Insertions and Deletions as Molecular Chronometers: A Unified Framework for Developmental Counting and Cancer

Perspective

Author: Dustin Lane

AI Author: Claude Opus 4.5

Email: lanedustin91@gmail.com

ABSTRACT

Cells must track their division history to coordinate development, maintain tissue homeostasis, and suppress cancer. We propose that insertions and deletions (indels) serve as molecular chronometers recording cellular age through four parallel counting systems: (1) trinucleotide repeat expansion, creating structural thresholds sensed by DNA damage response machinery; (2) homopolymer tract frameshifts, providing digital protein-loss signals through stochastic gene inactivation; (3) retrotransposon insertions, uniquely active in neurons; and (4) transcription-associated indels generating proteome diversity through polymerase slippage.

We further propose a "repeat-guided segregation" mechanism wherein strand-specific repeat expansion during S-phase, combined with asymmetric parental histone inheritance, biases asymmetric stem cell division—directing DNA with higher counter values toward differentiating daughters while preserving lower-count genomes in stem cells. G-quadruplex and i-motif structures provide molecular readout of repeat length; ATR/CHK1 signaling directs segregation of G4-bearing chromatids; and the PIDDosome enforces thresholds when repeat burden exceeds asymmetric segregation capacity.

This framework reframes cancer not as simple mutation accumulation, but as escape from multiple counting systems—with tissue-specific vulnerability patterns reflecting which counters dominate in each tissue. The dramatic acceleration of cancer onset in constitutional mismatch repair deficiency, where counting speeds up approximately 100-fold, underscores how tightly tuned these mechanisms normally are. Indeed, counter function appears deeply integrated with cellular metabolism through three convergent hubs—α-ketoglutarate, NAD^+, and succinyl-CoA—that couple metabolic state to epigenetic regulation. Cross-species analysis strengthens this view: despite 30-fold variation in mammalian lifespan, end-of-life mutation burden varies only 3-fold, suggesting evolution has calibrated counter rates to match species longevity.

The framework generates testable predictions including measurable repeat length differences between sister chromatids, correlation between CENP-A loading and repeat length, and disrupted asymmetric division in MMR-deficient stem cells.

Keywords: indels, repeat expansion, microsatellite instability, asymmetric cell division, stem cells, cancer, DNA damage response, molecular clock, G-quadruplex, ATR/CHK1, PIDDosome, metabolic hubs, epigenetic asymmetry, 5-hydroxymethylcytosine

1. INTRODUCTION: THE COUNTING PROBLEM

1.1 The Cellular Counting Problem

Multicellular organisms face a fundamental challenge: cells must "know" how many times they have divided to coordinate developmental timing, tissue homeostasis, and aging. A hematopoietic stem cell that has divided 500 times should behave differently than one that has divided 50 times. A transit-amplifying cell must cease proliferation after a defined number of divisions to prevent tissue overgrowth. Yet the molecular mechanisms by which cells count their divisions remain incompletely understood.

The classical answer to this counting problem is telomere attrition. With each cell division, telomeres shorten by 50-200 base pairs due to the end-replication problem, eventually triggering replicative senescence through the Hayflick limit (Hayflick & Moorhead, 1961; Harley et al., 1990). Telomere length thus serves as a molecular clock counting divisions toward an inevitable endpoint. However, telomere-based counting cannot fully explain developmental timing (reviewed in Shay & Wright, 2019). Stem cells in different tissues undergo vastly different numbers of divisions yet maintain appropriate differentiation timing. Telomerase activity varies across tissues and developmental stages. And telomere length alone cannot explain how cells distinguish between the 10th and 100th division when both remain far from the senescence threshold.

Biology has evolved multiple timing mechanisms beyond telomere attrition. Epigenetic clocks, exemplified by the Horvath clock, track methylation changes at specific CpG sites that correlate remarkably with chronological age across tissues and individuals (Horvath, 2013). Circadian oscillators count daily cycles through transcription-translation feedback loops involving CLOCK, BMAL1, and their repressors (Takahashi, 2017). Developmental timers in segmentation clocks use Notch/Wnt oscillations to count somite formation intervals (Hubaud & Pourquié, 2014). Cell cycle counters track division through degradation of CDK inhibitors like p27. Each of these mechanisms serves distinct purposes, and critically, organisms employ them in parallel rather than relying on any single counter. This redundancy suggests that accurate counting is so essential to multicellular life that evolution has favored multiple overlapping systems.

The importance of cellular counting becomes apparent when counting mechanisms fail. Bypass of the Hayflick limit through telomerase reactivation is a hallmark of cancer—cells that refuse to stop dividing in part because they have disabled their division counter (Hanahan & Weinberg, 2011). Conversely, premature counting leads to accelerated aging: Hutchinson-Gilford progeria syndrome patients show epigenetic age acceleration—detectable with sufficiently sensitive clocks—, with corresponding tissue dysfunction (Horvath et al., 2018). The symmetry of these failures is instructive. Too-slow counting permits indefinite proliferation (cancer). Too-fast counting causes premature senescence (progeria, aplastic anemia). Normal tissue homeostasis requires counting mechanisms that are precisely calibrated to organismal lifespan—a calibration that appears remarkably conserved across mammalian species despite 30-fold variation in lifespan.

We use the term "molecular chronometer" in this paper's title to emphasize the time-tracking function, while "counter" predominates in the text to emphasize discrete event tracking. Both terms are appropriate: chronometers measure elapsed time, while counters track discrete

events such as cell divisions. The indel-based systems described here encompass both modes—repeat expansion counts divisions (discrete events), while transcription-associated errors reflect cumulative transcriptional activity (time-like). The framework thus bridges chronometric and counting perspectives on cellular aging.

1.2 Indels as Counting Mechanisms

Insertions and deletions—collectively termed indels—have traditionally been viewed as pathological errors (reviewed in Mullaney et al., 2010). Unlike point mutations that substitute one base for another, indels add or remove genetic material, often with severe consequences. Frameshift indels in coding regions typically destroy protein function. Repeat expansions cause devastating neurological diseases. The prevailing view has treated indels as threats to be suppressed by DNA repair machinery.

A critical question arises: are these indels merely passive damage, or do they constitute active counting mechanisms? Several lines of evidence favor the active counting interpretation.

First, repeat expansion diseases exhibit a two-threshold model inconsistent with simple damage accumulation. If repeat expansion were simply accumulated damage, one would expect a continuous relationship between repeat length and pathology. Instead, these diseases exhibit TWO distinct thresholds. The first is an inherited permissive threshold—the repeat length above which further somatic expansion can occur (approximately 36 CAG repeats in Huntington's disease; Huntington's Disease Collaborative Research Group, 1993). The second is an intracellular pathogenic threshold—the length at which cells become dysfunctional, estimated at approximately 115 CAG repeats (95% CI: 70-165) based on modeling of disease onset (Donaldson et al., 2021). Most patients inherit 40-50 repeats but develop symptoms in midlife, implying that the cells that degenerate have accumulated substantially longer expansions somatically. The existence of two discrete thresholds—one for counting initiation, one for pathogenic readout—supports an active counting interpretation over passive damage accumulation.

Second, the expansion rate itself is not constant but shows nonlinear threshold behavior. Studies in induced pluripotent stem cells (iPSCs) from myotonic dystrophy patients identified an interval of 57-126 repeats where expansion rates dramatically increase (Du et al., 2013). Clones with longer repeats expanded faster, creating positive feedback. Furthermore, expansion ceased upon differentiation into embryoid bodies or neurospheres, indicating that the "counting" is linked to proliferative state. This threshold-gated acceleration is inconsistent with random damage but consistent with a calibrated counting system.

Third, NAD+ provides a quantitative readout of cumulative DNA damage at the molecular level. In human tissues, NAD+ levels show strong negative correlation with age ($r=-0.71$ in males, $p=0.001$), while PARP activity—reflecting DNA damage response—correlates positively with age ($r=0.77$, $p<0.0001$) (Massudi et al., 2012). PARP hyperactivation consumes NAD+, creating a molecular counter where cumulative damage is recorded as cumulative NAD+ depletion. Below a threshold, cells maintain function; below a critical NAD+ level, Sirtuin activity fails and cells enter senescence (Verdin, 2015).

