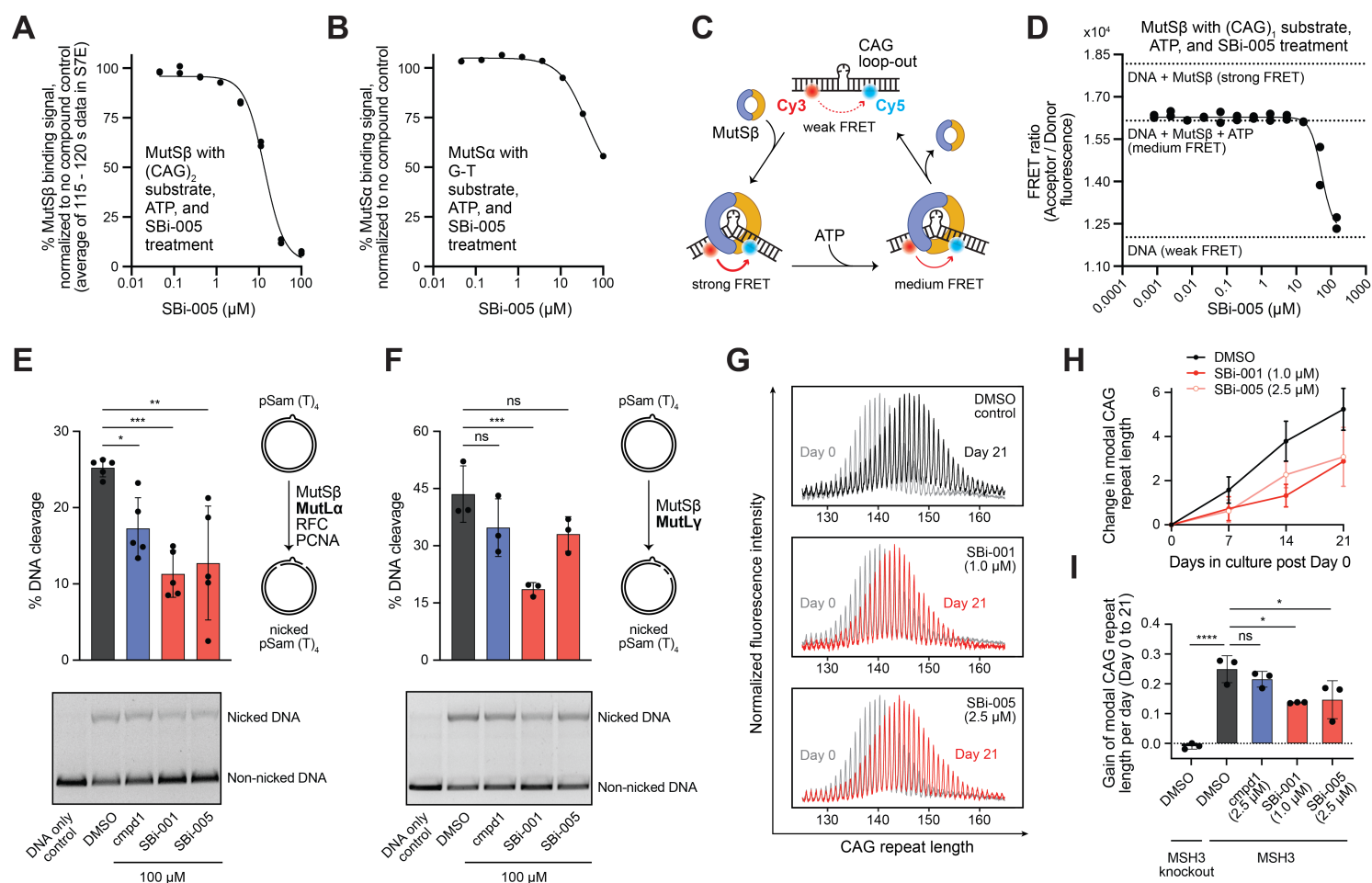
(C) Plots depict dose-dependent effects of SBI-001 on MutSβ-specific reporter activation in K562 PEmax parental (left) or MSH3-KO (right) cells relative to the mean of DMSO controls. Dots represent fold changes in luminescence presented as means of $n = 3$ biological replicates, error bars indicate standard deviations, and dashed lines indicate 3 standard deviations from the mean of DMSO controls.

(D) Testing SBI analogs for MutSβ-specific reporter activation (x-axis) and toxicity (y-axis). For activation of the MutSβ-specific reporter, each small molecule was tested at a range of concentrations (0 and 20 μM) and the maximal fold change relative to DMSO controls was reported. Since toxicity was monotonically related to drug concentrations, fold changes in live cells at 20 μM were reported (determined by Cell Titer-Glo). Dots represent means of $n = 3$ biological replicates.

(E) Comparisons of SBI analogs with a methyl group as R_1 and a different R_2 (left) or with a nitrogen-containing 5-7 member ring as R_2 and a different R_1 (right) in K562 PEmax cells. Relevant functional groups (R_1 and R_2) are depicted (insets), wherein the 5-7 circle indicates a nitrogen-containing 5-7 member ring for which the other atoms on the ring can be substituted at any position. Each dot represents the maximal fold change in activation of the MutSβ-specific reporter relative to DMSO controls for one analog, selected from a range of concentrations, as in D. Bars represent means across analog groups and error bars indicate standard deviations. Mann-Whitney U test (left; ***, $p < 0.0001$) and Kruskal-Wallis test (right; ns, not significant, $p \geq 0.05$) were used for statistical analysis.

(F) Plot depicts dose-dependent effects of SBI-005 on MutSβ-specific reporter activation (blue line), MutSα-specific reporter activation (red line), and cell viability (black line, determined by Cell Titer-Glo) in K562 PEmax cells relative to the mean of DMSO controls. Dots represent fold changes in luminescence presented as means of $n = 3$ biological replicates and error bars indicate standard deviations. SBI-005 functional groups (R_1 and R_2) depicted (plot insets). Data in D-F are from the same experiment.

FIGURE 5



674 **Figure 5. SBI molecules inhibit MutSβ binding to mispaired DNA and attenuate CAG expansion**

675 (A) MutSβ binding to a 38 bp DNA containing a (CAG)₂ loop-out at different concentrations of SBI-005
676 measured by SPR (as illustrated in Figure S7D). Dots represent means of mass response signals in
677 resonance units (from 115 - 120 s, as in Figure S7E) from n = 2 biological replicates normalized to controls
678 with no compound added. Black line represents a sigmoidal fit.

679 (B) As in A but for MutSα binding to a 38 bp DNA containing a G-T mismatch.

680 (C) Schematic of FRET assay. Three cycling states from reactions performed in the presence of ATP are
681 depicted: MutSβ-unbound DNA (top, weak FRET), MutSβ-bound inducing DNA bending (bottom left, strong
682 FRET), and ATP-dependent MutSβ sliding along and/or falling off from DNA (bottom right, medium FRET).
683 When both MutSβ and ATP are present, cycling among the three states can produce medium FRET.

684 (D) MutSβ binding to a 70 bp DNA containing a (CAG)₁ loop-out at different concentrations of SBI-005
685 measured in the presence of ATP by FRET assay. Dots represent FRET ratios from n = 2 biological
686 replicates and the solid black line represents a sigmoidal fit. Control signals measured without compound
687 treatment are indicated by black dashed lines.

688 (E) *In vitro* cleavage of a plasmid containing a (T)₄ loop-out by recombinant MutSβ-MutLα in the presence
689 of DMSO, cmpd1, SBI-001, or SBI-005. Dots in barplot (upper left) represent percentages of cleavage, bars
690 indicate means of n = 5 biological replicates, and error bars indicate standard deviations. Ordinary one-way
691 Analysis of Variance (ANOVA) was used for statistical analysis followed by Dunnett's multiple comparisons
692 test (***, adjusted *p* value < 0.001; **, adjusted *p* value < 0.01; *, adjusted *p* value < 0.05). A representative
693 gel image (lower left) and a schematic of the experiment (upper right) are also included.

694 (F) *In vitro* cleavage of a plasmid containing a (T)₄ loop-out by recombinant MutSβ-MutLγ in the presence
695 of DMSO, cmpd1, SBI-001, or SBI-005. Dots in barplot (upper left) represent percentages of cleavage, bars
696 indicate means of n = 3 biological replicates, and error bars indicate standard deviations. Ordinary one-way
697 ANOVA was used for statistical analysis followed by Dunnett's multiple comparisons test (***, adjusted *p*
698 value < 0.001; ns, not significant, adjusted *p* value ≥ 0.05). A representative gel image (lower left) and a
699 schematic of the experiment (upper right) are also included.

700 (G) Plots depict representative distributions of CAG repeat lengths from a cellular model of repeat
701 expansion (i.e., RPE-1 hTERT cells with exogenous mHTT exon 1) at initial sampling (defined as Day 0,
702 grey, two days post treatment with DMSO, SBI-001, or SBI-005) and 21 days later. Modal CAG repeat
703 lengths were ~130-140 at initial sampling.

704 (H) Changes in modal CAG repeat length observed in a cellular model of repeat expansion after 21 days of
705 growth in the presence of DMSO, SBI-001, and SBI-005 since Day 0 (vehicle and compounds replenished
706 every 3-4 days). Dots represent means of n = 3 biological replicates and error bars indicate standard
707 deviations.

708 (I) Per day changes in modal CAG repeat length observed in a cellular model of repeat expansion after 21
709 days of growth in the presence of DMSO, SBI-001, and SBI-005 since Day 0 (vehicle and compounds
710 replenished every 3-4 days). Dots represent per day averages, bars indicate means of n = 3 biological
711 replicates, and error bars indicate standard deviations. Ordinary one-way ANOVA was used for statistical
712 analysis followed by Tukey's multiple comparisons test (****, adjusted *p* value < 0.0001; *, adjusted *p* value
713 < 0.05; ns, not significant, adjusted *p* value ≥ 0.05). Data in G-I are from the same experiment.

722 EXPERIMENTAL MODEL AND SUBJECT DETAILS

723 K562 PEmax parental cells and MLH1-KO cells were previously described.^{56,65} Lenti-X 293T cells (632180)
 724 were from Takara. HeLa (CCL-2), U-2 OS (HTB-96), HCT 116 (CCL-247), and RPE-1 hTERT (CRL-4000)
 725 were from ATCC. HAP1 (parental, C631; MSH3-KO, HZGHC000656c001; MSH6-KO, HZGHC000653c003;
 726 MSH2-KO, HZGHC024799c006; MLH1-KO, HZGHC000343c021) were from Horizon. K562 PEmax cells
 727 were cultured in RPMI 1640 medium, HEPES (Gibco, 22400089) supplemented with 10% (v/v) fetal bovine
 728 serum (FBS; Gibco, A52567-01) and 1x Penicillin-Streptomycin-Glutamine (Gibco, 10378-016), also
 729 referred to as “complete RPMI medium”. K562 PEmax cells were grown between 4e5 to 1e6 cells / mL.
 730 Lenti-X 293T and HeLa cells were cultured in Dulbecco’s modified Eagle’s medium (Corning, 10-013-CV)
 731 supplemented with 10% (v/v) FBS (Gibco, A52567-01) and 1x Penicillin- Streptomycin (Corning,
 732 30-002-CI). HAP1 cells were cultured in Iscove’s Modified Dulbecco’s medium (IMDM; Cytiva HyClone™,
 733 SH30228.01) supplemented with 10% (v/v) FBS (Gibco, A52567-01) and 1x Penicillin-Streptomycin
 734 (Corning, 30-002-CI). HCT-116 and U-2 OS cells were cultured in McCoy’s 5A (Modified) Medium (Gibco,
 735 16600-082) supplemented with 10% (v/v) FBS (Gibco, A52567-01) and 1x Penicillin-Streptomycin
 736 (Corning, 30-002-CI). RPE-1 hTERT cells were cultured in DMEM/F12 medium with L-Glutamine and 15
 737 mM HEPES (Gibco, 11330-032) supplemented with 10% (v/v) FBS (Gibco, A52567-01), 1x
 738 Penicillin-Streptomycin (Corning, 30-002-CI), and 10 µg/mL hygromycin B (Goldbio, H270-2.5), with the
 739 following exception: In select experiments (Figures 5G-5I and S7H), we used a second RPE-1 hTERT cell
 740 line, called RPE-1 hTERT (CAG)₁₂₉, details of which can be found below (*Cell-based CAG repeat*
 741 *expansion assay*). For experiments herein, this cell line was cultured in Dulbecco-modified essential
 742 medium (DMEM), high glucose (Gibco, 11965092), supplemented with 10% (v/v) FBS (Gibco, A52567-01),
 743 1% (v/v) MEM NEAA (Gibco, 11140050), and 1x Penicillin-Streptomycin-Glutamine (Gibco, 10378-016). All
 744 cells used in this study were cultured at 37 °C in a humidified incubator with 5% CO₂ (v/v). Cell lines were
 745 authenticated by either suppliers or short tandem repeat profiling by ATCC and routinely tested negative for
 746 mycoplasma.

747 Small molecules used in this study include cmpd1 (purchased from Ambeed, A1110529), PMS2i
 748 (synthesized by WuXi Apptec Co., Ltd.), Ribavirin (purchased from MedChemExpress, HY-B0434),
 749 Levocarnitine propionate hydrochloride (purchased from MedChemExpress, HY-B0932), and Liothyronine
 750 (purchased from MedChemExpress, HY-A0070A). Other small molecules were ordered from ChemDiv Inc.
 751 Prior to cell-based experiments, small molecules were dissolved in dimethyl sulfoxide (DMSO, EMD
 752 Millipore, 317275-500ML). A summary of compounds used in this study can be found in Table S9.