Fourth, damage sensing is causally integrated with epigenetic response. The ATR kinase, activated by replication stress at expanded repeats, directly phosphorylates and activates TET3,

enhancing 5-hydroxymethylcytosine generation (Jiang et al., 2017). This causal link between damage sensing and epigenetic modification indicates that the system is not passively accumulating lesions but actively converting damage signals into heritable epigenetic states. An evolved counting system would be expected to show such integration; random damage would not.

Yet accumulating evidence reveals that programmed indels serve essential developmental functions. V(D)J recombination in lymphocytes requires RAG1/2-mediated DNA cleavage and imprecise rejoining, generating the antibody diversity essential for adaptive immunity (Schatz & Ji, 2011). Meiotic recombination initiated by SPO11-induced double-strand breaks involves resolution pathways that generate indels, ensuring genetic diversity in offspring (Lam & Keeney, 2015). In muscle differentiation, transient DNA breaks within a base excision repair framework promote transcriptional activation of myogenic regulators (Al-Khalaf et al., 2016). These examples demonstrate that controlled DNA lesions can sculpt differentiation trajectories. If programmed indels serve developmental functions, might “error” indels serve similar roles? We propose that the indels generated through replication slippage, mismatch repair processing, and transcription errors are not merely tolerated damage but constitute a sophisticated counting system that tracks cellular division history.

Several properties distinguish indels from point mutations and make them particularly suitable as cellular counters (Mullaney et al., 2010). First, indels exhibit threshold behavior: point mutations accumulate linearly and produce graded effects on protein function, whereas indels, particularly frameshifts, create discrete thresholds—a gene is either in-frame or out-of-frame, functional or destroyed. Second, repeat expansions produce structural consequences, crossing structural thresholds where DNA secondary structures (G-quadruplexes, R-loops, hairpins) suddenly become stable; these structures are directly sensed by the DNA damage response machinery, providing a built-in readout mechanism. Third, repair processing amplifies indel formation: the mismatch repair system paradoxically amplifies rather than suppresses certain indel types (Pearson et al., 2005), creating positive feedback that accelerates counting as thresholds approach. Fourth, indels are largely irreversible: while point mutations can revert, frameshift mutations effectively cannot—a second frameshift rarely restores function because it produces a different amino acid sequence. Finally, indel rates are tissue-specific, varying dramatically across tissues (Milholland et al., 2017) due to differences in replication rate, transcription activity, MMR proficiency, and chromatin state.

The central insight underlying our framework is that indel mutation rates are not constant but increase substantially with substrate length. The precise mathematical form of this relationship has been debated in the literature: some studies report exponential growth (Lai & Sun, 2003; Brinkmann et al., 1998), others find a linear relationship (Kruglyak et al., 1998), and still others suggest more complex functions that resist simple parameterization (Calabrese & Durrett, 2003). Importantly, all models agree on the fundamental observation that longer repeats mutate at substantially higher rates than shorter repeats—the critical requirement for the counting mechanism we propose. Whether the relationship is exponential, steeply linear, or polynomial, the consequence is the same: repeat tracts experience accelerating instability as they lengthen, creating natural thresholds below which the repeat is relatively stable and above which instability becomes increasingly probable. This steep length-dependence transforms linear inputs (cell divisions) into non-linear outputs (threshold crossings), with mutation probability increasing by orders of magnitude across biologically relevant length ranges.

Quantitative estimates of this length-dependence come from multiple approaches. Lai and Sun (2003) developed a mathematical model using microsatellite distributions in the human genome to estimate slippage mutation rates without assuming any a priori relationship to repeat unit number. Their analysis suggested exponential scaling, with expansion occurring more frequently for short microsatellites and contraction more frequently for long microsatellites—a pattern that explains the observed scarcity of long microsatellites in the genome. Kruglyak and colleagues (1998), analyzing yeast microsatellites, proposed a linear relationship but still documented substantial increases in mutation rate with length. Leopoldino and Pena (2003) observed that repeat length explains approximately 80% of the variance in mutation rate for autosomal tetranucleotide microsatellites, compatible with either power or exponential relationships. For trinucleotide repeats specifically, yeast studies demonstrate expansion rates of approximately 10^{-5} per cell per generation for $(CTG)_{25}$ and $(CGG)_{25}$ tracts (Miret et al., 1997; Samadashwily et al., 2001), with mutation in lagging strand synthesis machinery increasing expansion rates by nearly two orders of magnitude—consistent with replication-dependent expansion mechanisms. The threshold length at which instability accelerates also varies by sequence context, with estimates ranging from as few as 3 repeats (Dieringer & Schlötterer, 2003) to 8-10 repeats (Rose & Falush, 1998; Baptiste et al., 2013), and some studies finding no discrete threshold at all (Leclercq et al., 2010). This variability likely reflects genuine biological differences across repeat types, genomic contexts, and species.

1.3 A Framework of Four Parallel Counters

We propose that cells employ at least four parallel indel-based counting systems, each with distinct mechanisms and threshold types. The first counter involves trinucleotide repeat expansion, in which replication-coupled expansion creates structural thresholds through G-quadruplex and R-loop formation, detected by ATR/ATM signaling. The second counter involves homopolymer tract frameshifts, where mononucleotide run instability generates frameshift mutations that inactivate genes in a stochastic, length-dependent manner. The third counter involves retrotransposon insertions, particularly LINE-1 activity in neurons, which generates insertional diversity with functional consequences. The fourth counter involves transcription-associated indels, where RNA polymerase slippage during transcription generates frameshift transcripts, contributing to proteome diversity (Figure 1A).

These four counters operate in parallel, with tissue-specific dominance patterns that explain why different tissues are vulnerable to different diseases (Figure 1B). Cancer, in this framework, represents escape from multiple counting systems—a cell that has disabled the checkpoints, frameshifted the tumor suppressors, and evaded the consequences that normally follow threshold crossing (Figure 1D).

The existence of multiple parallel counting systems might seem redundant, but evolutionary theory suggests this redundancy is adaptive. The primary advantage is robustness against single-point failures: a counting system based on a single mechanism would be vulnerable to mutations that disable that mechanism, and parallel counters require multiple independent failures for complete evasion—explaining why carcinogenesis typically requires 5-7 driver mutations accumulating over decades (Vogelstein et al., 2013). Parallel counters also enable tissue-specific tuning: different tissues have different renewal requirements, and parallel counters with tissue-specific dominance allow each organ to count on its own timescale.

Furthermore, multiple counters achieve separation of concerns: trinucleotide repeats count replication events, transcription indels count transcriptional activity, and L1 insertions count time in post-mitotic cells. Finally, parallel counters provide evolutionary flexibility: somatic mutation rates scale inversely with lifespan across mammals, suggesting that counter calibration is itself under selection (Cagan et al., 2022).

The molecular machinery that interprets these counter signals to execute fate decisions—differentiation, senescence, or apoptosis—represents a distinct layer of cellular logic that will be addressed in future work. Our focus here is on the counting mechanisms themselves: what accumulates, at what rate, and through what molecular processes.