753 METHOD DETAILS

754 **Expression plasmids.** Plasmids for transfection or electroporation of pegRNAs were cloned using
 755 pU6-pegRNA-GG-acceptor (Addgene, 132777). Briefly, pegRNA sequences (affixed to a modified mouse
 756 U6 promoter) were PCR amplified from gene fragments (Twist Bioscience), digested with BamHI-HF and
 757 XhoI, purified, and then assembled into BamHI-HF- and XhoI-digested backbone. Plasmids for transfection
 758 or electroporation of *MSH3*, *MSH6*, and *MSH2* were ordered from GenScript (OHu20501D, OHu25314D,
 759 and OHu20167D in pcDNA3.1-C-(k)DYK). In these initial plasmids, MMR cDNAs were tagged with FLAG
 760 (DYKDDDDK, C-terminal). We removed this tag from each vector, creating modified plasmids for expression
 761 of untagged MSH3, MSH6, and MSH2 (pYJ067, pYJ080, and pYJ189, respectively). Expression of MMR
 762 genes delivered by transfection or electroporation was driven by the CMV promoter. Finally, plasmids for
 763 expression of MMR proteins via lentiviral transduction were assembled using cDNAs from pYJ067, pYJ080,

pYJ189 or plasmids containing MutL subunits ordered from Twist Bioscience and FUGW lentiviral backbone (Addgene, #14883). For expression of PE7 proteins via lentiviral transduction, PE7 were PCR amplified from pCMV-PE7 (Addgene, #214812) and assembled with a pU6-sgRNA EF1Alpha-puro-T2A-BFP (Addgene #60955)-derived lentiviral backbone, IRES sequences (Addgene, #210590), and miRFP680 encoding sequences.⁹⁷ All plasmids were prepared with NucleoSpin Plasmid, Mini kit (Macherey-Nagel, 740588.250), ZymoPURE II Plasmid Midiprep Kit (Zymo Research, D4201), or ZymoPURE II Plasmid Maxiprep kit (Zymo Research, D4203).

Backbone plasmids for self-targeting libraries. Throughout this study, we used the PE2 approach without the nicking guide RNA.⁴⁶ We generated two plasmids to enable our self-targeting prime editing screens. These plasmids, which served as backbones for library cloning (*Cloning of HEK3 (IYJ003) and FANCF (IYJ004) self-targeting libraries*), have several key components: (1) a pegRNA expression cassette composed of a modified mouse U6 promoter, spacer for HEK3 or FANCF (5'-GGCCCAGACTGAGCACGTGA-3' for HEK3 and 5'-GGAATCCCTTCTGCAGCACC-3' for FANCF), and an SpCas9 guide RNA scaffold sequence (5'-GTTTTAGAGCTAGAAATAGCAAGTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC-3');⁹⁸ (2) a cassette designed to enable subsequent library construction composed of an Esp3I recognition sequence, mRFP1, a SbfI recognition sequence, an XhoI recognition sequence, the reverse complement of the target sequence from endogenous HEK3 or FANCF, where the pegRNAs can bind and install the edits, and an i7 annealing region (5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3'); and (3) a selection marker cassette composed of an EF1a promoter, puromycin resistance gene (PuroR), T2A, and TagBFP.

To build these two plasmids, we modified an existing pegRNA expression plasmid (pAC025, Addgene, 227597).⁵⁶ Notably, inclusion of mRFP1 in the second cassette enables differentiation of colonies with inserts that replace sequences between the Esp3I and XhoI recognition sequences (white colonies) from those without (red colonies) during library construction.

Design of HEK3 (IYJ003) and FANCF (IYJ004) self-targeting libraries. We constructed two self-targeting libraries (IYJ003 and IYJ004) for our prime editing screens (Table S1). These libraries contained the following key elements (5' to 3'; Figures 1C, S1A, and S1B): (1) a pegRNA expression cassette driven by a modified mouse U6 promoter, for expression of individual pegRNAs with a single spacer sequence, standard constant region of SpCas9 guide RNA scaffold,⁹⁸ and variable extension sequences (6000 per library), and (2) a single target sequence (90 nt of HEK3 for IYJ003 and FANCF for IYJ004) adjacent to a 12-mer barcode (generated by an R package called DNABarcodes⁹⁹). Each of these libraries specified 6000 edits: 2539 (HEK3) or 2542 (FANCF) insertions, 3263 (HEK3) or 3283 (FANCF) substitutions, and 96 (HEK3) or 97 (FANCF) deletions, and had 102 (HEK3) or 78 (FANCF) negative controls. Across the two libraries, edits had the same general patterning:

Insertions. In each library, we included insertions ranging from 1 bp to 25 bp with starting positions programmed directly adjacent to the predicted nick made by PEmax (+1 position) or six base pairs away (+7). All library insertions were designed from 30 "parent" sequences, containing 26 standard utility sequences (e.g., FLAG tag) and 4 repeat sequences related to human disease such as (CAG)_n. To generate a diverse but related set of insertions, we included each "parent" sequence, their reverse complement, and successive truncations thereof (i.e., full length sequence minus 1 from the 3' end repeatedly until only 1 nt from the 5' end remained) into each library in each start position. Then, to keep

library elements unique, we removed any repeated insertions. All insertions, within each of the two libraries, were designed with the same primer binding site sequences (PBS, 5'-CGTGCTCAGTCTG for HEK3 and 5'-GCTGCAGAAGGGAT for FANCF) and with reverse transcription templates (RTT) containing the same homology regions (5'-TGGAGGAAGCAGGGCTTCCTTTCTGCCATCA-3' for HEK3 +1 position, 5'-TGGAGGAAGCAGGGCTTCCTTTCTG-3' for HEK3 +7 position, 5'-GAGACCGCCAGAAGCTCGGAAAAGCGATCCAGGT-3' for FANCF +1 position, 5'-GAGACCGCCAGAAGCTCGGAAAAGCGAT-3' for FANCF +7 position).

Substitutions. We included substitutions of four types in each library: (1) All possible consecutive substitutions of 1-5 bp, starting directly adjacent to the predicted PEmax nick (+1 position) or two (+3) or four (+5) positions away. For each start site, these substitutions include 3 edits of 1 bp, 9 edits of 2 bp, 27 edits of 3 bp, 81 edits of 4 bp, and 243 edits of 5 bp for a total of 363 edits per start site and 1089 total edits; (2) All possible consecutive substitutions of 6 bp, starting at position +1 for a total of 729 such edits; (3) A select set of randomly generated consecutive substitutions ranging in length from 7-25 bp. These edits include 50 substitutions per length starting at position +1 for a total of 950 such edits; (4) A select set of non-consecutive substitutions, wherein the first substitution is at position +1 and the last is 4 or 5 bp away. These edits include 306 “complex substitutions” of 5 bp and 243 of 6 bp, respectively, for a total of 549 edits. Here too, we removed any repeated edits.

Deletions. In each library, we included deletions ranging from 1 bp to 25 bp with starting positions programmed directly adjacent to the predicted PEmax nick (+1 position) or two (+3) or four (+5) or six (+7) base pairs away. These deletions include 1 edit of each length and 25 edits per start site for a total of 100 such edits. Here too, we removed any repeat edits.

Negative controls. We also included two sets of negative controls in each library: (1) “non-annealing” pegRNAs containing scrambled extensions that should not anneal to the target site and (2) “no edit” pegRNAs with extensions that perfectly match the target site.

Cloning of HEK3 (IYJ003) and FANCF (IYJ004) self-targeting libraries. To construct our self-targeting libraries, pegRNA extension sequences for either HEK3 or *FANCF* containing edits described above were ordered as oligonucleotide pools (Twist Bioscience) and cloned into one of two plasmids (*Backbone plasmids for self-targeting libraries*). Oligonucleotides within these pools contained the following elements (5' to 3'): an annealing region for PCR amplification, an Esp3I recognition site, the pegRNA extension, a 7-mer polyT (transcription termination signal), a BamHI recognition site, an SbfI recognition site, a 12-mer barcode that uniquely corresponds to the paired pegRNA extension, a XhoI site, and another annealing region for PCR amplification. For cloning purposes, extensions and barcodes with Esp3I, BamHI, SbfI, or XhoI recognition sites were removed prior to ordering.

Cloning of each library was performed separately in several steps: (1) Sequences within each oligonucleotide pool were amplified by PCR using a pair of primers covalently labelled with biotin on their 5'-ends (oYJ379 and oYJ380; Table S4). PCR mixtures were assembled with NEBNext Ultra II Q5 Master Mix (M0544L), and reactions were performed at 98 °C for 30 s, then 12 cycles of [98 °C for 10 s, 62 °C for 20 s, and 72 °C for 30 s], followed by 72 °C for 2 min; (2) Amplified oligonucleotide pools were then purified using Oligo Clean & Concentrator (Zymo, D4060), according to manufacturer's instructions, and sequentially digested with FastDigest XhoI and Esp3I (Thermo Scientific, FD0694 and FD0454); (3) Digested adaptors (with biotin) were removed using Dynabeads™ MyOne™ Streptavidin T1 from the

847 Dynabeads™ Streptavidin Trial Kit (Invitrogen, 65801D) and remaining products (“inserts”) were purified
848 with Oligo Clean & Concentrator; (4) Purified inserts were next ligated into corresponding backbones
849 linearized by sequential digestion with FastDigest XhoI and Esp3I (Thermo Scientific, FD0694 and
850 FD0454) and purified with Zymoclean Gel DNA Recovery Kit (Zymo, D4001). Ligation mixtures were
851 assembled with ~150 ng of backbone, ~38 ng inserts, 1.5 μ L T4 DNA ligase (NEB, M0202T), and water to
852 20 μ L, and reactions were performed at 16 °C for 16 hrs; (5) Ligated product was then purified using
853 SPRIselect reagent (Beckman Coulter, B23318) with a 0.7X left side selection following manufacturer’s
854 instructions and electroporated into Endura DUOs electrocompetent cells (Biosearch Technologies,
855 60242-2) using a Gene Pulser Xcell electroporation system (BIO-RAD, 1652660); (6) After recovery at 37
856 °C, bacteria were grown in 200 mL LB broth supplemented with 70 μ g/mL carbenicillin for 24 hours at 30
857 °C; (7) Finally, plasmid libraries were extracted using ZymoPURE II Plasmid Maxiprep Kit (Zymo, D4203).

858 To assess library quality, pegRNA cassettes and target regions were amplified and sequenced.
859 Amplification was performed with a flanking 5’ primer
860 (5'-AATGATACGGCGACCACCGAGATCTACACGCACAAAAGGAACTCACCCT-3') and 3' indexing primer
861 (5'-CAAGCAGAAGACGGCATACGAGATNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCC-3')
862 using the following program: 98 °C for 30 s, then 12 cycles of [98 °C for 10 s, 62 °C for 20 s, and 72 °C for
863 30 s], followed by 72 °C for 2 min. PCR products were purified using SPRIselect, quantified by a Qubit 4
864 Fluorometer, analyzed with an Agilent 2100 Bioanalyzer, and sequenced using a NovaSeq 6000 SP
865 Reagent Kit v1.5 (300 cycles; Illumina, 20028400) with R1 = 190 nt (using oSL055,
866 5'-GTGTGTTTTGAGACTATAAGTATCCCTTGGAGAACCACCTTGTTG-3' as a custom sequencing primer),
867 I1 = 8 nt i7 index, and R2 = 135 nt. The libraries were well-balanced (Figure S1C).

868 **Lentivirus packaging and transduction.** Lentivirus was prepared and cells were transduced using
869 standard protocols. Materials for preparation were Lenti-X 293T cells, transfer plasmid(s), a packaging
870 plasmid mix containing pALD-Rev-A, pALD-GagPol-A, and pALD-VSV-G-A (Aldevron), TransIT-LT1 (Mirus,
871 MIR 2300), Opti-MEM I Reduced Serum Medium (Gibco, 31985070), and ViralBoost reagent (ALSTEM,
872 VB100). Virus-containing supernatant was typically stored at -80 °C. Prior to transduction, all cells were
873 isolated by centrifugation and resuspended into media with 6-8 μ g / mL polybrene (Santa Cruz
874 Biotechnology, sc-134220). K562 PEmax cells were transduced with centrifugation (“spinfection” at 1000 x
875 g for 2 hours at room temperature in 6-well plates), and after centrifugation, cells were resuspended by
876 swirling. Cells were incubated with virus for at least 48 hours before passaging or analysis. Transductions
877 performed for prime editing screens are described with additional details below (*Self-targeting prime editing*
878 *screens*).

879 **Cloning MutSa/ β -specific reporter constructs.** Reporter plasmid sequences are included in Table S7.
880 Our MutSa/ β -specific reporters have two key components: (1) a pegRNA expression cassette and (2) a
881 reporter gene cassette that can be induced by prime editing (Figures 2D and 2E). To build the pegRNA
882 expression cassette, we assembled a modified mouse U6 promoter (gene fragment, Twist Biosciences),
883 spacer sequence, modified SpCas9 guide RNA scaffold
884 (5'-GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAA
885 GTGGCACCAGAGTCGGTGC-3'),¹⁰⁰ relevant extension sequence, trimmed evopreq1 riboswitch
886 (5'-CGCGGTTCTATCTAGTTACGCGTTAAACCACTAGAA-3', only for MutSa reporters),¹⁰¹ and polyT
887 transcription termination signal (5' to 3'). To build the reporter gene cassette, we assembled an SFFV
888 promoter, the corresponding pegRNA target site, the NanoLuc luciferase gene¹⁰² from pTRE3G-NLucP
889 (Addgene, 162595), an IRES2 from pLenti-DsRed_IRES_eGFP (Addgene, 92194), and mCherry from

890 pHR-SFFV-KRAB-dCas9-P2A-mCherry (Addgene, 60954) (5' to 3'). We cloned these reporters in two
891 steps: First, we inserted all elements except the pegRNA sequence, SFFV promoter, and corresponding
892 pegRNA target into a lentiviral backbone using NEBuilder HiFi DNA assembly. The resulting plasmid
893 contained two BsmBI restriction digestion sites between the mU6 promoter and luciferase gene. Second,
894 we inserted pegRNA sequences, SFFV promoter, and targets (5' to 3') using gene fragments (Twist
895 Biosciences) between BsmBI sites with NEBuilder HiFi DNA assembly. Our best performing MutS β - and
896 MutS α -specific reporters, designs #T5 and #4, respectively (Figures 2D-2G), were cloned into two
897 backbones, pALD-lentiEGFP-A and FUGW (Addgene #14883; generating pYJ461 and pYJ464).

898 Our MutS α -specific reporters are selectively responsive to loss of MutS α -driven MMR; however, these
899 reporters generally demonstrate lower dynamic range than our MutS β -specific reporters (Figures 2F, 2G,
900 3A-3C, S4A, S4B, and S4I). This difference may be caused by higher baseline editing efficiencies in
901 MMR-proficient cell lines with the MutS α -specific reporters, which would impose a more limited ceiling on
902 the maximal fold change that could be observed from perturbing MMR. To improve the dynamic range of
903 our MutS α -specific reporters, future efforts may identify MutS α -specific edits with lower baseline
904 efficiencies or remove evopreQ1, which is known to enhance editing efficiencies, from those reporters.

905 **Generation of K562 cells with MMR gene disruptions and stable expression of PEmax.** K562 PEmax
906 parental cells⁶⁵ and the *MLH1* knockout derivative⁶⁶ (herein, K562 PEmax *MLH1*-KO) were previously
907 described. We made six additional MMR knockout cell lines in K562 cells for this study: K562 PEmax
908 *MSH3*-KO (*MSH3* disrupted), K562 PEmax *MSH6*-KO (*MSH6* disrupted), K562 PEmax *MSH2*-KO (*MSH2*
909 disrupted), K562 PEmax *PMS2*-KO (*PMS2* disrupted), K562 PEmax *PMS1*-KO (*PMS1* disrupted), and
910 K562 PEmax *MLH3*-KO (*MLH3* disrupted). To generate each of these cell lines, we electroporated K562
911 PEmax parental cells (~5.5e5) with 62 pmol of Alt-R S.p. Cas9 Nuclease V3 (Integrated DNA Technologies,
912 1081058) and 100 pmol of Alt-R CRISPR-Cas9 sgRNA (pre-complexed for 20 min at room temperature)
913 using SE Cell Line 96-well Nucleofector Kit (Lonza, V4SC-1096) with program FF-120, according to
914 manufacturer's instructions. Sequences of sgRNAs targeted to *MSH3* (Hs.Cas9.MSH3.1.AA), *MSH6*
915 (Hs.Cas9.MSH6.1.AF), *MSH2* (Hs.Cas9.MSH2.1.AA), *PMS2* (Hs.Cas9.PMS2.1.AB), *PMS1*
916 (Hs.Cas9.PMS1.1.AA), and *MLH3* (Hs.Cas9.MLH3.1.AA) are included in Table S2. Seven days later, single
917 cells were isolated by fluorescence-activated cell sorting (BD, FACSAria Fusion). Cells were expanded and
918 those with high percentage of EGFP expression (i.e., similar to the K562 PEmax parental cells) were
919 characterized by amplicon sequencing and western blot analysis (Figures S1D-S1N).

920 **Self-targeting prime editing screens.** Prime editing screens performed with our IYJ003 and IYJ004
921 self-targeting libraries were conducted in replicates. For each replicate, we transduced 100e6 cells, K562
922 PEmax parental or MMR knockout derivative cell lines at an MOI of 0.3 using lentivirus prepared as
923 described above (*Lentivirus packaging and transduction*). Prior to transduction, cells were isolated by
924 centrifugation and resuspended in 25 mL complete medium supplemented with 8 μ g/mL polybrene (final
925 density of 4e6 cells/mL). Transductions were performed with centrifugation in 6-well plates (16e6 cells/well,
926 1000 xg for 2 hours at room temperature). After centrifugation, cells were recovered in an 37 °C incubator
927 for 30 mins, and then resuspended to 5e5 cells/mL using complete RPMI medium to a final volume of 200
928 mL. Cells were cultured in a Multitron (Infors HT) at 37 °C, 85% humidity, 63 rpm, and with 5% (v/v) CO₂.
929 After 48 hours, 100 mL complete RPMI medium supplemented with 3 μ g/mL puromycin (Goldbio,
930 P-600-500) and 0.5% DMSO (1/200 v/v) was added to each culture. Cells were cultured for another 5 days,
931 splitting as needed, and collected for analysis 7 days after transduction. After collection, cells were washed
932 with PBS and stored at -80 °C until sample preparation.

To prepare screen samples for sequencing, cell pellets of ~100e6 cells were thawed (one pellet per replicate sample). Genomic DNA (gDNA) was then extracted using NucleoSpin Blood XL kit (Macherey-Nagel, 740950.50) according to manufacturer's instructions, and library elements were amplified for sequencing. We assembled 60-100 PCR reactions per replicate sample (100 μ L total per reaction) using NEBNext Ultra II Q5 Master Mix (New England Biolabs, M0544X). Each PCR reaction contained 10 μ g gDNA and 0.1 nmol of each of two primers: one complementary to part of the mU6 promoter (oSL059, 5'-AATGATACGGCGACCAACGAGATCTACACGCACAAAAGGAACTCACCT-3') and one complementary to an i7 annealing region on the construct and bearing a standard Illumina i7 index (oYJ023-oYJ034, exact sequences included in Table S4, 5'-CAAGCAGAAGACGGCATACGAGATNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTC-3'). PCR reactions were performed at 98 °C for 1 min, then 22 cycles of [98 °C for 10 s, 65 °C for 30 s, and 72 °C for 45 s], followed by 72 °C for 5 min. PCR products were purified using SPRIselect reagent, quantified by a Qubit 4 Fluorometer, pooled with equal molarity, and evaluated using an Agilent 2100 Bioanalyzer (Agilent, G2939B). Final samples were sequenced using a NovaSeq 6000 S1 300 cycles kit (Illumina, 20028317) with a pair-end read structure of R1 = 184 nt (using oSL055 as a custom sequencing primer), I1 = 8 nt, and R2 = 144 nt. R1 sequenced the pegRNAs, I1 sequenced i7 indexes, and R2 sequenced the target sites and barcodes.

Generation of RPE-1 hTERT cells with MMR gene disruptions. We made three MMR knockout cell lines in RPE-1 hTERT cells for this study: RPE-1 hTERT MSH3-KO (*MSH3* disrupted), RPE-1 hTERT MSH6-KO (*MSH6* disrupted), RPE-1 hTERT PEmax MSH2-KO (*MSH2* disrupted). To generate each of these cell lines, we electroporated RPE-1 hTERT parental cells (~2e5) with 76 pmol of Alt-R S.p. Cas9 Nuclease V3 (Integrated DNA Technologies, 1081058) and 125 pmol of Alt-R CRISPR-Cas9 sgRNA (pre-complexed for 20 min at room temperature) using SE Cell Line 96-well Nucleofector Kit (Lonza, V4SC-1096) with program EA-104, according to manufacturer's instructions. sgRNAs sequences are included in Table S2. Twelve days later, single cells were isolated by fluorescence-activated cell sorting (BD, FACSAria Fusion). Cells were expanded and characterized by amplicon sequencing and western blot analysis (Figures S3C and S3D).

Ectopic expression of MMR proteins in K562 PEmax cells, related to Figures 3B, 3C, and S4C-S4H. We designed rescue constructs that allowed us to express MMR proteins at qualitatively different levels. Our "strong construct" drove MMR gene expression from a UBC promoter. Our "weak construct" drove MMR gene expression from an IRES. For relevant experiments, K562 PEmax parental and MMR knockout cells were transduced with these constructs, each harboring a control gene (emiRFP/TagBFP) or an MMR gene, at an MOI < 0.4, using lentivirus prepared as described above (*Lentivirus packaging and transduction*). Transductions were performed in triplicate for each group. After centrifugation, cells were cultured for 48 hours. Cells were then selected by 3 μ g/mL puromycin and expanded daily and, after a week, transduced with MutSa/ β -specific reporters and assayed for reporter signals (*MutSa/ β -specific reporter assays performed with MMR knockout cells*).

Western blot analysis. To prepare samples, we isolated ~5e5 cells by centrifugation and lysed them in 50 μ L 1X Bolt™ LDS Sample Buffer (Invitrogen, B0007) and 50 mM Bolt™ Sample Reducing Agent (Invitrogen, B0009). Lysates were incubated at 70 °C for 10 min and then at 85 °C for 3 min with agitation (1250 rpm). To degrade DNA and RNA, lysates were supplemented with 5 mM MgCl₂ and 1.25 U μ L⁻¹ GENIUS™ Nuclease (Santa Cruz Biotech, sc-202391) and incubated at 37 °C for 30-45 min with agitation (1250 rpm). Samples were stored at -20 °C prior to further analysis. For SDS-PAGE, samples were thawed,

976 mixed, loaded onto Bolt™ Bis-Tris Plus Mini 4-12% 15-well gels (Invitrogen, NW04125BOX), and run for 40
977 min at 150 V using MES SDS running buffer (Invitrogen, B0002) with either PageRuler™ Plus Prestained
978 Protein Ladder (Thermo Scientific, 26619) or Precision Plus Protein Dual Color standards (Bio-Rad,
979 161-0374). Proteins were transferred onto nitrocellulose (Invitrogen, IB23001) using an iBlot™ 2 Dry
980 Blotting System (Invitrogen, IB21001) following manufacturer's instructions. Resulting membranes were
981 blocked in 5% non-fat dry milk in TBST (Santa Cruz Biotech, sc-2324) for 1 h at room temperature.
982 Proteins were detected using antibodies included in Table S3 using standard procedures.

983 **Cell electroporation, related to Figures 2A-2C, S3B, and S3E.** We electroporated plasmids into K562
984 PEmax cells using the SE Cell Line 96-well Nucleofector Kit (Lonza, V4SC-1096) and program FF-120,
985 following manufacturer's instructions (Figures 2A and 2B). Typically, ~5e5 cells were used per reaction with
986 250 ng pegRNA-expression plasmid and/or 500 ng expression plasmid for MMR proteins or control
987 proteins. Afterward, cells were allowed to recover for 10 min at 37 °C. Cells were then mixed with
988 pre-warmed, antibiotic-free RPMI 1640 medium supplemented with 10% FBS and cultured for 3 days prior
989 to collection. We electroporated plasmids into HAP1 cells using the SE Cell Line 96-well Nucleofector Kit
990 (Lonza, V4SC-1096) and program DZ-113, following manufacturer's instructions (Figures 2C and S3B).
991 Typically, ~3e5 cells were used per reaction with 150 ng pegRNA-expression plasmid, 450 ng prime
992 editor-expression plasmid, and/or 300 ng expression plasmid for MMR proteins or control proteins..
993 Afterward, cells were allowed to recover for 10 min at 37 °C. Cells were then mixed with pre-warmed,
994 antibiotic-free IMDM medium supplemented with 10% FBS and cultured for 3 days prior to collection. We
995 electroporated plasmids into RPE-1 hTERT cells using the SE Cell Line 96-well Nucleofector Kit (Lonza,
996 V4SC-1096) and program EA-104, following manufacturer's instructions (Figure S3E). In this case, ~1.5e5
997 cells were nucleofected with 500 ng prime editor-expression plasmid and 200 ng pegRNA-expression
998 plasmid. Afterward, cells were allowed to recover for 10 min at 37 °C. Cells were then mixed with
999 pre-warmed, antibiotic-free DMEM/F12 medium supplemented with 10% FBS and cultured for 3 days prior
1000 to collection. We note that rescue effects in Figures 2B and 2C were incomplete, especially in MSH6-KO
1001 cells, likely because MSH6 is highly abundant in many cell types,^{103,104} and ectopic expression may not
1002 have reached endogenous expression level. For all experiments listed here, three independent
1003 nucleofections were carried out for each experimental group (biological replicates n = 3).

1004 **Targeted amplicon sequencing, related to Figures 2A-2C, S1D-S1F, S1J-S1L, S3C, and S4B.** We
1005 performed amplicon sequencing to characterize K562 and RPE-1 hTERT knockout cell lines and quantify
1006 prime editing efficiencies. We first prepared cell lysates by plating ~150 µL cells into individual wells of
1007 96-well poly-D-lysine coated plates (Greiner Bio-One, 655946, for K562) or regular 96-well plates (Greiner
1008 Bio-One, 655180, for RPE-1 hTERT). Cells were incubated overnight (>12 h) to allow plate attachment,
1009 and then culture medium was replaced with 1X lysis buffer (~40-50 µL / well) composed of 10 mM Tris
1010 pH 8.0 (Gibco, AM9855G), 0.05% SDS (Invitrogen, 15553027), 25 µg / mL proteinase K (Invitrogen,
1011 AM2546), and nuclease-free water (Invitrogen, AM9939). Cells were lysed at 37 °C for 90 min, and
1012 resulting lysates were transferred into a PCR plate (Thermo Scientific, AB2400), incubated at 80 °C for 45
1013 min to heat-inactivate proteins, and used for PCR amplification immediately or after storage at -20 °C.
1014 Lysates for quantifying prime editing efficiencies were prepared with the following differences: (1) we
1015 isolated edited K562 cell populations by centrifugation (200 x g for 5 min at room temperature) and lysed
1016 cells with 1X lysis buffer (~60 µL / 500,000 cells), or (2) we removed medium from adherent cells and lysed
1017 directly with 1X lysis buffer (~40 µL).

1018 Amplicon sequencing libraries were prepared using two rounds of PCR. In the first round (PCR1), targeted
 1019 loci were amplified with primers containing an annealing region ($T_m = 56-64^\circ\text{C}$) and an adaptor sequence
 1020 (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3' or 5'-TGGAGTTCAGACGTGTGCTCTTCCGATCT
 1021 -3'; Table S4). PCR1 reaction mixtures were assembled with 0.1 nmol of each primer, 1.5 μL of cell lysate,
 1022 10 μL of NEBNext Ultra II Q5 Master Mix (New England Biolabs, M0544X), and water up to 20 μL . PCR1
 1023 reactions were performed at 98°C for 30 s, then 29-30 cycles of [98°C for 10 s, 61°C for 20 s, and 72°C
 1024 for 40 s], followed by 72°C for 2 min. PCR1 products were confirmed by agarose gel electrophoresis. In the
 1025 second round (PCR2), we extended the adapter region on each amplicon to add P5, i5, P7, and i7
 1026 sequences. PCR2 reaction mixtures were assembled with 0.1 nmol of each primer, 1.5 μL of PCR1
 1027 product, 10 μL Phusion U Multiplex PCR Master Mix (Thermo Scientific, F564S), and water up to 20 μL .
 1028 PCR2 reactions were performed at 98°C for 2 min, then 8 cycles of [98°C for 10 s, 61°C for 20 s, and
 1029 72°C for 40 s], followed by 72°C for 2 min. After amplification, PCR2 products were pooled by volume and
 1030 purified by gel extraction using Zymoclean Gel DNA Recovery Kit (Zymo Research, D4001). Final products
 1031 were then quantified by a Qubit 4 Fluorometer (Thermo Scientific, Q33238), examined by Agilent 2100
 1032 Bioanalyzer (Agilent, G2939B), and sequenced using a MiSeq Micro (Illumina, MS-103-1002) or Nano
 1033 (Illumina, MS-103-1001) kit with R1 = 300 nt, i7 = 8 nt, and i5 = 8 nt read structure.