Figure 1: The Four-Counter Framework

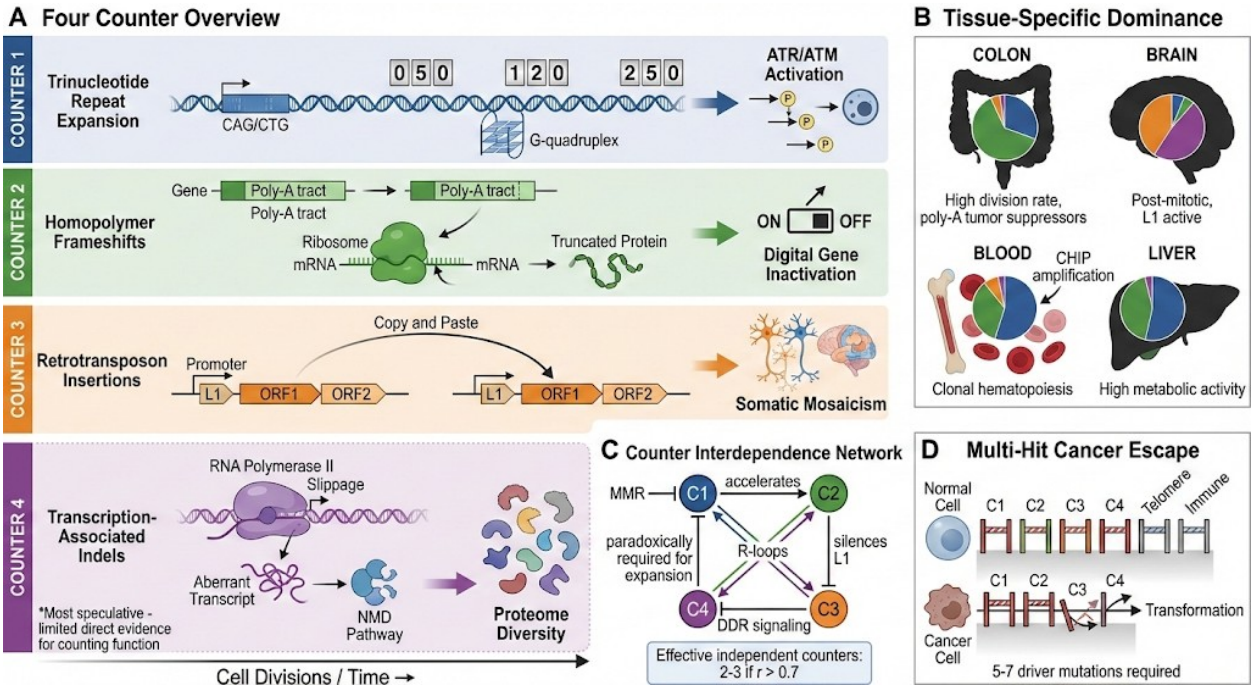

Figure 1. Four parallel indel-based counting systems and their tissue-specific dominance patterns. (A) Schematic overview of the four counters: Counter 1 (trinucleotide repeat expansion) operates through replication-coupled expansion creating G4/i-motif structural thresholds sensed by ATR/ATM; Counter 2 (homopolymer frameshifts) generates digital protein-loss signals through stochastic gene inactivation; Counter 3 (retrotransposon insertions) creates somatic mosaicism primarily in neurons; Counter 4 (transcription-associated indels) produces proteome diversity through polymerase slippage. (B) Tissue-specific counter dominance: colon (Counter 2 dominant—high division rate plus poly-A tracts in tumor suppressors); brain (Counters 3 and 4 dominant—post-mitotic neurons accumulate L1 insertions and transcription errors); blood (Counter 1 with clonal amplification through CHIP); liver (Counters 1 and 2 with high metabolic activity). (C) Counter interdependence diagram showing mechanistic connections: MMR deficiency accelerates Counter 2 while paradoxically reducing Counter 1 expansion; transcription (Counter 4) and replication (Counter 1) couple through R-loop formation; L1 activity (Counter 3) is regulated by methylation status

affected by DDR signaling from Counter 1. Arrows indicate positive regulation; blunted lines indicate negative regulation. (D) Multi-hit cancer escape model: transformation requires disabling multiple counters, explaining why carcinogenesis typically requires 5-7 driver mutations accumulating over decades.

1.4 Epistemological Framework: Distinguishing Counting from Damage

Before proceeding, we must explicitly address a fundamental interpretive challenge: distinguishing evolved counting functions from epiphenomenal damage accumulation. Two interpretations of the evidence are possible:

The first interpretation (the strong claim) holds that Evolution has selected indel accumulation mechanisms *specifically* to count cellular divisions and trigger appropriate fate responses. Under this view, the molecular machinery that generates indels and responds to them constitutes an evolved timing system analogous to circadian clocks or developmental oscillators. The second interpretation (the weak claim) suggests that Indels accumulate as unavoidable consequences of DNA replication and transcription, and cellular responses to this damage have been co-opted by evolution to trigger fate transitions. Under this view, indels are primarily damage that cells have learned to exploit informationally.

These interpretations are not mutually exclusive, and the truth likely lies along a continuum. However, distinguishing them matters for how we interpret the evidence and design experiments. Several criteria could distinguish evolved counting from exploited damage:

Criterion 1: Rate calibration. If indel accumulation rates were tuned by evolution to match organismal lifespan, this would support evolved counting. The Cagan et al. (2022) finding that mutation rates scale inversely with lifespan across mammals is consistent with this criterion, though alternative explanations (metabolic rate differences, body temperature, repair capacity scaling) cannot be excluded.

Criterion 2: Threshold specificity. If thresholds for cellular responses correspond to biologically meaningful division numbers rather than arbitrary damage levels, this would support evolved counting. The two-threshold structure of repeat expansion diseases—with permissive and pathogenic thresholds corresponding to distinct cellular states—is consistent with this criterion.

Criterion 3: Readout integration. If damage-sensing pathways show evidence of being specifically wired to fate-determination machinery beyond generic stress responses, this would support evolved counting. The ATR-TET3 phosphorylation axis linking replication stress to epigenetic modification (Jiang et al., 2017) exemplifies such integration.

Criterion 4: Counter independence. If multiple indel types function as genuinely parallel counters with distinct tissue-specific roles, this would support evolved redundancy. However, as we discuss in Section 1.5, the counters show significant interdependence that complicates this interpretation.

Throughout this manuscript, we present evidence favoring the counting interpretation while acknowledging that definitive proof requires experimental discrimination between these

hypotheses. Specifically, we cannot exclude the possibility that all observations reflect biophysical constraints on DNA stability plus selection for damage tolerance, without invoking counting as a primary function. The testable predictions in Section 9 are designed in part to distinguish these interpretations.

We use "counter" and "counting" as descriptive terms for the pattern of accumulation and threshold-crossing, without claiming that this terminology resolves the question of whether counting is the evolved function versus an emergent property of damage accumulation. The framework's value lies in generating testable predictions regardless of which interpretation proves correct.

1.5 Counter Interdependence

The four counters are presented as parallel systems with tissue-specific dominance, but their interdependence must be acknowledged (Figure 1C). MMR deficiency accelerates Counter 2 (frameshifts) but also affects Counter 1 (repeat expansion) through the paradoxical MMR role in expansion—MMR-deficient cells show reduced repeat expansion because the "toxic oxidation cycles" that drive expansion require functional MMR. Transcription (Counter 4) and replication (Counter 1) share mechanical coupling through R-loop formation at transcribed repeat sequences. L1 activity (Counter 3) is regulated by DNA methylation status, which is itself affected by DNA damage response signaling that responds to repeat expansion.

This interdependence has implications for the framework's predictions. If counters are highly correlated—rising and falling together due to shared regulatory mechanisms—then the "multi-hit escape" model for cancer (Section 8) becomes less compelling, as disabling one counter might partially disable others. Conversely, if counters are genuinely independent, their parallel operation provides robust tumor suppression requiring multiple independent failures.

Empirical resolution requires measuring correlations between counter values across single cells. Technologies such as long-read single-cell sequencing combined with L1 insertion detection could address whether individual cells show correlated or independent counter progression. If counter correlations are high ($r > 0.7$ across single cells), then the effective number of independent counters may be closer to 2-3 than 4, and disabling one counter would substantially compromise the others. If correlations are low ($r < 0.3$), the counters operate largely independently, and the multi-hit escape model predicts that cancer requires independent disabling of each. Based on mechanistic considerations, Counters 1 and 4 (repeat expansion and transcription errors) are likely coupled at R-loop-forming loci, while Counters 2 and 3 (frameshifts and L1 insertions) may be more independent. These predictions remain to be tested systematically.

2. COUNTER 1: TRINUCLEOTIDE REPEAT EXPANSION

2.1 Molecular Mechanisms of Repeat Expansion

Trinucleotide repeats are among the most unstable sequences in the genome, and their instability increases dramatically with length (reviewed in Usdin et al., 2015). Comparative genomic analysis has shown that rates of tandem insertions and deletions at microsatellite loci strongly deviate from background rates, with no detectable lower threshold length for slippage (Leclercq et al., 2010). Even short repeats experience slippage, and the rate increases

continuously with length. The mathematical relationship can be approximated as $\mu(n) \propto e^{(kn)}$, where μ is the mutation rate, n is the repeat length in units, and k is a sequence-specific constant. The biological consequence is that repeat expansion is self-accelerating: each expansion event increases the probability of subsequent expansion, creating a positive feedback loop that drives repeats toward pathological lengths.

The rate of repeat expansion varies by sequence context, cell type, and physiological state. Germline expansion rates have been precisely measured in repeat expansion disease families. CAG repeats in Huntington's disease expand by approximately 1-2 units per transmission in the male germline (The Huntington's Disease Collaborative Research Group, 1993). CTG repeats in myotonic dystrophy are highly variable, with expansions of 50-1000+ units possible in single transmission events (Brook et al., 1992). CGG repeats in Fragile X show that premutation alleles (55-200 repeats) expand to full mutations (>200 repeats) in approximately 95% of maternal transmissions (Verkerk et al., 1991; Oberlé et al., 1991). Somatic expansion rates are tissue-specific and generally lower than germline rates, but accumulate over a lifetime. Studies of dinucleotide repeats have demonstrated expansion bias for microsatellites up to 20 units in somatic cells, with MutSα and MutLα complexes providing protection against expansion mutations (Baptiste et al., 2013).

Repeat expansion occurs primarily during DNA replication. During lagging strand synthesis, Okazaki fragment processing creates transient single-stranded regions where repeat sequences can form secondary structures (Mirkin, 2007). These structures escape polymerase proofreading and, if not removed by flap endonuclease or MMR, become incorporated as expansions. Studies of early development have shown that endogenous mutational rates are higher in the first cell division but decrease to approximately one mutation per cell per division later in life (Park et al., 2021). The mismatch repair (MMR) system plays a paradoxical role: while MMR normally corrects replication errors, it actually promotes expansion at trinucleotide repeats through "toxic oxidation cycles." Overexpressing MSH2 favors the formation of heteroduplex regions, leading to increased contractions and expansions of CAG/CTG repeat tracts during replication (Viterbo et al., 2016). This MMR involvement creates the strand-specificity that becomes critical for our segregation hypothesis (Section 6).

2.2 Structural Thresholds and DDR Activation

As repeats lengthen, they cross critical thresholds for secondary structure formation. G-rich sequences like (CGG)n form G-quadruplex (G4) structures—stable four-stranded configurations (Biffi et al., 2013) where guanine bases form planar quartets. G4 stability increases cooperatively with the number of stacked quartets; below approximately 30 CGG repeats, G4s are transient, while above approximately 55 repeats (the Fragile X premutation threshold), G4s persist long enough to stall replication forks. C-rich template strands favor R-loop formation during transcription—RNA:DNA hybrids that displace the non-template strand. R-loops promote trinucleotide repeat deletion through base excision repair activities, preferentially leading to repeat deletion during BER by disrupting the balance between TNR addition and removal (Laverde et al., 2020). CAG and CTG sequences also form intrastrand hairpin structures where mismatched C-A or T-G base pairs create bulged loops. These structures form cooperatively, with stability increasing non-linearly with length; below approximately 30-35 CAG repeats, secondary structures are transient and easily resolved, while above this threshold, structures persist long enough to stall replication forks and activate DNA damage signaling.

Restarted replication forks exhibit low fidelity through repeat sequences, and recombination-dependent replication contributes to trinucleotide repeat instability (Gold et al., 2021).

When repeat-associated secondary structures stall replication forks, they activate the DNA damage response (DDR). ATR kinase is recruited to RPA-coated single-stranded DNA at stalled forks, initiating a signaling cascade through CHK1 phosphorylation (Cimprich & Cortez, 2008). Genome-wide studies have identified structure-forming repeats as principal sites of fork collapse upon ATR inhibition, demonstrating that the genomic feature most strongly associated with ATR dependence is repetitive DNA with high structure-forming potential (Shastri et al., 2018). Persistent structures can convert to double-strand breaks, activating ATM and its downstream effectors including p53. This DDR activation provides the mechanism by which cells "read" the repeat counter. A cell with short repeats experiences minimal replication stress and continues proliferating normally; a cell with long repeats experiences chronic ATR activation, checkpoint engagement, and increased probability of fate transition—differentiation, senescence, or apoptosis depending on cellular context. The threshold is not absolute but probabilistic: each cell division with a long repeat carries a probability of triggering consequential DDR signaling, and as repeats lengthen, this probability increases until fate transition becomes nearly inevitable.

2.3 Repeat Expansion Diseases as Counting Dysfunction

The approximately 50 known repeat expansion diseases (reviewed in Paulson, 2018) can be organized by their threshold characteristics and tissue specificity. The polyglutamine diseases involve CAG expansions in coding regions that produce toxic polyglutamine tracts: Huntington's disease (HTT gene) has a threshold of approximately 36 repeats and primarily affects the striatum; the spinocerebellar ataxias (SCA1, 2, 3, 6, 7, 17) have thresholds ranging from 35-55 repeats and primarily affect the cerebellum; Kennedy disease (AR gene) has a threshold of approximately 38 repeats and affects motor neurons (La Spada et al., 1991). The RNA toxicity diseases involve expansions in non-coding regions where the repeat-containing RNA itself is pathogenic: myotonic dystrophy type 1 (DMPK 3'UTR) has a threshold of approximately 50 CTG repeats and affects muscle and heart; Fragile X-associated tremor/ataxia syndrome (FMR1 5'UTR) occurs in the 55-200 CGG premutation range; C9orf72-associated ALS/FTD has a threshold of approximately 30 GGGGCC repeats (DeJesus-Hernandez et al., 2011). The silencing diseases involve expansions that cause transcriptional shutdown: Fragile X syndrome occurs when more than 200 CGG repeats cause promoter methylation and gene silencing; Friedreich's ataxia involves GAA expansion in the first intron of FXN, causing heterochromatin formation and reduced frataxin expression (Campuzano et al., 1996).

2.4 G-Quadruplex and i-Motif: The Structural Readout of Repeat Length

The transition from stable to unstable repeat behavior reflects a fundamental structural threshold: the formation of non-B DNA secondary structures that are directly sensed by the DNA damage response machinery. Two complementary structures—G-quadruplexes (G4) on G-rich strands and i-motifs on C-rich strands—provide the molecular mechanism by which cells "read" repeat length and convert it into biological signals.

G-quadruplexes form when four guanine bases associate through Hoogsteen hydrogen bonding to create planar G-quartets, which stack to form remarkably stable four-stranded structures

(Rhodes & Lipps, 2015). At CGG repeats (Fragile X), GGGGCC repeats (C9orf72-ALS), and the G-rich strand of CAG/CTG repeats, G4 stability increases cooperatively with repeat length. Below approximately 30-35 repeats, G4 structures are transient and readily resolved by helicases; above this threshold, G4 persistence increases dramatically (Usdin & Woodford, 1995), creating replication fork barriers that activate ATR signaling. Critically, G4-stabilizing compounds such as pyridostatin increase the frequency of replication stress and nonrandom DNA segregation (Xing et al., 2020), demonstrating that G4 structures are sufficient to trigger the segregation machinery.

On the complementary C-rich strand, i-motif structures form through intercalated cytosine-cytosine+ base pairs under slightly acidic conditions. While historically considered pH-dependent artifacts, recent evidence demonstrates that i-motifs form in the nuclei of human cells, with formation that is cell-cycle and pH dependent, and occur in regulatory regions including promoters and telomeres (Zeraati et al., 2018). At CTG repeats (myotonic dystrophy) and CCG repeats, i-motifs create strand-specific replication impediments that contribute to repeat instability.

The strand asymmetry of these structures has profound implications for sister chromatid differentiation. During replication of a CAG/CTG repeat, the G-rich template strand can form G4 structures that stall the replication fork, while the C-rich template strand forms i-motifs. Because leading and lagging strand synthesis handle these structures differently, and because MMR processing of slipped intermediates is strand-specific, the two daughter duplexes emerge from S-phase with measurably different properties (Lopes et al., 2011). The sister chromatid that inherited the G4-forming strand during replication carries a distinct structural signature —increased DDR protein recruitment (RPA, FANCD2, γH2AX), altered histone modifications, and differential CENP-A loading potential. This structural asymmetry provides the molecular basis for the repeat-guided segregation mechanism detailed in Section 6.

Each disease can be reinterpreted through our counting framework. Huntington's disease represents a counter running too fast—the expanded allele crosses structural thresholds earlier in life than normal alleles would. The phenomenon of anticipation—earlier onset in successive generations—reflects inherited counter acceleration. A crucial feature of repeat expansion is somatic mosaicism—repeat lengths vary across tissues and even among cells within the same tissue, reflecting the counting process itself. In Huntington's disease, somatic mosaicism correlates with disease severity. The Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium demonstrated that the timing of onset is influenced by polymorphic variation at multiple DNA maintenance genes, suggesting that the onset-determining property of the uninterrupted CAG repeat is its propensity for length instability leading to somatic expansion (GeM-HD Consortium, 2019). This mosaicism means that repeat length is not a fixed property inherited at birth but a dynamic value that continues to increment throughout life. The counter is still counting—and different cells, with different division histories and microenvironments, accumulate different counter values. This reframing suggests that repeat expansion diseases are not fundamentally different from normal aging; they represent the same counting mechanism operating on accelerated timescales, crossing thresholds in years rather than decades.