1034 **MutSa/ β -specific reporter assays performed with MMR knockout cells, related to Figures 2F, 2G,**
 1035 **3A-3C, S4A, S4B, S4I, and S4J.** Reporter plasmids were packed into lentivirus and delivered into the cells
 1036 by lentiviral transduction (*Lentivirus packaging and transduction*). For reporter assays in K562 PEmax cell
 1037 lines, cells were transduced with the reporter lentivirus at an MOI of ~ 1.5 (Day 0). After 2-4 days of culture,
 1038 cells were mixed and analyzed by flow cytometry (Attune NxT) to determine the live cell number and
 1039 mCherry positive cell number, which indicates successful reporter transduction. Another aliquot of cells
 1040 (100 μL) were plated into 96-well white assay plates (Corning, 3917), mixed with equal volume (100 μL) of
 1041 reconstituted Nano-Glo Luciferase Assay reagents (Promega, N1150), and luminescence signals were
 1042 read out using a plate reader (Agilent BioTek Synergy HTX, S1LFA-SN or BioTek Cytation 1) using the
 1043 following parameters: excitation: plug; emission: hole; optics position: top; gain: between 80-120;
 1044 integration time: 0:00.10 MM:SS.ss; read height: 4.5 mm (for HTX) or 7 mm (for Cytation 1); read speed:
 1045 normal; delay: 100 msec. For experiments related to Figures 2F, 2G, and S4B, cells were assayed 72
 1046 hours post transduction. For experiments related to Figures 3A and S4I, cells were assayed 96 hours post
 1047 transduction. For experiments related to Figures 3B and 3C, cells were assayed 48 hours post
 1048 transduction. For experiments related to Figure S4A, cells were assayed at either 72 (x-axis data) or 96
 1049 (y-axis data) hours post-transduction. We note that data presented in Figures 3A, S4A, and S4B are from
 1050 the same experiments.

1051 For reporter assays in RPE-1 hTERT cells (Figure S4J), RPE-1 hTERT parental and derivative MMR
 1052 knockout cells were transduced with the reporter lentivirus at an MOI of ~ 1.5 and expanded for a week.
 1053 Cells ($\sim 2 \times 10^5$) were then electroporated with 500 ng pCMV-PEmax using the SE Cell Line 96-well
 1054 Nucleofector Kit (Lonza, V4SC-1096) with program EA-104, according to manufacturer's instructions. Cells
 1055 were trypsinized and resuspended in DMEM/F12 medium 2 days after electroporation, and then
 1056 resuspended cells were analyzed by flow cytometry as above for K562 cells. Another aliquot of cells (100
 1057 μL) were plated and assayed also as described above. For these experiments, we normalized the reporter
 1058 signals to the number of mCherry positive cells per sample before calculating the fold change.

1059 Replicates for these experiments were established as follows: for experiments related to Figures 2F and
 1060 2G, a single transduction was performed for each experimental group and cells were sampled

post-transduction and assayed in independent wells as replicates (n = 2 or 3). For experiments related to Figures 3A-3C, S4A, S4B, S4I, and S4J, three independent transductions were performed (n = 3) and each was independently assayed.

MutS β -specific reporter assays for siRNA or cDNA overexpression, related to Figures S4K and S4L. MutS β -specific reporter constructs were transduced into the K562 PEmax cells as described above (*MutSa/ β -specific reporter assays performed with MMR knockout cells*). Cells (5e5) were nucleofected with indicated amounts of siRNA (Figure S4K; Table S2) or 250 ng indicated plasmids (Figure S4L) 2 days after transduction using the SE Cell Line 96-well Nucleofector Kit (Lonza, V4SC-1096) with program FF-120 according to manufacturer's instructions. After 3 days of recovery, cells were analyzed by flow cytometry and NanoGlo luciferase reagents, as described above. For these experiments, we normalized the resulting luminescence to numbers of mCherry positive cells in each sample, as described above.

MutSa/ β -specific reporter assays performed with drug treatments in 96-well plates, related to Figures 3D, 3E, 4B, 4C, S5A, S5C-S5G, S5K, S5L, S6A-S6E. For reporter assays in K562 PEmax cells, cells were transduced with reporter constructs at an MOI of ~1.5 (Figures 3D, 3E, 4B, 4C, S5A, S5E, S5G, S5K, S5L, and S6A-S6C) or at indicated MOIs (Figure S5F). Transduced cells were treated with DMSO (control) or indicated small molecules at the indicated concentrations 2 days post transduction, cultured for another 2 days (Figures 3D right, 3E, 4B, 4C, S5A, S5E-S5G, S5K, S5L, and S6A-S6C) or for indicated amounts of time (Figure 3D left) and then analyzed with NanoGlo luciferase reagents. For the experiments depicted in Figure S5G left, different luciferase reagents were also used (Renilla-Glo for Renilla Luc and ONE-Glo for Firefly Luc, Promega, E2710 and E6110 respectively).

For reporter assays in HAP1, HeLa, U-2 OS, Lenti-X 293T, and HCT 116 cells (Figures S5C and S5D), cells were transduced with reporter constructs at an MOI of 1-2. Transduced cells were then cultured and split as needed for 4-5 days, and then transfected with pCMV-PEmax-EGFP or pCMV-PE7-EGFP, by TurboFectin 8.0 (OriGene, TF81001) for HAP1, Transit-HeLaMONSTER (mirus bio, MIR2904) for HeLa, Lipofectamine 2000 (Invitrogen, 11668027) for U-2 OS and Lenti-X 293T, and Lipofectamine LTX (Invitrogen, 15338030) for HCT 116, all following manufacturers' protocols. Transfected cells were treated with DMSO (control) or cmpd1 2 days later, plated into 96-well white assay plates with clear bottom (Corning, 3610), cultured for another 2 days, and then analyzed with NanoGlo luciferase reagents also as described above.

For reporter assays in RPE-1 hTERT cells (Figures S6D and S6E), cells were transduced with PE7 at an MOI of ~0.3. Cells were expanded after transduction and miRFP680 positive cells (with prime editor constructs) were sorted by fluorescence-activated cell sorting (BD, FACS Aria Fusion). Cells were expanded again and then transduced with the MutS β -specific reporter lentiviruses at an MOI of ~1.5. Transduced cells were treated with small molecules at indicated concentrations 2 days post transduction, cultured in 96-well white assay plates with clear bottoms (Corning, 3610) for another 2 days, and then analyzed by NanoGlo luciferase reagent, also as described above.

Replicates for these experiments were established as follows: Cells were mixed with indicated concentrations of drugs and split into separate wells of 96-well plates. Luminescence was examined separately for each well and resulting data were treated as individual replicates. Experiments in Figures 4B, 4C, S5A, S5K, S5L, and S6A-S6C were performed with 3 such replicates; experiments in Figures 3D, 3E, S5C-S5F, S5G left, S6D, and S6E were performed with 4 such replicates; and experiments Figure S5G

1102 (right, 96-well plates) were performed with 18 such replicates. We note that, for small molecule-related
1103 experiments, we did not normalize luminescence to numbers of mCherry positive cells in each sample.

1104 **Cell viability assays (standard throughput), related to Figure S5B.** Cells were treated with drugs and 2
1105 days later were transferred into wells of 96-well white assay plates (Corning, 3917), mixed with equal
1106 volume (100 μ L) of CellTiter-Glo® 2.0 Cell Viability Assay reagents (Promega G9241), and the
1107 luminescence signals were read out by a plate reader (Agilent BioTek Synergy HTX, S1LFA-SN or BioTek
1108 Cytation 1) as described above (*MutSa/ β -specific reporter assays performed with MMR knockout cells*).
1109 Reported cell viability measurements (i.e., fold changes of live cells of drug treated versus DMSO controls)
1110 were calculated by normalizing resulting luminescence signals to the mean of luminescence from
1111 DMSO-treated cells.

1112 **Composition of small molecule libraries.** Small molecule libraries were provided by the Princeton
1113 University Small Molecule Screening Center. Altogether, we screened two libraries: the bioactive collection,
1114 which consisted of 5,923 small molecules, and the general collection, which consisted of 47,780 individual
1115 small molecules and 4,480 pooled mixture with 4 stereoisomers in each well. The total number of small
1116 molecules screened was therefore $5,923 + 47,780 + 4,480 \times 4 = 71,623$ small molecules.

1117 **Primary small molecule screen, related to Figures 3F, 3G, and S5H-S5J. Preparing small molecules.** To
1118 aliquot compounds for screening, small molecules from each collection were added individually to wells of
1119 white, flat-bottom 384-well plates (Thermo Scientific, 164610) at either 0.02 μ L or 0.04 μ L per well (from
1120 10 mM stock solutions dissolved in DMSO) by an Echo acoustic liquid handler (Beckman Coulter, Labcyte
1121 Echo 550). On each plate, we left the first and last two columns (1, 2, 23, 24) empty and, immediately
1122 before screening, added DMSO or cmpd1 to those wells (DMSO, 1 and 23; cmpd1, 2 and 24; 0.02 μ L per
1123 well). Bioactive collection small molecules were screened at a final concentration of 10 and 20 μ M, while
1124 general collection compounds were screened at a final concentration of 10 μ M. As an internal standard, we
1125 ran a control plate with only titrated cmpd1 alongside each batch of plates in the screen.

1126 **Preparing and processing cells.** K562 PEmax cells were transduced with the MutS β -specific reporter at an
1127 MOI of ~ 1.5 . After 48 hours of recovery, transduced cells were analyzed by flow cytometry (AttuneNXT,
1128 Invitrogen, A24858) and transduction at the targeted MOI was confirmed by mCherry expression. Cells
1129 were then diluted to 5×10^5 cells/mL in fresh RPMI complete medium and dispensed into
1130 compound-containing 384-well plates at 20 μ L/well using a Multidrop Combi reagent dispenser (Thermo
1131 Scientific, 5840300) with small tube metal tip dispensing cassette (Thermo Scientific, 24073295). Cells
1132 were allowed to grow for 48 hours and were then assayed for luminescence using Nano-Glo Luciferase
1133 Assay reagents (Promega, N1150) dispensed at 20 μ L per well. Plates rested at room temperature for 5-10
1134 min and then luminescence was read out using a plate reader (Agilent BioTek Synergy HTX, S1LFA-SN)
1135 with the following parameters: excitation, plug; emission, hole; optics position, top; gain, 120; integration
1136 time, 0:00.20 MM:SS.ss; read height, 1.00 mm. Resulting data were exported into .xlsx files using the
1137 Agilent Gen5 software for processing and analysis (QUANTIFICATION AND STATISTICAL ANALYSIS).
1138 We note that, for the primary small molecule screen, only one replicate was performed.

1139 **Secondary screening of small molecules and individual testing of analogs and resynthesized hits in**
1140 **high-throughput.** Hits from the bioactive collection, resynthesized small molecules, and compound
1141 analogs were examined in high-throughput with procedures similar to those used in primary screening, with
1142 the following exceptions: (1) compounds in these tests were evaluated across a range of concentrations

and with biological replicates, and (2) compounds in these tests were examined with both MutS α - and MutS β -specific reporters. Hits from the bioactive collection (Table S8) were dispensed in volumes of 2.5 nL, 5 nL, 10 nL, 20 nL, 40 nL, and 80 nL into individual wells of white, flat-bottom 384-well plates (2 wells per condition) from 10 mM stock solutions dissolved in DMSO to reach final concentrations of 1.25 μ M to 40 μ M after subsequent addition of 20 μ L of cells. Other small molecules tested in 384-well plates (Figures 4A, 4D-4F, and S6F-S6K) were dispensed in volumes of 2.5 nL, 5 nL, 10 nL, 15 nL, 20 nL, and 40 nL into individual wells of white, flat-bottom 384-well plates (3 wells per condition) from 10 mM stock solutions dissolved in DMSO to reach final concentrations of 1.25 μ M to 20 μ M after subsequent addition of 20 μ L of cells. When testing small molecules in this format, cells without reporters were often similarly processed and used to examine compound toxicities. Cell viability measurements were made by adding 20 μ L of CellTiter-Glo 2.0 Cell Viability Assay reagent (Promega, G9242) to the cells and quantifying luminescence on a plate reader (Agilent BioTek Synergy HTX, S1LFA-SN) with the following parameters: excitation, plug; emission, hole; optics position, top; gain, 120; integration time, 0:00.20 MM:SS.ss; read height, 1.00 mm. Other procedures were the same as described above (*Primary small molecule screen*).

Reverse transcription-quantitative polymerase chain reaction (RT-qPCR). For RT-qPCR experiments, total RNA was extracted using Monarch® Total RNA Miniprep Kit (NEB, T2010S) and quantified by NanoDrop. Total RNA was then used for RT-qPCR reactions using Power SYBR™ Green RNA-to-CT™ 1-Step Kit (Applied Biosystems, 4389986), following manufacturer's instructions. Each 20 μ L reaction contained 50 ng of RNA, 8 pmol of each of the forward and reverse primers (Table S4), 0.16 μ L of 125X RT enzyme mix, and 10 μ L of Power SYBR Green Master Mix. Reactions were then run on a ViiA 7 Real-Time PCR System (Applied Biosystem) with the following program: 48 °C for 30 min, 95 °C for 10 min, then 40 cycles of [95 °C for 15 s, 65 °C for 1 min], followed by 95 °C for 15 s, 60 °C for 1 min, increase to 95 °C at 0.05 °C/s, and 95 °C for 15 s. $\Delta\Delta$ Ct values¹⁰⁵ were calculated and converted to percentage of expression level. For these experiments, biological replicates represent three independent samples of compound-treated cells, and for each of those replicates, 4 technical replicates using the same reaction master mix were aliquoted into 4 wells in the assay plate were also included.

DNA intercalation assay, related to Figure S7C. SPR measurements were performed on a Cytiva Biacore 8K+ (Danaher Life Sciences LLC, USA) using Series S Streptavidin-coated biochips (Danaher Life Sciences LLC, USA). Measurements were performed in the following assay buffer: 50 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl₂, 0.5 mM TCEP, 500 μ M ATP, 0.05% (v/v) Tween-20, 0.05% BSA, and 1% DMSO. After standard chip conditioning according to the manufacturer's instructions, 1.5 μ M 5'-biotinylated 38 bp (CAG)₂ DNA was captured in the assay buffer to a target level of 2500 RU. For this experiment, the sample and chip compartment were set to T = 20 °C. All compounds were diluted in 100% DMSO using a Biomek FXP (Beckman Coulter Inc, USA) from 10 mM stocks in a 3-fold dilution series to 1.52 μ M. The compound dilution series was transferred to 384-micro well plates as 1:100 dilutions in the assay buffer using an Integra Viaflo 384 system (INTEGRA Biosciences GmbH, Germany). Test compounds were then injected into the SPR flowcell. Contact duration of association was 120 s, dissociation duration 120 s at a flow rate of 30 μ L/min with a data collection rate of 10 Hz. To calculate the binding signals, a 5 seconds window, 115-120 s after initial injection of compounds, was taken to calculate an average signal (background signal from DNA binding to the flow cell subtracted) for each small molecule at each concentration.

SPR assay, related to Figures 5A, 5B, S7D, and S7E. SPR measurements were performed with equipment described above (*DNA intercalation assay*). Measurements were performed in the following

assay buffer: 50 mM HEPES, pH 7.5, 150 mM KCl, 5 mM MgCl₂, 0.5 mM TCEP, 100 μM ATP, 0.05% (v/v) Tween-20. After standard chip conditioning according to the manufacturer's instructions, 0.35 nM 5'-biotinylated 38 bp (CAG)₂ DNA was captured in running buffer without DMSO to a target level of 25 RU for MutSβ measurements; analogously for MutSα measurements, 5'-biotinylated 38 bp G-T mismatch DNA was captured. For these experiments, the sample and chip compartment were set to T = 20°C. All compounds were diluted in 100% DMSO using a Biomek FXP (Beckman Coulter Inc, USA) from 10 mM stocks in a 3-fold dilution series to 4.57 μM. The compound dilution series were transferred to 384-micro well plates as 1:100 dilutions in assay buffer with either 15 nM MutSβ or 50 nM MutSα using an Integra Viaflo 384 system (INTEGRA Biosciences GmbH, Germany). After 30 minutes of incubation, MutSα or MutSβ only control and mixtures with the compound dilution series were injected according to the SPR schema. Contact duration of association was 120 s (0-120 s, schematized in Figure S7D), dissociation duration 240 s at a flow rate of 30 μL/min with a data collection rate of 10 Hz. After dissociation, surface regeneration was performed by injection of 0.1% SDS for 30 seconds. To calculate the binding signals, the "analyte binding late" region of the data (5 seconds window, 115-120 s, schematized in Figure S7E) was taken to calculate an average signal (background signal from DNA binding to the flow cell subtracted) for each sample at each concentration. The calculated binding signals were then normalized to the results from no compound control (0 μM).

FRET assay, related to Figures 5C, 5D, and S7F. Assay was performed with or without ATP as described⁸¹ with some modifications: The assay buffer was modified to contain 1 mM TCEP instead of DTT, and 100 nM protein and 100 μM ATP was used. The assay was carried out in a 384-well low volume round bottom black plate (Corning 4514) in 8 μL total assay volume. Assay materials were prepared as follows: First, a compound serial dilution plate was generated via Biomek FX Laboratory Automation Workstation with 12-point, 3-fold dilution starting from 20 mM top concentration of each compound in DMSO in a 384-well plate (Greiner). Second, MutSβ was added at 4 μL to the assay plate with matrix pipette (Thermo scientific). Third, compounds were dipped at 58 nL from the compound serial dilution plate to the assay plate with CyBio pintool such that the maximum concentration of compounds on the assay plate was 145 μM. For MutSβ+DNA or free DNA controls without compound, an equivalent amount of DMSO was dipped instead of compounds from the compound plate. Samples were then mixed briefly and incubated for 30 min at room temperature. Meanwhile probe plate containing Cy3-Cy5 labeled 70 bp (CAG)₁ loop DNA or such DNA plus ATP was prepared and 4 μL of probe (labeled DNA +/- ATP) was transferred via Integra Viaflo 384 from the probe plate to the assay plate such that the final labeled DNA concentration in the assay was 5 nM. Samples were mixed and fluorescence of labeled DNA was recorded in donor (Ex 545 nm/Em 590 nm) and acceptor channels (Ex 545 nm/Em 680 nm) in kinetic mode over 70 min. The FRET ratio (fluorescence in acceptor channel / fluorescence in donor channel) was recorded for steady state.

Plasmid-based mismatch repair assay, related to Figures 5E, 5F, and S7G. Preparations of the proteins and plasmid DNA substrates were described previously.⁶⁹ Reactions were carried out in a final volume of 15 μL in buffer containing 25 mM Tris-acetate (pH 7.5), 0.1 mg/mL recombinant albumin (New England Biolabs, B9200S), 5 mM MgCl₂, and 2 mM ATP. Each reaction contained 80 ng of the 3.2 kb plasmid DNA substrate pSam (T)₄. No additional NaCl or KCl was added. The only salt present came from the recombinant protein storage buffer, giving a final NaCl concentration of approximately 20 mM. Protein storage buffer was added as needed to ensure the same final NaCl concentration in all samples. Where indicated, inhibitors were used at 100 μM. MutSβ (75 nM) or FAN1 (75 nM) was added on ice, and reactions were then incubated with the indicated inhibitor for 5 min at 37 °C. Next, 75 nM MutLγ or MutLα, 75 nM RFC, and 75 nM PCNA (where indicated) were added, and the reactions were incubated for a further 45 min at 37 °C. Reactions were stopped by adding 5 μL STOP solution (150 mM EDTA, 2% SDS,

1231 30% glycerol, and 0.01% bromophenol blue) together with 1 μ L Proteinase K (Roche, 18 mg/mL), followed
 1232 by incubation at 50 °C for 1 h. Products were separated on 1% agarose gels containing GelRed (Biotium)
 1233 in TAE buffer (40 mM Tris, 20 mM acetate, and 1 mM EDTA) for 90 min at 6 V/cm. Gels were imaged using
 1234 a Quantum gel imager (Vilber). Band intensities were quantified with ImageJ2 (NIH) and data were plotted
 1235 and analyzed in Prism 10 (GraphPad).

1236 **Cell-based CAG repeat expansion assay, related to Figures 5G-5I.** RPE-1 hTERT (CAG)₁₂₉ cells carry a
 1237 stably integrated mHTT gene exon 1 (originally introduced with lentiviral transduction). We plated these
 1238 cells at 2.5e5 cells per well in 6 well plates and added the indicated compounds at indicated
 1239 concentrations. After 48 hours, initial samples (Day 0) were collected. Cells were then subcultured for 21
 1240 additional days, by splitting 1:4 every 3 to 4 days. Compounds were replenished during reseeding.
 1241 Samples were taken every 7 days. Genomic DNA was then extracted from each sample using the DNeasy
 1242 Blood & Tissue kit (Qiagen, 69509), and CAG repeat tracts within the integrated mHTT gene exon 1 were
 1243 amplified by PCR using 100 ng of genomic DNA, AmpliTaq Gold 360 Master Mix (Applied Biosystems,
 1244 4398876), a 6-FAM labelled forward primer (5'-CCTTCGAGTCCCTCAAGTCCTT-3'), and a reverse primer
 1245 (5'-CGGCTGAGGCAGCAGCGGCTGT-3') with the following program: 95 °C for 10 min, then 35 cycles of
 1246 [95 °C for 30 s, 58 °C for 30 s, and 72 °C for 90 s], followed by 72 °C for 7 min. PCR products were
 1247 denatured in Hi-Di™ Formamide (Thermo Scientific, 4440753) and analyzed by capillary electrophoresis
 1248 using the 3730xl DNA Analyzer with MapMarker® 1000 labelled with ROX (Eurogentec,
 1249 MW-0195-80ROX). Raw data from capillary electrophoresis was analysed using the R package “trace:
 1250 Tandem Repeat Analysis by Capillary Electrophoresis”
 1251 (<https://cran.r-project.org/web/packages/trace/index.html>). Briefly, ladder standard peak was aligned with
 1252 CAG repeat tract peaks and modal repeat length was called by determining the modal peak above 20% of
 1253 the maximum peak signal. Per day changes in modal CAG repeat lengths were calculated as follows:
 1254 (modal CAG repeats length on Day 21 - modal CAG repeats length on Day 0) / 21.

1255 **Cell cycle analysis by Click EdU assay, related to Figure S7H.** RPE-1 hTERT (CAG)₁₂₉ cells were
 1256 treated with compounds as described above (*Cell-based CAG repeat expansion assay*) and collected for
 1257 cell cycle analysis on Day 14 (drug treatment for 16 days total). Cells were incubated with EdU for 1 hour at
 1258 37 °C. Cells were then fixed and EdU was labeled using the Click-iT™ EdU Alexa Fluor™ 647 Flow
 1259 Cytometry Assay Kit (Thermo Scientific, C10419) following manufacturer's instructions. DNA was labeled
 1260 by incubating resulting EdU-labeled cells in 2 μ g/mL propidium iodide and 100 μ g/mL RNase A. For
 1261 analysis, 20,000 events were recorded per sample by flow cytometry (Cytotflex, Beckman technology).

1262 QUANTIFICATION AND STATISTICAL ANALYSIS

1263 **Analysis of targeted amplicon sequencing.** To quantify prime editing outcomes from targeted amplicon
 1264 sequencing, we used a previously reported approach.⁶⁵ Briefly, we aligned sequencing reads to
 1265 corresponding reference sequences (Table S4) with CRISPResso2¹⁰⁶ (2.2.11) using HDR batch mode. We
 1266 defined the intended outcome as the expected HDR amplicon sequence (“--expected_hdr_amplicon_seq”) and used the following settings: the minimum average read quality was 30 (“--min_average_read_quality 30”), and the quantification window size was 20 (“--quantification_window_size 20”) and centered at the
 1269 pegRNA nick (“--quantification_window_center -3”). We discarded reads with indels (i.e., unwanted editing
 1270 outcomes, “--discard_indel_reads”). We defined intended editing efficiencies as the fractions of reads
 1271 containing intended edits without indels and calculated those efficiencies as follows: (number of
 1272 non-discarded HDR-aligned reads) / (number of reads that aligned all amplicons) X 100%.

Analysis of self-targeting prime editing libraries (IYJ003 and IYJ004). Plasmid libraries were cloned and sequenced as described above (*Cloning of HEK3 (IYJ003) and FANCF (IYJ004) self-targeting libraries*). Reads from validation sequencing were expected to contain the following:

Read 1 (5' to 3'): spacer (20 nt), sgRNA scaffold (76 nt), extension (variable for each edit), and polyT (5'-TTTTTTTCC-3') sequences.

Read 2 (5' to 3'): target site (90 nt), barcode (12 nt) flanked by XhoI and BamHI recognition sequences.

Quality assessment was performed on read pairs with average Phred ≥ 30 in two multipart steps. **Step 1: Assigning reads to library elements.** (1) Spacer and scaffold regions (20 + 76 nt) were searched on read 1 to match the expected sequences, regardless of the starting position. Reads with no spacer-scaffold region found were discarded. (2) The polyT sequence was searched downstream of the spacer-scaffold region on read 1, and we extracted the sequence in between the spacer-scaffold and polyT. Reads for which the extracted sequence perfectly matched a designed extension sequence were retained and assigned to a library element while others were discarded. (3) For reads assigned to a library element, the corresponding barcode was extracted from between the XhoI and BamHI recognition sequences on read 2. Extracted barcodes were compared to expected sequences, and reads for which the extracted barcode matched the expected barcode up to one Levenshtein distance were retained. **Step 2: Verification of target site sequences.** (1) After confidently assigning reads to library elements, we verified the target site sequences (90 nt) by extracting them from the beginning of read 2. Reads were retained if the extracted target site sequence differed from the expected target site sequence by fewer than 4 bp.

To assess the uniformity of these libraries, we calculated a Gini coefficient using read counts for each element, after removing any recombined elements.

Analysis of self-targeting prime editing screens (performed with IYJ003 and IYJ004). For analysis of screen data, reads were assigned to library elements essentially as described for Step 1 in *Analysis of self-targeting prime editing libraries (IYJ003 and IYJ004)*. For this analysis, a separate FASTQ file was created for each library element containing read 2 sequences retained after Step 1. These files were processed using Cutadapt¹⁰⁷ to trim off the restriction enzyme recognition sequences and barcode, leaving the 90 nt target sequence. Reads shorter than 10 bp after trimming were discarded. Trimmed reads were then analyzed using CRISPResso2 (version 2.3.1) with the following parameters:

```
--exclude_bp_from_left 0 --exclude_bp_from_right 0 --amplicon_seq [reference/unedited sequence] -g [protospacer sequence] -e [HDR/edited sequence] -q 30 -w 0 --discard_indel_reads.
```

This analysis used trimmed read 2 sequences for alignment and classification into three categories: reference, HDR, or ambiguous. A minimum average Phred quality score of 30 was required, and reads falling below that score were removed entirely from any downstream calculation. The window size was set to 0 so that each read would be classified based on its full sequence. Reads containing indels were not considered HDR events but were retained in the total read count (defined below as the number of reads that aligned to all amplicons). Note that in this analysis, we use “indel(s)” to refer to unintended insertions and/or deletions introduced during prime editing, library preparation, or sequencing, which is different from intended insertion or deletion prime edits.

As enough sequencing depth is required for accurate calculation of editing efficiencies, we discarded all elements with less than 500 reads aligning to all amplicons.