3. COUNTER 2: HOMOPOLYMER FRAMESHIFTS

3.1 Homopolymer Instability and Digital Counting

Homopolymer tracts—runs of identical nucleotides such as AAAAAAA or CCCCCCC—are hotspots for insertion and deletion mutations (Levinson & Gutman, 1987; Ellegren, 2004). During replication, polymerase slippage can add or remove single nucleotides, with probability increasing with tract length. MutSα and MutLα complexes protect against expansion mutations, but deletion rates in homopolymeric runs increase with increasing length of the mononucleotide repeat (Baptiste et al., 2013). The length-dependence is steep: a run of 6 adenines is far more stable than a run of 10, which is far more stable than a run of 15. Quantitative analyses demonstrate that each additional nucleotide in a homopolymer substantially increases the frameshift mutation rate, with genome-wide studies finding strong length-dependence across both human and chimpanzee microsatellites (Kelkar et al., 2008). This steep length-rate relationship parallels that observed for trinucleotide repeats, though the exact mathematical form varies by sequence context and genomic location.

Trinucleotide repeats provide analog counting—a graded signal that increases continuously with repeat length. Homopolymer frameshifts provide digital counting through the slipped-strand mispairing mechanism first described by Streisinger et al., 1966—a binary switch from functional to non-functional protein. Consider a gene with a 10-nucleotide poly-A tract in its coding region: each cell division carries a small probability of frameshift mutation, and once a frameshift occurs, the gene is inactivated. This is a one-way transition—the counter has "clicked" and cannot be reset. The two systems carry different information content: analog counters track continuous variables such as division number and time, while digital counters track discrete events such as gene inactivation and pathway loss. They also differ in noise characteristics: analog signals are subject to continuous fluctuation, whereas digital signals are immune to noise once the threshold is crossed, since a frameshifted gene is unambiguously non-functional. The cell therefore has access to two complementary information channels: continuous measurement of "how far along" through repeat length, and discrete events marking "something happened" through frameshifts.

3.2 Frameshift Targets and Microsatellite Instability (MSI) Acceleration

Certain genes are particularly vulnerable to frameshift inactivation due to long homopolymer tracts in their coding regions (reviewed in Yamamoto & Imai, 2015). Among tumor suppressors with long homopolymers, TGFBR2 contains an (A)10 tract that is frequently frameshifted in MSI-H colorectal cancers; targeted microsatellite studies report rates of 70-90% (Markowitz et al., 1995; Parsons et al., 1995). BAX contains a (G)8 tract and is frameshifted in approximately 50% of MSI-H cancers using targeted analysis (Rampino et al., 1997). MSH3 and MSH6 both contain homopolymer tracts, and their frameshift creates feedback acceleration of the counting system. DNA repair genes MRE11 and RAD50 contain long tracts and are frequently mutated in MSI-H colorectal cancers (Miquel et al., 2007). Immune recognition genes including B2M and JAK1 are frequently targeted, with mutations in these genes conferring resistance to immune checkpoint blockade (Budczies et al., 2017). This target landscape reveals that the digital counter preferentially affects growth control, DNA repair, and immune recognition—precisely the pathways that must be inactivated for cancer progression.

The MRE11 frameshift deserves particular attention because it constrains how cancer cells can achieve telomere immortalization. MRE11 contains a poly(T)11 tract in intron 4 that is frequently mutated in MSI tumors; targeted microsatellite studies report mutation rates of 70-90% in MSI-H colorectal cancers (Giannini et al., 2002). This frameshift disrupts the MRN complex (MRE11-RAD50-NBS1), which is essential for the Alternative Lengthening of Telomeres (ALT) pathway—a recombination-based mechanism for telomere maintenance (Jiang et al., 2005). Because ALT requires intact MRN for the homology-directed recombination that extends telomeres, MSI tumors with disrupted MRE11 are effectively locked into telomerase-dependent immortalization. This explains the striking tissue-specific pattern of telomere maintenance mechanism selection: epithelial carcinomas, which frequently arise through MSI pathways, show only 5-15% ALT prevalence, while sarcomas and neural tumors with intact MRN show 25-60% ALT (Heaphy et al., 2011). The Counter 2 frameshift thus not only accelerates cancer progression but also constrains which counter escape routes remain available for immortalization.

Microsatellite instability-high (MSI-H) cancers provide a natural experiment in counter acceleration. When MMR is defective, homopolymer frameshifts accumulate at dramatically elevated rates—a phenomenon first discovered by multiple groups who identified the mutator phenotype underlying hereditary colorectal cancer (Ionov et al., 1993; Thibodeau et al., 1993). In MMR-proficient cells, frameshift rates at homopolymer tracts of 8-10 nucleotides are in the range of 10^-5 to 10^-6 per cell division; in MMR-deficient cells, these rates increase by approximately 100- to 1,000-fold (Parsons et al., 1993; Boyer et al., 2002). The clinical consequences in constitutional MMR deficiency (CMMRD) are stark: cumulative cancer incidence by age 18 years reaches 90% (Ercan et al., 2024). Lynch syndrome, caused by heterozygous germline mutations in MMR genes, represents an intermediate state of counter acceleration (Fishel et al., 1993; Leach et al., 1993; Papadopoulos et al., 1994). Prospective data show cumulative colorectal cancer incidence at age 70 of approximately 50% for MLH1 and MSH2 carriers, but lower (approximately 10-15%) for MSH6 and PMS2 carriers (Møller et al., 2022).

3.3 Therapeutic and Developmental Implications

Beyond cancer, stochastic frameshift inactivation may serve normal developmental functions. Consider a stem cell niche where cells express a gene essential for maintaining stemness. If this gene contains a homopolymer tract, each division carries a small probability of inactivation. Over time, cells at the population level will lose this gene stochastically, with the rate determined by tract length and MMR efficiency. This provides a population-level timer: even if individual cells cannot “know” how many times they have divided, the population as a whole will show a predictable fraction of inactivated cells at any time point. Natural selection could tune this timer by adjusting homopolymer tract lengths within critical genes—a mechanism for encoding cell-autonomous differentiation probability into DNA sequence.

MSI-H tumors respond remarkably well to immune checkpoint inhibitors, and this response directly relates to the frameshift counter. Each frameshift mutation generates a neo-open reading frame producing novel peptides recognized as foreign by T cells. A subset of MSI-H associated highly recurrent frameshift mutations yield shared immunogenic neoantigens that can serve as therapeutic targets (Roudko et al., 2021). This explains why pembrolizumab was the first cancer drug approved by tumor molecular profile (MSI-H) rather than tissue of origin

—the frameshift counter creates a common vulnerability across all MSI-H cancers regardless of where they arise (Le et al., 2015; Le et al., 2017). The therapeutic implication is profound: the very mechanism by which cancer cells escape growth control (frameshift inactivation of tumor suppressors) simultaneously creates their vulnerability to immunotherapy (neoantigen generation).

4. COUNTER 3: RETROTRANSPOSON INSERTIONS

4.1 LINE-1 as a Neural Counting Mechanism

Long interspersed nuclear elements (LINEs), particularly LINE-1 (L1), are active retrotransposons comprising approximately 17% of the human genome (Lander et al., 2001; Moran et al., 1996). While L1 is heavily silenced in most somatic tissues through DNA methylation and piRNA pathways, it remains active in specific contexts that suggest regulated rather than chaotic activity. Recent work has demonstrated that L1 cis-regulatory elements contribute to parvalbumin interneuron development and transcriptome diversity through SOX6-dependent activation of novel L1 promoter-driven transcript isoforms (Bodea et al., 2024). This finding reveals that L1 activity is not merely tolerated but serves functional developmental roles in the nervous system.

Neurons present a paradox for counting systems. As post-mitotic cells, they cannot use division-coupled mechanisms like repeat expansion, yet they must maintain function for decades. Seminal work demonstrated that engineered human L1 elements can retrotranspose in neuronal precursor cells, generating somatic mosaicism in the mammalian brain (Muotri et al., 2005; Coufal et al., 2009). However, the frequency of endogenous L1 retrotransposition in human neurons remains actively debated. Initial single-neuron sequencing studies found that most cortical and caudate neurons lack detectable somatic insertions, estimating fewer than 0.6 unique insertions per neuron and concluding that "L1 is not a major generator of neuronal diversity in cortex and caudate" (Evrony et al., 2012). Subsequent work using different methodologies estimated substantially higher rates—approximately 13.7 somatic L1 insertions per hippocampal neuron, enriched in neuronal stem cell enhancers and protein-coding genes (Upton et al., 2015). This 20-fold discrepancy reflects methodological differences including cell population selection, whole-genome amplification approaches, and criteria for validating candidate insertions (Faulkner & Garcia-Perez, 2017). The true frequency likely varies by brain region, developmental stage, and individual genetic background (reviewed in Suarez et al., 2018). Regardless of the precise rate, even conservative estimates extrapolate to millions of unique L1 insertions across the ~86 billion neurons in a human brain, representing a substantial reservoir of somatic genetic diversity. Several molecular features make neurons permissive to L1 activity: open chromatin at L1 loci, reduced methylation at evolutionarily young L1 elements, diminished piRNA pathway activity, and transcription factor-mediated activation during development.

4.2 Insertion Bias and Functional Consequences

L1 insertions are not randomly distributed across the genome but show pronounced biases toward regions of accessible chromatin, including actively expressed genes and their enhancers (Bodea et al., 2018). This biased distribution means that L1 insertions preferentially affect neural genes—precisely the genes whose modification would most directly influence neuronal

phenotype. Rather than viewing this as purely pathological, the pattern suggests L1 insertions create phenotypic diversity where it matters most for neural function.

Each L1 insertion is a permanent genetic change that accumulates over the neuron's lifetime (Upton et al., 2015). L1 sequences contain their own enhancers and can serve as alternative promoters, creating novel transcripts unique to neurons with specific insertions. The consequence is a population of neurons with slightly different genetic capabilities, and this somatic mosaicism may contribute to the individuality of neural circuits and the heterogeneity of aging phenotypes.

4.3 L1 Activity Beyond the Brain and Inflammatory Consequences

While neurons show the highest L1 activity, retrotransposition occurs in other contexts. L1 is highly active in preimplantation embryos before global silencing is established, creating constitutional L1 mosaicism (Kano et al., 2009). Many tumors reactivate L1 through promoter hypomethylation, and tumor-specific L1 insertions can inactivate tumor suppressors, representing pathological counter acceleration (Lee et al., 2012). Low-level L1 activity has also been detected in aged liver, muscle, and other tissues, particularly in contexts of cellular stress or senescence.

The connection between L1 activity and inflammaging provides another mechanism by which this counter is "read." Cytosolic DNA, including L1 cDNA intermediates, activates the cGAS-STING pathway, triggering type I interferon production and inflammatory signaling (Miller et al., 2021). As L1 silencing weakens with age, increased retrotransposition generates more cytosolic DNA, activating chronic low-grade inflammation characteristic of aging tissues. Reverse transcriptase inhibitors have shown efficacy in reducing inflammation in aging models, consistent with L1 activity contributing to the inflammatory phenotype (De Cecco et al., 2019). The counter thus produces both cell-autonomous effects (genomic insertions) and non-cell-autonomous effects (inflammatory signaling).

5. COUNTER 4: TRANSCRIPTION-ASSOCIATED INDELS

5.1 Frameshift During Transcription: Quantifying the Error Rate

RNA polymerase II, like DNA polymerase, can slip on homopolymeric templates. This slippage generates transcripts with frameshift errors—insertions or deletions relative to the DNA template. Unlike replication errors that become permanent mutations, transcription errors are transient but their cumulative effects are substantial. Of the four counters proposed in this framework, transcription-associated indels have the least direct experimental support for a counting function. While transcription slippage is well-documented and quality control systems are characterized, the proposal that cells use transcription error burden as a functional readout of cellular state—rather than simply damage to be managed—remains more speculative than Counters 1-3, which have clearer evidence for threshold-dependent cellular responses.

Transcription insertions prevail in homopolymeric runs of A and T, while transcription deletions arise in more complex sequences; continuation of transcription after slippage depends on nucleotide complementarity within the RNA:DNA hybrid at the new template location (Traverse & Ochman, 2017). The rates are significant: transcription insertion rates

approach 10^-6 per nucleotide per transcription event at susceptible homopolymer sites, while transcription deletion rates are approximately 10^-7 per nucleotide but occur in more complex sequence contexts. A highly expressed gene transcribed 10,000 times per day generates dozens of frameshifted transcripts daily (Schwanhäusser et al., 2011). Unlike replication errors that affect all subsequent cell progeny, transcription errors affect only the proteins produced from that specific erroneous transcript, but given that each gene may be transcribed thousands of times, the population-level effect is substantial.

5.2 RNA Quality Control: The Surveillance System

Cells deploy multiple surveillance mechanisms to detect and eliminate aberrant transcripts (reviewed in Lykke-Andersen & Jensen, 2015). Nonsense-mediated decay (NMD) recognizes and degrades transcripts with premature stop codons, as occur with frameshift, and NMD requires translation to detect the aberrant stop codon, so the cell must first attempt (and fail) to make protein from the defective transcript. No-go decay triggers transcript degradation when ribosomes stall during translation; frameshift-induced stop codons or rare codon stretches can cause such stalling. Non-stop decay targets transcripts lacking stop codons entirely, though this is less relevant for frameshift errors. The Dom34/Hbs1 rescue pathway rescues stalled ribosomes and recycles them, limiting the damage from aberrant transcripts. These quality control systems consume cellular resources—ATP for degradation, ribosomal capacity tied up on defective transcripts, and chaperone time handling mistranslated proteins—and as transcription error burden increases with cellular age, the cost of quality control rises proportionally.

5.3 When Surveillance Fails: Proteome Corruption

Quality control systems have limited capacity, and under high transcription error load—as occurs in aged cells with accumulated damage to transcription machinery—aberrant transcripts escape surveillance. Frameshift-induced stop codons produce truncated proteins that may be partially functional, non-functional, or dominantly toxic. Novel open reading frames downstream of frameshifts encode peptide sequences never seen during immune tolerance, and these neoepitopes can trigger autoimmune responses or, in cancer, provide targets for immunotherapy. Even when aberrant proteins are recognized and degraded, the degradation process consumes resources and can overwhelm proteasome capacity. The proportion of aberrant transcripts increases with expression level, since highly expressed genes generate more total errors simply through more transcription events, and neurons, with their high transcriptional activity and long lifespans, accumulate particularly high transcription error burdens.

5.4 Expression-Level Scaling and Potential Adaptive Functions

This counter scales with transcriptional activity rather than cell division, making it complementary to division-coupled counters. Importantly, Counter 4 differs conceptually from Counters 1-3: transcription-associated indels do not accumulate genomically (each transcript is independent) but rather create a *rate* of proteome corruption that increases with cellular age as transcription machinery becomes less faithful and quality control systems become overwhelmed. This counter thus measures instantaneous error rate rather than cumulative genomic damage, providing complementary information about cellular state. While Counters 1-

3 record division history in DNA sequence, Counter 4 reflects the current functional burden on cellular proteostasis—a rate that correlates with but is mechanistically distinct from genomic counting.

Dividing cells with low transcription, such as early embryos and quiescent stem cells, experience low transcription error burden despite many divisions, whereas post-mitotic cells with high transcription, such as neurons and cardiomyocytes, accumulate high transcription error burden despite no divisions. Dividing cells with high transcription, such as activated lymphocytes and transit-amplifying cells, see both division-coupled and transcription-coupled counters advance rapidly. The biological consequence is age-related protein quality decline concentrated in highly active tissues: the very cells that work hardest, transcribing most genes most often, accumulate the most errors and suffer the earliest functional decline. This may explain why neurons and cardiomyocytes, despite being "protected" from division-coupled mutation, nonetheless show pronounced age-related dysfunction.

While transcription errors are generally considered pathological, controlled transcription variation might serve adaptive purposes. Low-level sequence variation in proteins could provide phenotypic buffering against environmental changes, representing a form of bet-hedging at the proteome level. Continuous low-level generation of neoepitopes might maintain immune vigilance, priming T cell populations against potential cancer-derived antigens. The UPR activation caused by transcription errors might pre-condition cells for subsequent stress, a form of hormesis. These speculative functions remain to be tested, but they suggest that the transcription error "counter" might be tuned to provide optimal, rather than minimal, error rates.