CRISPResso2 generated a file for each library element with two rows containing read counts aligned to the (1) reference sequence and (2) the HDR sequence (which in this case, is the reference sequence containing the intended prime edit). Columns of interest were Reads_aligned_all_amplicons (total number of reads that are able to align to either the reference sequence or HDR sequence), Modified (number of reads aligned to the reference or HDR sequence with mismatches), Unmodified (number of reads aligned to the reference or HDR sequence without mismatches), and Discarded (number of reads aligned to the reference or HDR sequence that contain indels, with or without mismatches). Editing efficiencies were calculated as: (number of unmodified HDR-aligned reads) / (number of reads that aligned all amplicons).

During analysis, we considered that library elements may obtain edits from unassigned pegRNAs when multiple constructs enter the same cell, and moreover, recombination events and sequencing errors may produce similar effects. We therefore estimated a “background editing rate” for each library element that was designed to make an edit (i.e., “editing elements”) using elements that were designed not to introduce any edit (i.e., “negative control elements”). That is, we determined the frequencies at which modifications from the former set of elements unexpectedly appeared across the latter. To calculate these frequencies, we first applied CRISPResso2 across our negative controls. We performed this analysis without the -e parameter, as no HDR sequences were expected for those elements. Similar to editing elements, negative controls with fewer than 500 reads after library assignment were discarded.

For each negative control element, CRISPResso2 generated a table summarizing all read patterns observed among assigned read 2 sequences, recorded the number of read counts per pattern, and indicated read patterns containing an indel (denoted as DISCARDED_Reference). We then further assigned these read patterns into categories using custom scripts as follows: (1) NaN_mutation: a read pattern containing substitutions that does not perfectly match any editing element in the library; (2) NaN_indel: a read pattern containing indels that does not perfectly match any editing element in the library; (3) Unedited: a read pattern that perfectly matches the reference sequence; (4) Assigned_others: a read pattern that perfectly matches the HDR sequence corresponding to an editing element in the library.

We used the Assigned_others category to calculate the frequencies at which modifications from each editing element appeared across all negative controls, and defined that as the background editing rate for the corresponding element. As a stringent filter, we excluded any library element whose intended editing efficiency, in any replicate of screen data from any of the MMR knockout cell lines, was < 3 standard deviations above the mean of its background editing rate. This filter removed 28 elements from downstream analysis for HEK3 Screens 1 and 2 and removed 8 elements for the FANCF Screen.

Analysis of the small molecule screen. Screen data were compressed from raw read .xlsx files into a summary .xlsx file for statistical analysis. For each plate a Z'-factor was calculated based on the means (μ) and standard deviations (σ) of raw luminescence signals from negative (DMSO) and positive (cmpd1) control wells:

$$Z' = 1 - [3 * (\sigma_p + \sigma_n) / |\mu_p - \mu_n|]$$

1349 where p and n denote positive and negative controls, respectively.

1350 These Z'-factors were used to assess and maintain data quality throughout our screen. That is, plates with
1351 a $Z' < 0.2$ (rounded to one decimal place) were rerun until a $Z' \geq 0.2$ was obtained. Out of 204 plates in
1352 our screen, 9 (4.41%) had an initial $Z' < 0.2$, and the first rerun of all 9 substantially improved the Z' to
1353 satisfy the $Z' \geq 0.2$ requirement. For clarity, only plates with a $Z' \geq 0.2$ were used for downstream
1354 analysis. For technical reasons, positive controls were not added to the first plate in the screen (index
1355 number 1), and we therefore excluded that plate from downstream analysis.

1356 We next used a relative reporter activation metric to score compounds in our screen, herein called “fraction
1357 of reporter activation relative to cmpd1 control”. This metric was calculated from raw luminescence signals
1358 on a per-plate basis as:

1359
$$\text{fraction of reporter activation relative to cmpd1 control} = (C - \mu_n) / (\mu_p - \mu_n)$$

1360 where C represents the raw luminescence of the test compound and p and n denote positive and negative
1361 controls, respectively.

1362 This metric directly compares the activities of each test compound to signals from our positive and negative
1363 controls, such that a score of 0 would mean that cells treated with test compound produced signals
1364 equivalent to those treated with vehicle (DMSO), and a score of 1 would mean that cells treated with test
1365 compound produced signals equivalent to those treated with cmpd1. We also calculated this metric for
1366 each well of DMSO-treated cells and determined hits as compounds that scored > 3 standard deviations
1367 from the mean of all such wells, across all plates in the screen. Compounds scoring above this threshold
1368 were then reevaluated in secondary testing using resynthesized compounds or new stocks from vendors
1369 (*Secondary screening of small molecules and individual testing of resynthesized hits and analogs in*
1370 *high-throughput*).

1371 **Analysis from secondary testing of small molecules in high-throughput.** For determining
1372 MutS β -specific reporter effects in these experiments, we used procedures essentially the same as
1373 described above (*Analysis of the small molecule screen*). In this analysis, a small molecule significantly
1374 activated the MutS β -specific reporter if the corresponding fraction of reporter activation relative to cmpd1
1375 control was ≥ 3 standard deviations from the mean of DMSO controls (0.075) at either 10 or 20 μM (Table
1376 S8). For determining MutS α -specific reporter effects, we used an alternative metric, as we lacked a positive
1377 control for calculating a relative metric. Instead, we calculated a Z-score:

1378
$$Z = (S - \mu_n) / \sigma_n$$

1379 where S represents the raw luminescence of the test compound and n denotes negative control.

1380 In this analysis, a small molecule significantly activated the MutS α -specific reporter if the corresponding
1381 Z-score was ≥ 3 (Table S8).

1387 REFERENCES

- 1388 1. Walker, F.O. (2007). Huntington's disease. *Lancet* 369, 218–228.
1389 [https://doi.org/10.1016/S0140-6736\(07\)60111-1](https://doi.org/10.1016/S0140-6736(07)60111-1).
- 1390 2. The Huntington's Disease Collaborative Research Group (1993). A novel gene containing a trinucleotide
1391 repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971–983.
1392 [https://doi.org/10.1016/0092-8674\(93\)90585-e](https://doi.org/10.1016/0092-8674(93)90585-e).
- 1393 3. Hannan, A.J. (2018). Tandem repeats mediating genetic plasticity in health and disease. *Nat. Rev.*
1394 *Genet.* 19, 286–298. <https://doi.org/10.1038/nrg.2017.115>.
- 1395 4. Tabrizi, S.J., Schobel, S., Gantman, E.C., Mansbach, A., Borowsky, B., Konstantinova, P., Mestre, T.A.,
1396 Panagoulas, J., Ross, C.A., Zauderer, M., et al. (2022). A biological classification of Huntington's
1397 disease: the Integrated Staging System. *Lancet Neurol.* 21, 632–644.
1398 [https://doi.org/10.1016/S1474-4422\(22\)00120-X](https://doi.org/10.1016/S1474-4422(22)00120-X).
- 1399 5. Handsaker, R.E., Kashin, S., Reed, N.M., Tan, S., Lee, W.-S., McDonald, T.M., Morris, K., Kamitaki, N.,
1400 Mullally, C.D., Morakabati, N.R., et al. (2025). Long somatic DNA-repeat expansion drives
1401 neurodegeneration in Huntington's disease. *Cell* 188, 623–639.e19.
1402 <https://doi.org/10.1016/j.cell.2024.11.038>.
- 1403 6. Mätlik, K., Baffuto, M., Kus, L., Deshmukh, A.L., Davis, D.A., Paul, M.R., Carroll, T.S., Caron, M.-C.,
1404 Masson, J.-Y., Pearson, C.E., et al. (2024). Cell-type-specific CAG repeat expansions and toxicity of
1405 mutant Huntingtin in human striatum and cerebellum. *Nat. Genet.* 56, 383–394.
1406 <https://doi.org/10.1038/s41588-024-01653-6>.
- 1407 7. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium (2025). Genetic modifiers of somatic
1408 expansion and clinical phenotypes in Huntington's disease highlight shared and tissue-specific effects.
1409 *Nat. Genet.* 57, 1426–1436. <https://doi.org/10.1038/s41588-025-02191-5>.
- 1410 8. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium (2015). Identification of Genetic
1411 Factors that Modify Clinical Onset of Huntington's Disease. *Cell* 162, 516–526.
1412 <https://doi.org/10.1016/j.cell.2015.07.003>.
- 1413 9. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. Electronic address:
1414 gusella@helix.mgh.harvard.edu and Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium
1415 (2019). CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell*
1416 178, 887–900.e14. <https://doi.org/10.1016/j.cell.2019.06.036>.
- 1417 10. Namuli, K.L., Slike, A.N., Hollebeke, M.A., and Wright, G.E.B. (2025). Genomic characterization of
1418 Huntington's disease genetic modifiers informs drug target tractability. *Brain Commun.* 7, fcae418.
1419 <https://doi.org/10.1093/braincomms/fcae418>.
- 1420 11. Laabs, B.-H., Klein, C., Pozojevic, J., Domingo, A., Brüggemann, N., Grütz, K., Rosales, R.L.,
1421 Jamora, R.D., Saranza, G., Diesta, C.C.E., et al. (2021). Identifying genetic modifiers of age-associated
1422 penetrance in X-linked dystonia-parkinsonism. *Nat. Commun.* 12, 3216.
1423 <https://doi.org/10.1038/s41467-021-23491-4>.
- 1424 12. Moss, D.J.H., Pardiñas, A.F., Langbehn, D., Lo, K., Leavitt, B.R., Roos, R., Durr, A., Mead, S.,
1425 TRACK-HD investigators, REGISTRY investigators, et al. (2017). Identification of genetic variants
1426 associated with Huntington's disease progression: a genome-wide association study. *Lancet Neurol.* 16,
1427 701–711. [https://doi.org/10.1016/S1474-4422\(17\)30161-8](https://doi.org/10.1016/S1474-4422(17)30161-8).

13. Lee, J.-M., Huang, Y., Orth, M., Gillis, T., Siciliano, J., Hong, E., Mysore, J.S., Lucente, D., Wheeler, V.C., Seong, I.S., et al. (2022). Genetic modifiers of Huntington disease differentially influence motor and cognitive domains. *Am. J. Hum. Genet.* *109*, 885–899. <https://doi.org/10.1016/j.ajhg.2022.03.004>.
14. Lin, Y., Dion, V., and Wilson, J.H. (2006). Transcription promotes contraction of CAG repeat tracts in human cells. *Nat. Struct. Mol. Biol.* *13*, 179–180. <https://doi.org/10.1038/nsmb1042>.
15. Goold, R., Hamilton, J., Menneteau, T., Flower, M., Bunting, E.L., Aldous, S.G., Porro, A., Vicente, J.R., Allen, N.D., Wilkinson, H., et al. (2021). FAN1 controls mismatch repair complex assembly via MLH1 retention to stabilize CAG repeat expansion in Huntington's disease. *Cell Rep.* *36*, 109649. <https://doi.org/10.1016/j.celrep.2021.109649>.
16. Bunting, E.L., Donaldson, J., Cumming, S.A., Olive, J., Broom, E., Miclăuș, M., Hamilton, J., Tegtmeyer, M., Zhao, H.T., Brenton, J., et al. (2025). Antisense oligonucleotide-mediated MSH3 suppression reduces somatic CAG repeat expansion in Huntington's disease iPSC-derived striatal neurons. *Sci. Transl. Med.* *17*, eadn4600. <https://doi.org/10.1126/scitranslmed.adn4600>.
17. Ferguson, R., Goold, R., Coupland, L., Flower, M., and Tabrizi, S.J. (2024). Therapeutic validation of MMR-associated genetic modifiers in a human ex vivo model of Huntington disease. *Am. J. Hum. Genet.* *111*, 1165–1183. <https://doi.org/10.1016/j.ajhg.2024.04.015>.
18. Mouro Pinto, R., Murtha, R., Azevedo, A., Douglas, C., Kovalenko, M., Ulloa, J., Crescenti, S., Burch, Z., Oliver, E., Kesavan, M., et al. (2025). In vivo CRISPR-Cas9 genome editing in mice identifies genetic modifiers of somatic CAG repeat instability in Huntington's disease. *Nat. Genet.* *57*, 314–322. <https://doi.org/10.1038/s41588-024-02054-5>.
19. Wang, N., Zhang, S., Langfelder, P., Ramanathan, L., Gao, F., Plascencia, M., Vaca, R., Gu, X., Deng, L., Dionisio, L.E., et al. (2025). Distinct mismatch-repair complex genes set neuronal CAG-repeat expansion rate to drive selective pathogenesis in HD mice. *Cell* *188*, 1524–1544.e22. <https://doi.org/10.1016/j.cell.2025.01.031>.
20. Aldous, S.G., Smith, E.J., Landles, C., Osborne, G.F., Cañibano-Pico, M., Nita, I.M., Phillips, J., Zhang, Y., Jin, B., Hirst, M.B., et al. (2024). A CAG repeat threshold for therapeutics targeting somatic instability in Huntington's disease. *Brain J. Neurol.* *147*, 1784–1798. <https://doi.org/10.1093/brain/awae063>.
21. O'Reilly, D., Belgrad, J., Ferguson, C., Summers, A., Sapp, E., McHugh, C., Mathews, E., Boudi, A., Buchwald, J., Ly, S., et al. (2023). Di-valent siRNA-mediated silencing of MSH3 blocks somatic repeat expansion in mouse models of Huntington's disease. *Mol. Ther. J. Am. Soc. Gene Ther.* *31*, 1661–1674. <https://doi.org/10.1016/j.ymthe.2023.05.006>.
22. Dragileva, E., Hendricks, A., Teed, A., Gillis, T., Lopez, E.T., Friedberg, E.C., Kucherlapati, R., Edelmann, W., Lunetta, K.L., MacDonald, M.E., et al. (2009). Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. *Neurobiol. Dis.* *33*, 37–47. <https://doi.org/10.1016/j.nbd.2008.09.014>.
23. Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol.* *7*, 335–346. <https://doi.org/10.1038/nrm1907>.
24. Kunkel, T.A., and Erie, D.A. (2015). Eukaryotic Mismatch Repair in Relation to DNA Replication. *Annu. Rev. Genet.* *49*, 291–313. <https://doi.org/10.1146/annurev-genet-112414-054722>.
25. Owen, B.A.L., Yang, Z., Lai, M., Gajec, M., Badger, J.D., Hayes, J.J., Edelmann, W., Kucherlapati, R., Wilson, T.M., and McMurray, C.T. (2005). (CAG)(n)-hairpin DNA binds to Msh2-Msh3 and changes properties of mismatch recognition. *Nat. Struct. Mol. Biol.* *12*, 663–670. <https://doi.org/10.1038/nsmb965>.

- 1471 26. Guo, J., Gu, L., Leffak, M., and Li, G.-M. (2016). MutS β promotes trinucleotide repeat expansion by
1472 recruiting DNA polymerase β to nascent (CAG) $_n$ or (CTG) $_n$ hairpins for error-prone DNA synthesis. *Cell*
1473 *Res.* 26, 775–786. <https://doi.org/10.1038/cr.2016.66>.
- 1474 27. Li, J., Wang, H., and Yang, W. (2023). Tandem MutS β binding to long extruded DNA trinucleotide
1475 repeats underpins pathogenic expansions. *bioRxiv*, 2023.12.12.571350.
1476 <https://doi.org/10.1101/2023.12.12.571350>.
- 1477 28. Peltomäki, P., Nyström, M., Mecklin, J.-P., and Seppälä, T.T. (2023). Lynch Syndrome Genetics and
1478 Clinical Implications. *Gastroenterology* 164, 783–799. <https://doi.org/10.1053/j.gastro.2022.08.058>.
- 1479 29. Lynch, H.T., Snyder, C.L., Shaw, T.G., Heinen, C.D., and Hitchins, M.P. (2015). Milestones of Lynch
1480 syndrome: 1895-2015. *Nat. Rev. Cancer* 15, 181–194. <https://doi.org/10.1038/nrc3878>.
- 1481 30. Miao, H.-K., Chen, L.-P., Cai, D.-P., Kong, W.-J., Xiao, L., and Lin, J. (2015). MSH3 rs26279
1482 polymorphism increases cancer risk: a meta-analysis. *Int. J. Clin. Exp. Pathol.* 8, 11060–11067.
- 1483 31. Adam, R., Spier, I., Zhao, B., Kloth, M., Marquez, J., Hinrichsen, I., Kirfel, J., Tafazzoli, A.,
1484 Horpaopan, S., Uhlhaas, S., et al. (2016). Exome Sequencing Identifies Biallelic MSH3 Germline
1485 Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. *Am. J. Hum. Genet.* 99,
1486 337–351. <https://doi.org/10.1016/j.ajhg.2016.06.015>.
- 1487 32. Villy, M.-C., Masliah-Planchon, J., Schnitzler, A., Delhomelle, H., Buecher, B., Filser, M.,
1488 Merchadou, K., Golmard, L., Melaabi, S., Vacher, S., et al. (2023). MSH3: a confirmed predisposing
1489 gene for adenomatous polyposis. *J. Med. Genet.* 60, 1198–1205.
1490 <https://doi.org/10.1136/jmg-2023-109341>.
- 1491 33. de Wind, N., Dekker, M., Claij, N., Jansen, L., van Klink, Y., Radman, M., Riggins, G., van der Valk,
1492 M., van't Wout, K., and te Riele, H. (1999). HNPCC-like cancer predisposition in mice through
1493 simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. *Nat. Genet.* 23, 359–362.
1494 <https://doi.org/10.1038/15544>.
- 1495 34. Kuraguchi, M., Yang, K., Wong, E., Avdievich, E., Fan, K., Kolodner, R.D., Lipkin, M., Brown, A.M.,
1496 Kucherlapati, R., and Edelmann, W. (2001). The distinct spectra of tumor-associated Apc mutations in
1497 mismatch repair-deficient Apc1638N mice define the roles of MSH3 and MSH6 in DNA repair and
1498 intestinal tumorigenesis. *Cancer Res.* 61, 7934–7942.
- 1499 35. Ruprecht, B., Wei, L., Zheng, L., Bodea, S., Mo, X., Maschberger, M., Stoeher, G., Hahne, H.,
1500 Cornella-Taracido, I., and Chi, A. (2022). Chemoproteomic profiling to identify activity changes and
1501 functional inhibitors of DNA-binding proteins. *Cell Chem. Biol.* 29, 1639-1648.e4.
1502 <https://doi.org/10.1016/j.chembiol.2022.10.008>.
- 1503 36. Goold, R., Donaldson, J., Gidney, F., Goff, P., Hamilton, J., Elmasri, M., Coupland, L., Flower, M.,
1504 and Tabrizi, S.J. (2026). Genetic or pharmacological disruption of the MSH3 Y245/K246 IDL binding
1505 pocket slows CAG repeat expansion. Preprint at bioRxiv, <https://doi.org/10.64898/2026.02.26.707948>.
- 1506 37. Halabi, A., Ditch, S., Wang, J., and Grabczyk, E. (2012). DNA mismatch repair complex MutS β
1507 promotes GAA·TTC repeat expansion in human cells. *J. Biol. Chem.* 287, 29958–29967.
1508 <https://doi.org/10.1074/jbc.M112.356758>.
- 1509 38. Keogh, N., Chan, K.Y., Li, G.-M., and Lahue, R.S. (2017). MutS β abundance and Msh3 ATP
1510 hydrolysis activity are important drivers of CTG·CAG repeat expansions. *Nucleic Acids Res.* 45,
1511 10068–10078. <https://doi.org/10.1093/nar/gkx650>.

39. Halabi, A., Fuselier, K.T.B., and Grabczyk, E. (2018). GAA•TTC repeat expansion in human cells is mediated by mismatch repair complex MutL γ and depends upon the endonuclease domain in MLH3 isoform one. *Nucleic Acids Res.* 46, 4022–4032. <https://doi.org/10.1093/nar/gky143>.
40. Flower, M., Elmasri, M., Goold, R., Sadler, F., Ferguson, R., Ungerleider, K., Balmus, G., and Tabrizi, S.J. (2024). Specific Measurement of MutS β and MutS α DNA Mismatch Repair Activity Using a Frameshift Reporter Assay. Preprint at Social Science Research Network, <https://doi.org/10.2139/ssrn.4741096>.
41. An, J., Towle, T., Aksit, M.A., Mohiuddin, M., Castaneda, S., Nakashima, R., Moccia, R., Bulawa, C., and Fleming, J. (2023). Development of a cellular reporter assay to measure activity of MutS β , a therapeutic target for Huntington's disease. Preprint at bioRxiv, <https://doi.org/10.1101/2023.09.07.555786>.
42. Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T., and Jiricny, J. (1996). hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. *Curr. Biol. CB* 6, 1181–1184. [https://doi.org/10.1016/s0960-9822\(02\)70685-4](https://doi.org/10.1016/s0960-9822(02)70685-4).
43. Genschel, J., Littman, S.J., Drummond, J.T., and Modrich, P. (1998). Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. *J. Biol. Chem.* 273, 19895–19901. <https://doi.org/10.1074/jbc.273.31.19895>.
44. Surtees, J.A., and Alani, E. (2006). Mismatch repair factor MSH2-MSH3 binds and alters the conformation of branched DNA structures predicted to form during genetic recombination. *J. Mol. Biol.* 360, 523–536. <https://doi.org/10.1016/j.jmb.2006.05.032>.
45. Bowen, N., Smith, C.E., Srivatsan, A., Willcox, S., Griffith, J.D., and Kolodner, R.D. (2013). Reconstitution of long and short patch mismatch repair reactions using *Saccharomyces cerevisiae* proteins. *Proc. Natl. Acad. Sci. U. S. A.* 110, 18472–18477. <https://doi.org/10.1073/pnas.1318971110>.
46. Anzalone, A.V., Randolph, P.B., Davis, J.R., Sousa, A.A., Koblan, L.W., Levy, J.M., Chen, P.J., Wilson, C., Newby, G.A., Raguram, A., et al. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157. <https://doi.org/10.1038/s41586-019-1711-4>.
47. Kramer, B., Kramer, W., and Fritz, H.J. (1984). Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*. *Cell* 38, 879–887. [https://doi.org/10.1016/0092-8674\(84\)90283-6](https://doi.org/10.1016/0092-8674(84)90283-6).
48. Dohet, C., Wagner, R., and Radman, M. (1985). Repair of defined single base-pair mismatches in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 82, 503–505. <https://doi.org/10.1073/pnas.82.2.503>.
49. Su, S.S., Lahue, R.S., Au, K.G., and Modrich, P. (1988). Mismatch specificity of methyl-directed DNA mismatch correction in vitro. *J. Biol. Chem.* 263, 6829–6835.
50. Iyer, R.R., Pluciennik, A., Burdett, V., and Modrich, P.L. (2006). DNA mismatch repair: functions and mechanisms. *Chem. Rev.* 106, 302–323. <https://doi.org/10.1021/cr0404794>.
51. Chen, P.J., Hussmann, J.A., Yan, J., Knipping, F., Ravisankar, P., Chen, P.-F., Chen, C., Nelson, J.W., Newby, G.A., Sahin, M., et al. (2021). Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* 184, 5635–5652.e29. <https://doi.org/10.1016/j.cell.2021.09.018>.
52. Ferreira da Silva, J., Oliveira, G.P., Arasa-Verge, E.A., Kagiou, C., Moretton, A., Timelthaler, G., Jiricny, J., and Loizou, J.I. (2022). Prime editing efficiency and fidelity are enhanced in the absence of mismatch repair. *Nat. Commun.* 13, 760. <https://doi.org/10.1038/s41467-022-28442-1>.

- 1553 53. Koeppe, J., Weller, J., Peets, E.M., Pallaseni, A., Kuzmin, I., Raudvere, U., Peterson, H., Liberante,
1554 F.G., and Parts, L. (2023). Prediction of prime editing insertion efficiencies using sequence features and
1555 DNA repair determinants. *Nat. Biotechnol.* 41, 1446–1456. <https://doi.org/10.1038/s41587-023-01678-y>.
- 1556 54. Park, J.-C., Kim, Y.-J., Han, J.H., Kim, D., Park, M.J., Kim, J., Jang, H.-K., Bae, S., and Cha, H.-J.
1557 (2023). MutS α and MutS β as size-dependent cellular determinants for prime editing in human embryonic
1558 stem cells. *Mol. Ther. Nucleic Acids* 32, 914–922. <https://doi.org/10.1016/j.omtn.2023.05.015>.
- 1559 55. Li, X., Chen, W., Martin, B.K., Calderon, D., Lee, C., Choi, J., Chardon, F.M., McDiarmid, T.A.,
1560 Daza, R.M., Kim, H., et al. (2024). Chromatin context-dependent regulation and epigenetic manipulation
1561 of prime editing. *Cell* 187, 2411–2427.e25. <https://doi.org/10.1016/j.cell.2024.03.020>.
- 1562 56. Cirincione, A., Simpson, D., Yan, W., McNulty, R., Ravisankar, P., Solley, S.C., Yan, J., Lim, F.,
1563 Farley, E.K., Singh, M., et al. (2025). A benchmarked, high-efficiency prime editing platform for
1564 multiplexed dropout screening. *Nat. Methods* 22, 92–101. <https://doi.org/10.1038/s41592-024-02502-4>.
- 1565 57. Kim-Yip, R.P., McNulty, R., Joyce, B., Mollica, A., Chen, P.J., Ravisankar, P., Law, B.K., Liu, D.R.,
1566 Toettcher, J.E., Ivakine, E.A., et al. (2024). Efficient prime editing in two-cell mouse embryos using
1567 PEmbryo. *Nat. Biotechnol.* 42, 1822–1830. <https://doi.org/10.1038/s41587-023-02106-x>.
- 1568 58. Sharma, M., Predeus, A.V., Kovacs, N., and Feig, M. (2014). Differential mismatch recognition
1569 specificities of eukaryotic MutS homologs, MutS α and MutS β . *Biophys. J.* 106, 2483–2492.
1570 <https://doi.org/10.1016/j.bpj.2014.04.026>.
- 1571 59. Marsischky, G.T., and Kolodner, R.D. (1999). Biochemical characterization of the interaction
1572 between the *Saccharomyces cerevisiae* MSH2-MSH6 complex and mispaired bases in DNA. *J. Biol.*
1573 *Chem.* 274, 26668–26682. <https://doi.org/10.1074/jbc.274.38.26668>.
- 1574 60. Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1996). Binding of insertion/deletion DNA
1575 mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. *Curr. Biol. CB* 6,
1576 1185–1187. [https://doi.org/10.1016/s0960-9822\(02\)70686-6](https://doi.org/10.1016/s0960-9822(02)70686-6).
- 1577 61. Zou, X., Koh, G.C.C., Nanda, A.S., Degasper, A., Urgo, K., Roumeliotis, T.I., Agu, C.A., Badja, C.,
1578 Momen, S., Young, J., et al. (2021). A systematic CRISPR screen defines mutational mechanisms
1579 underpinning signatures caused by replication errors and endogenous DNA damage. *Nat. Cancer* 2,
1580 643–657. <https://doi.org/10.1038/s43018-021-00200-0>.
- 1581 62. Zhao, X., Kumari, D., Miller, C.J., Kim, G.-Y., Hayward, B., Vitalo, A.G., Pinto, R.M., and Usdin, K.
1582 (2021). Modifiers of Somatic Repeat Instability in Mouse Models of Friedreich Ataxia and the Fragile
1583 X-Related Disorders: Implications for the Mechanism of Somatic Expansion in Huntington’s Disease. *J.*
1584 *Huntingt. Dis.* 10, 149–163. <https://doi.org/10.3233/JHD-200423>.
- 1585 63. Rajagopal, S., Donaldson, J., Flower, M., Hensman Moss, D.J., and Tabrizi, S.J. (2023). Genetic
1586 modifiers of repeat expansion disorders. *Emerg. Top. Life Sci.* 7, 325–337.
1587 <https://doi.org/10.1042/ETLS20230015>.
- 1588 64. Hayward, B., Kumari, D., Santra, S., van Karnebeek, C.D.M., van Kuilenburg, A.B.P., and Usdin, K.
1589 (2024). All three MutL complexes are required for repeat expansion in a human stem cell model of
1590 CAG-repeat expansion mediated glutaminase deficiency. *Sci. Rep.* 14, 13772.
1591 <https://doi.org/10.1038/s41598-024-64480-z>.
- 1592 65. Yan, J., Oyler-Castrillo, P., Ravisankar, P., Ward, C.C., Levesque, S., Jing, Y., Simpson, D., Zhao,
1593 A., Li, H., Yan, W., et al. (2024). Improving prime editing with an endogenous small RNA-binding protein.
1594 *Nature* 628, 639–647. <https://doi.org/10.1038/s41586-024-07259-6>.