6. THE REPEAT-GUIDED SEGREGATION HYPOTHESIS

This section develops a mechanistic hypothesis for how repeat expansion could guide asymmetric stem cell division. The evidence supporting this model spans three tiers: established findings include ATR/CHK1-dependent nonrandom DNA segregation under replication stress and G4 structure formation by expanded repeats; extrapolated claims include the proposal that repeat-induced G4 structures engage this same ATR/CHK1 pathway; and speculative components include the integrated "Segregate and Sacrifice" model connecting these elements into a coherent cellular counting mechanism. We present this hierarchy explicitly because the integrated model, while consistent with available evidence, has not been directly tested as a unified system (Figure 2).

Figure 2: Repeat-Guided Segregation Mechanism

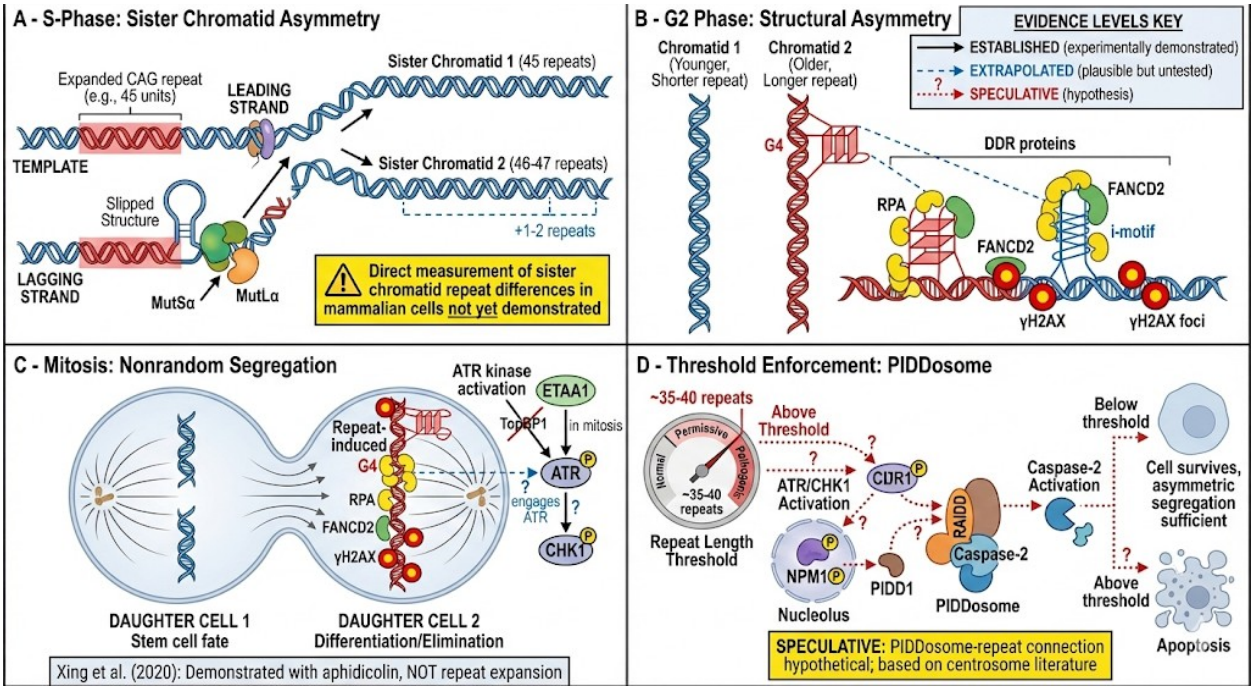

Figure 2. Molecular mechanism of repeat-guided asymmetric segregation. (A) S-phase: Replication of expanded CAG/CTG repeat creates sister chromatid asymmetry. The lagging strand template accumulates more slipped structures, processed by MMR into expansion bias. (B) G2 phase: G4 structures form on G-rich strand; i-motifs on C-rich strand. Differential structure stability creates distinct DDR protein recruitment. (C) Mitosis: ATR/CHK1 activation via ETAA1 (not TopBP1) directs nonrandom segregation. G4-bearing chromatid marked by RPA, FANCD2, γH2AX is directed to daughter destined for elimination. (D) Threshold enforcement: When repeat burden exceeds segregation capacity, ATR/CHK1 → NPM1 phosphorylation → PIDD1 release → caspase-2 activation → apoptosis.

6.1 Asymmetric Division and Strand-Specific Expansion

Stem cells must balance two competing imperatives: self-renewal to maintain the stem cell pool, and differentiation to generate functional progeny. Many stem cells achieve this balance through asymmetric division, producing one daughter that remains a stem cell and one that differentiates. The molecular mechanisms governing asymmetric division are well-characterized at the level of protein and organelle inheritance—Numb, Prospero, and other fate determinants segregate asymmetrically through interactions with the mitotic spindle (Knoblich, 2008). But a fundamental question remains: what distinguishes the two daughters at the level of their DNA?

The “immortal strand hypothesis,” proposed by John Cairns (Cairns, 1975; Cairns, 2006), suggested that stem cells preferentially retain template DNA strands across successive divisions, protecting themselves from replication-associated mutations. While this elegant hypothesis has generated decades of investigation, direct evidence remains elusive and

experimental tests have yielded mixed results (Winquist et al., 2014). We propose that repeat length differences between sister chromatids, generated during S-phase, provide one molecular basis for asymmetric DNA segregation. The chromatid with shorter repeats—the "younger" DNA—is preferentially retained in the stem cell, while the chromatid with longer repeats—the "older" DNA—is directed to the differentiating daughter.

During DNA replication, repeat expansion occurs preferentially on the newly synthesized strand, particularly during lagging strand synthesis. Templates with expanded repeats yield predominantly expansions when CAG serves as the lagging strand template, or predominantly deletions when CTG serves as the lagging strand template, demonstrating that replication direction determines mutation type (Panigrahi et al., 2002). This strand-specificity means that after S-phase, the two sister chromatids are not identical at repeat loci. Studies in *E. coli* have shown that small slipped structures recognized by the MMR system form primarily on the lagging strand of the replication fork (Schmidt et al., 2000); while this bacterial work demonstrates the principle of strand-specific structure formation, direct measurement of sister chromatid repeat length differences in mammalian cells remains an essential but unmet experimental need. If MMR processing of slipped structures in mammalian cells mirrors the bacterial mechanism, it would create a consistent length difference of approximately 1-2 repeat units between sister chromatids after each S-phase—but this prediction awaits validation through long-read single-cell sequencing of sister chromatids (Figure 2A).

6.1.1 Mammalian Evidence for Asymmetric Lesion Segregation

While sister chromatid repeat length asymmetry has not been directly measured in mammalian cells, substantial evidence demonstrates that DNA lesions persist through multiple cell divisions and segregate asymmetrically into daughter cells. Aitken et al. (2020) provided compelling evidence for this phenomenon in chemically-induced mouse liver tumors, observing chromosome-scale mutational strand biases spanning entire chromosomes (median 55 Mb segments). Their analysis demonstrated that >95% of mutation-causing lesions segregate for at least one mitosis before resolution.

Critically, this lesion segregation phenomenon is not specific to chemical mutagens. Aitken et al. demonstrated identical patterns in human cells exposed to UV light, tobacco smoke components (BPDE), and chemotherapeutic agents (temozolomide), establishing that asymmetric segregation is "a unifying property of exogenous mutagens" in mammalian cells. This universality suggests that asymmetric segregation is a fundamental cellular mechanism, not an artifact of specific damage types.

Several features of lesion segregation directly support the repeat-based counting model. First, persistence across divisions: DNA lesions template multiple rounds of replication, producing multi-allelic variation at 8% of mutated sites (Aitken et al., 2020). This demonstrates that damaged templates are not immediately repaired or eliminated but instead segregate through successive divisions—precisely the behavior we propose for repeat-expanded sister chromatids. Second, sister chromatid exchange detection: The mutational asymmetry patterns allowed fine mapping (~20 kb resolution) of sister chromatid exchange events, with a median of 27 SCEs per diploid genome. SCE frequency correlated with mutation load, suggesting that homologous recombination-mediated repair responds to accumulated damage burden. Third, differential cell fates based on segregation timing: Aitken et al. could infer the cell generation post-

mutagenesis from which tumors developed based on multi-allelic patterns. Tumors arising from first-generation daughter cells showed uniformly high multi-allelic rates; those from later generations showed segmental patterns reflecting progressive dilution of lesion-containing strands. This demonstrates that lesion inheritance patterns shape clonal expansion—a central prediction of the counting framework.