- 1595 66. Mathis, N., Allam, A., Tálas, A., Kissling, L., Benvenuto, E., Schmidheini, L., Schep, R.,
1596 Damodharan, T., Balázs, Z., Janjuha, S., et al. (2025). Machine learning prediction of prime editing
1597 efficiency across diverse chromatin contexts. *Nat. Biotechnol.* 43, 712–719.
1598 <https://doi.org/10.1038/s41587-024-02268-2>.
- 1599 67. Kadyrov, F.A., Dzantiev, L., Constantin, N., and Modrich, P. (2006). Endonucleolytic function of
1600 MutL α in human mismatch repair. *Cell* 126, 297–308. <https://doi.org/10.1016/j.cell.2006.05.039>.
- 1601 68. Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. *Annu. Rev. Biochem.* 74, 681–710.
1602 <https://doi.org/10.1146/annurev.biochem.74.082803.133243>.
- 1603 69. Senoussi, I., Mengoli, V., Cerana, A., Rinaldi, A., Marco, A., Reginato, G., Moro, S.G., Acharya, A.,
1604 Roy, M., Jayachandran, A., et al. (2025). Mechanism of trinucleotide repeat expansion by MutS β -MutL γ
1605 and contraction by FAN1. *Nat. Commun.* 16, 9445. <https://doi.org/10.1038/s41467-025-64485-w>.
- 1606 70. Kadyrova, L.Y., Gujar, V., Burdett, V., Modrich, P.L., and Kadyrov, F.A. (2020). Human MutL γ , the
1607 MLH1-MLH3 heterodimer, is an endonuclease that promotes DNA expansion. *Proc. Natl. Acad. Sci. U.*
1608 *S. A.* 117, 3535–3542. <https://doi.org/10.1073/pnas.1914718117>.
- 1609 71. Saito, S., Kato, S., Arai, U., En, A., Tsunozumi, J., Mizushima, T., Tateishi, K., and Adachi, N.
1610 (2025). HR eye & MMR eye: one-day assessment of DNA repair-defective tumors eligible for targeted
1611 therapy. *Nat. Commun.* 16, 4239. <https://doi.org/10.1038/s41467-025-59462-2>.
- 1612 72. Zhai, T., Mazzucato, P., Ricciardi, C., Christiani, D.C., Liang, L., Samson, L.D., Chaim, I.A., and
1613 Nagel, Z.D. (2025). Comprehensive Measurement of Inter-Individual Variation in DNA Repair Capacity in
1614 Healthy Individuals. *Adv. Sci.* 12, e13893. <https://doi.org/10.1002/adv.202513893>.
- 1615 73. Nagel, Z.D., Margulies, C.M., Chaim, I.A., McRee, S.K., Mazzucato, P., Ahmad, A., Abo, R.P., Butty,
1616 V.L., Forget, A.L., and Samson, L.D. (2014). Multiplexed DNA repair assays for multiple lesions and
1617 multiple doses via transcription inhibition and transcriptional mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.*
1618 111, E1823-1832. <https://doi.org/10.1073/pnas.1401182111>.
- 1619 74. Takedachi, A., Matsuishi, E., Mizusaki, S., Nagasawa, T., Fujikane, R., Hidaka, M., Iwai, S., and
1620 Kuraoka, I. (2022). Novel plasmids for the fluorescence-based evaluation of DNA mismatch repair in
1621 human cells. *Mutat. Res.* 824, 111779. <https://doi.org/10.1016/j.mrfmmm.2022.111779>.
- 1622 75. Ito, S., Shiraishi, M., Tsuchihashi, K., Takatsuka, R., Yamamoto, J., Kuraoka, I., and Iwai, S. (2018).
1623 Fluorescence detection of DNA mismatch repair in human cells. *Sci. Rep.* 8, 12181.
1624 <https://doi.org/10.1038/s41598-018-30733-x>.
- 1625 76. Lei, X., Zhu, Y., Tomkinson, A., and Sun, L. (2004). Measurement of DNA mismatch repair activity in
1626 live cells. *Nucleic Acids Res.* 32, e100. <https://doi.org/10.1093/nar/gnh098>.
- 1627 77. Hellen, C.U., and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules.
1628 *Genes Dev.* 15, 1593–1612. <https://doi.org/10.1101/gad.891101>.
- 1629 78. Kim, H.K., Yu, G., Park, J., Min, S., Lee, S., Yoon, S., and Kim, H.H. (2021). Predicting the
1630 efficiency of prime editing guide RNAs in human cells. *Nat. Biotechnol.* 39, 198–206.
1631 <https://doi.org/10.1038/s41587-020-0677-y>.
- 1632 79. Blagg, J., Riou, P., Hervieu, A., Piumatti, E., Rodriguez-Plata, M.T., Battuello, P., Peall, A., Amodio,
1633 V., Vitiello, P.P., Nightingale, D.J.H., et al. (2026). Pharmacological inhibition of PMS2 induces MMR
1634 deficiency and response to immune checkpoint blockade. *Cancer Discov.*
1635 <https://doi.org/10.1158/2159-8290.CD-26-0003>.

- 1636 80. Zhang, J.H., Chung, T.D., and Oldenburg, K.R. (1999). A Simple Statistical Parameter for Use in
1637 Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* 4, 67–73.
1638 <https://doi.org/10.1177/108705719900400206>.
- 1639 81. Lee, J.-H., Thomsen, M., Daub, H., Thieulin-Pardo, G., Steinbacher, S., Sztyler, A., Dahiya, V.,
1640 Neudegger, T., Dominguez, C., Iyer, R.R., et al. (2025). Elucidation of multiple high-resolution states of
1641 human MutS β by cryo-EM reveals interplay between ATP/ADP binding and heteroduplex DNA
1642 recognition. *Nucleic Acids Res.* 53, gkaf604. <https://doi.org/10.1093/nar/gkaf604>.
- 1643 82. McLean, Z.L., Gao, D., Correia, K., Roy, J.C.L., Shibata, S., Farnum, I.N., Valdepenas-Mellor, Z.,
1644 Kovalenko, M., Rapuru, M., Morini, E., et al. (2024). Splice modulators target PMS1 to reduce somatic
1645 expansion of the Huntington's disease-associated CAG repeat. *Nat. Commun.* 15, 3182.
1646 <https://doi.org/10.1038/s41467-024-47485-0>.
- 1647 83. Räschle, M., Marra, G., Nyström-Lahti, M., Schär, P., and Jiricny, J. (1999). Identification of
1648 hMutLbeta, a heterodimer of hMLH1 and hPMS1. *J. Biol. Chem.* 274, 32368–32375.
1649 <https://doi.org/10.1074/jbc.274.45.32368>.
- 1650 84. Kadyrova, L.Y., Kadyrov, F.F., Hayward, B., Usdin, K., and Kadyrov, F.A. (2026). Mechanism of
1651 MutL β -dependent DNA expansions. *Proc. Natl. Acad. Sci. U. S. A.* 123, e2601397123.
1652 <https://doi.org/10.1073/pnas.2601397123>.
- 1653 85. Iyer, R.R., Pluciennik, A., Napierala, M., and Wells, R.D. (2015). DNA triplet repeat expansion and
1654 mismatch repair. *Annu. Rev. Biochem.* 84, 199–226.
1655 <https://doi.org/10.1146/annurev-biochem-060614-034010>.
- 1656 86. Horn, T., Sandmann, T., Fischer, B., Axelsson, E., Huber, W., and Boutros, M. (2011). Mapping of
1657 signaling networks through synthetic genetic interaction analysis by RNAi. *Nat. Methods* 8, 341–346.
1658 <https://doi.org/10.1038/nmeth.1581>.
- 1659 87. Bassik, M.C., Kampmann, M., Lebbink, R.J., Wang, S., Hein, M.Y., Poser, I., Weibezahn, J.,
1660 Horlbeck, M.A., Chen, S., Mann, M., et al. (2013). A systematic mammalian genetic interaction map
1661 reveals pathways underlying ricin susceptibility. *Cell* 152, 909–922.
1662 <https://doi.org/10.1016/j.cell.2013.01.030>.
- 1663 88. Laufer, C., Fischer, B., Billmann, M., Huber, W., and Boutros, M. (2013). Mapping genetic
1664 interactions in human cancer cells with RNAi and multiparametric phenotyping. *Nat. Methods* 10,
1665 427–431. <https://doi.org/10.1038/nmeth.2436>.
- 1666 89. Wong, A.S.L., Choi, G.C.G., Cheng, A.A., Purcell, O., and Lu, T.K. (2015). Massively parallel
1667 high-order combinatorial genetics in human cells. *Nat. Biotechnol.* 33, 952–961.
1668 <https://doi.org/10.1038/nbt.3326>.
- 1669 90. Wong, A.S.L., Choi, G.C.G., Cui, C.H., Pregernig, G., Milani, P., Adam, M., Perli, S.D., Kazer, S.W.,
1670 Gaillard, A., Hermann, M., et al. (2016). Multiplexed barcoded CRISPR-Cas9 screening enabled by
1671 CombiGEM. *Proc. Natl. Acad. Sci. U. S. A.* 113, 2544–2549. <https://doi.org/10.1073/pnas.1517883113>.
- 1672 91. Han, K., Jeng, E.E., Hess, G.T., Morgens, D.W., Li, A., and Bassik, M.C. (2017). Synergistic drug
1673 combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat. Biotechnol.*
1674 35, 463–474. <https://doi.org/10.1038/nbt.3834>.
- 1675 92. Du, D., Roguev, A., Gordon, D.E., Chen, M., Chen, S.-H., Shales, M., Shen, J.P., Ideker, T., Mali, P.,
1676 Qi, L.S., et al. (2017). Genetic interaction mapping in mammalian cells using CRISPR interference. *Nat.*
1677 *Methods* 14, 577–580. <https://doi.org/10.1038/nmeth.4286>.

- 1678 93. Shen, J.P., Zhao, D., Sasik, R., Luebeck, J., Birmingham, A., Bojorquez-Gomez, A., Licon, K.,
1679 Klepper, K., Pekin, D., Beckett, A.N., et al. (2017). Combinatorial CRISPR-Cas9 screens for de novo
1680 mapping of genetic interactions. *Nat. Methods* 14, 573–576. <https://doi.org/10.1038/nmeth.4225>.
- 1681 94. Horlbeck, M.A., Xu, A., Wang, M., Bennett, N.K., Park, C.Y., Bogdanoff, D., Adamson, B., Chow,
1682 E.D., Kampmann, M., Peterson, T.R., et al. (2018). Mapping the Genetic Landscape of Human Cells.
1683 *Cell* 174, 953–967.e22. <https://doi.org/10.1016/j.cell.2018.06.010>.
- 1684 95. Park, J.-C., Uhm, H., Kim, Y.-W., Oh, Y.E., Lee, J.H., Yang, J., Kim, K., and Bae, S. (2025).
1685 AI-generated MLH1 small binder improves prime editing efficiency. *Cell* 188, 5831–5846.e21.
1686 <https://doi.org/10.1016/j.cell.2025.07.010>.
- 1687 96. Putnam, C.D. (2021). Strand discrimination in DNA mismatch repair. *DNA Repair* 105, 103161.
1688 <https://doi.org/10.1016/j.dnarep.2021.103161>.
- 1689 97. Matlashov, M.E., Shcherbakova, D.M., Alvelid, J., Baloban, M., Pennacchietti, F., Shemetov, A.A.,
1690 Testa, I., and Verkhusha, V.V. (2020). A set of monomeric near-infrared fluorescent proteins for
1691 multicolor imaging across scales. *Nat. Commun.* 11, 239. <https://doi.org/10.1038/s41467-019-13897-6>.
- 1692 98. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J.,
1693 Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat.*
1694 *Biotechnol.* 31, 827–832. <https://doi.org/10.1038/nbt.2647>.
- 1695 99. Buschmann, T., and Bystriykh, L.V. (2013). Levenshtein error-correcting barcodes for multiplexed
1696 DNA sequencing. *BMC Bioinformatics* 14, 272. <https://doi.org/10.1186/1471-2105-14-272>.
- 1697 100. Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn,
1698 E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by
1699 an optimized CRISPR/Cas system. *Cell* 155, 1479–1491. <https://doi.org/10.1016/j.cell.2013.12.001>.
- 1700 101. Nelson, J.W., Randolph, P.B., Shen, S.P., Everette, K.A., Chen, P.J., Anzalone, A.V., An, M., Newby,
1701 G.A., Chen, J.C., Hsu, A., et al. (2022). Engineered pegRNAs improve prime editing efficiency. *Nat.*
1702 *Biotechnol.* 40, 402–410. <https://doi.org/10.1038/s41587-021-01039-7>.
- 1703 102. Hall, M.P., Unch, J., Binkowski, B.F., Valley, M.P., Butler, B.L., Wood, M.G., Otto, P., Zimmerman, K.,
1704 Vidugiris, G., Machleidt, T., et al. (2012). Engineered luciferase reporter from a deep sea shrimp utilizing
1705 a novel imidazopyrazinone substrate. *ACS Chem. Biol.* 7, 1848–1857.
1706 <https://doi.org/10.1021/cb3002478>.
- 1707 103. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å.,
1708 Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human
1709 proteome. *Science* 347, 1260419. <https://doi.org/10.1126/science.1260419>.
- 1710 104. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M.,
1711 Kampf, C., Wester, K., Hober, S., et al. (2010). Towards a knowledge-based Human Protein Atlas. *Nat.*
1712 *Biotechnol.* 28, 1248–1250. <https://doi.org/10.1038/nbt1210-1248>.
- 1713 105. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time
1714 quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408.
1715 <https://doi.org/10.1006/meth.2001.1262>.
- 1716 106. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R.,
1717 Joung, J.K., Bauer, D.E., et al. (2019). CRISPResso2 provides accurate and rapid genome editing
1718 sequence analysis. *Nat. Biotechnol.* 37, 224–226. <https://doi.org/10.1038/s41587-019-0032-3>.

- 1719 107. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.
1720 EMBnet.journal 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>.