6.1.2 Multigenerational Tracking of Heritable DNA Damage

Panagopoulos et al. (2025) provide the most direct evidence to date for multigenerational damage inheritance and asymmetric segregation in mammalian cells. Using CRISPR-engineered cells with endogenously tagged 53BP1 and PCNA, they tracked DNA lesions through up to four cell generations in asynchronously growing human cells.

Several findings directly support the repeat-based counting model. Heritable damage induces sister cell asymmetry: DNA damage experienced in G1 induced marked heterogeneity between sister cells in the next generation, including differences in S-phase commitment and replication timing. The authors observe that "even a single, transient genotoxic event in the parental DNA may result in a persistent genomic scar that affects cell cycle commitment and replication timing in the next cell generation, leading to increased asymmetry between sister cells." This demonstrates that damage can be asymmetrically inherited, creating divergent cell fates—precisely the outcome we propose for repeat-expanded sister chromatids.

The key mechanistic insight from Panagopoulos et al. is that damage inheritance is not merely passive segregation but involves active recognition and processing by the DDR machinery. Their observation that "53BP1 nuclear bodies marking regions of inherited endogenous replication stress appeared primarily after cell division in G1 and were cleared as cells entered S phase" suggests a surveillance window during which cells assess inherited damage burden before committing to replication. We propose that repeat-expanded sequences engage this same surveillance mechanism through G4-induced replication stress, triggering asymmetric segregation when expansion exceeds threshold levels.

6.2 G4/i-Motif Asymmetry as the Segregation Signal

The structural asymmetry described in Section 2.4 provides the molecular tag that distinguishes sister chromatids. When expanded repeats replicate, the two daughter duplexes inherit different structural propensities based on which strand served as template. The chromatid where the G-rich strand was the lagging strand template accumulates more G4-stabilizing damage and recruits more DDR proteins, creating a readable difference that the cell machinery can exploit (Figure 2B).

This structural asymmetry manifests as differential protein recruitment. G4 structures recruit RPA to coat single-stranded regions exposed during replication stress. FANCD2, a key component of the Fanconi anemia pathway, localizes to G4-induced stalled forks. Phosphorylated H2AX (γH2AX) marks chromatin surrounding replication impediments. These proteins co-segregate with the more damaged chromatid through mitosis (Xing et al., 2020), providing a visible marker of repeat-induced stress that persists from S-phase through cell division.

The critical insight is that G4 formation is length-dependent and threshold-gated (Figure 3C). Below the pathogenic threshold, G4 structures are transient and resolved without consequence. Above the threshold, G4 persistence triggers sustained ATR signaling that marks the affected chromatid for segregation to the differentiating daughter. The pathogenic threshold in repeat expansion diseases may therefore represent not merely a toxicity boundary, but the point at which G4-mediated segregation can no longer protect the stem cell pool from accumulated repeat burden.

6.3 ATR/CHK1 as the Mitotic Segregation Executor

The ATR-CHK1 signaling axis, canonically associated with ssDNA break and the replication checkpoint, serves a distinct function in mitosis: directing the segregation of damaged chromatids to specific daughter cells. This mechanism, termed nonrandom DNA segregation (NDS), was directly demonstrated by Xing et al. (2020), who showed that replication stress increases asymmetric chromosome segregation from a baseline of approximately 7% to 15-18%, with damaged DNA preferentially inherited by one daughter cell that is subsequently eliminated. Importantly, Xing et al. induced replication stress using aphidicolin (a DNA polymerase inhibitor) and oncogene overexpression, not repeat expansion or G4 structures. Our proposal that repeat-induced replication stress engages this same pathway is mechanistically plausible—both create stalled forks and single-stranded DNA that activate ATR—but remains untested in the context of endogenous repeat expansion.

The molecular mechanism involves a mitosis-specific ATR activation pathway distinct from S-phase checkpoint signaling. During unperturbed S-phase, ATR is activated by TopBP1 at stalled replication forks. However, in mitosis, ETAA1 serves as the primary ATR activator (Bass et al., 2016), and Claspin—the canonical CHK1 activator in S-phase—is degraded at mitotic entry through SCF-βTrCP-mediated ubiquitination (Mamely et al., 2006). This cofactor switch ensures that mitotic ATR/CHK1 signaling produces different outputs than S-phase checkpoint activation: rather than arresting the cell cycle, mitotic ATR/CHK1 directs chromosome segregation.

Three lines of evidence establish ATR/CHK1 as essential for repeat-guided segregation. First, chemical inhibition of ATR abolishes nonrandom segregation, reducing asymmetric inheritance to baseline levels even under replication stress. Second, CHK1 inhibition produces the identical phenotype, placing CHK1 downstream of ATR in this pathway. Third, ATM inhibition has no effect on NDS, demonstrating pathway specificity—the double-strand break response kinase ATM is dispensable, while the replication stress kinase ATR is required.

The biological consequence is a “segregate and sacrifice” strategy: the daughter cell receiving the G4-bearing, DDR-marked chromatid undergoes cell cycle arrest or apoptosis, while the daughter receiving the cleaner chromatid continues proliferating. Xing et al. (2020) describe how the damaged daughter undergoes “cell-cycle arrest and cell death”—we term this protective logic “Segregate and Sacrifice” to capture its essence. By partitioning damage asymmetrically and eliminating the damaged lineage, stem cells preserve genomic integrity across generations despite continuous repeat expansion (Figure 2C).

6.3.1 Direct Evidence for ATR/CHK1 in Damage Segregation

These phenotypes are consistent with ATR/CHK1 normally functioning to: (a) coordinate damage recognition with segregation timing; (b) ensure asymmetric partitioning of damaged chromatids; and (c) activate threshold checkpoints when damage burden is excessive.

Recent work has revealed that cells possess multiple parallel mechanisms for asymmetric damage inheritance, with different damage types engaging distinct pathways. Ferrand et al. (2025) demonstrated that UV-induced dipyrimidine damage (CPDs) segregates non-randomly through mitosis in human cells, with an average 15% bias between daughter cells—significantly exceeding random expectation and approaching the theoretical maximum given initial sister chromatid asymmetry. Critically, this UV damage segregation is **ATR-independent**: ATR inhibition does not reduce the CPD segregation bias, distinguishing this mechanism from the ATR/CHK1-dependent pathway described by Xing et al. (2020) for replication stress-induced damage.

Instead, UV damage segregation operates through a chromatin-marking pathway involving two sequential mechanisms: (1) PARP/HPF1-mediated ADP-ribosylation on H3S10, which directly competes with mitotic phosphorylation at this residue; and (2) incorporation of newly synthesized histones at damage sites, which are poor substrates for Aurora B kinase. Both mechanisms prevent local H3S10 phosphorylation, creating a chromatin "bookmark" that guides asymmetric segregation of damaged chromosomes. Functionally, daughter cells inheriting more UV damage show reduced expression of the stemness marker Sox2 in iPSCs, suggesting this mechanism may protect stem cell compartments by preferentially directing damage to differentiating daughters.

Multiple damage-type-specific inheritance pathways—ATR/CHK1 for replication stress, PARP/DDB2/chromatin marking for UV damage—converge on protecting lineages from accumulated damage (Ferrand et al., 2025).

6.3.2 Genomic Evidence for Sister Chromatid Asymmetry

Independent support for the molecular basis of sister chromatid asymmetry comes from Zhang et al. (2026), who analyzed 2,588 cancer genomes and identified a widespread phenomenon they term "breakage-replication/fusion." Their key finding is that replication of a single DSB end generates two non-identical "sister" DNA ends. Adjacent parallel breakpoints—the genomic signature of sister DNA end processing—account for 12% of all rearrangement junctions (35,422 of 296,088 junctions analyzed).

Critically, Zhang et al. demonstrated that leading versus lagging strand synthesis of DSB ends creates predictably different products. The size differences between adjacent parallel breakpoints (ranging up to 20 kb, with most within several hundred base pairs) reflect the extent of 5'-resection of ancestral DSB ends. Strand-coordinated base substitutions (APOBEC-mediated kataegis) restricted to the offset region between parallel breakpoints directly established their origin from staggered DSB ends with 3' overhangs.

This mechanism is directly analogous to how we propose MMR-mediated processing of slipped structures creates repeat length differences between sister chromatids. In both cases:

1. **Replication creates asymmetry**: Leading and lagging strand synthesis produce non-identical products from identical templates.

2. **Asymmetry persists through clonal expansion**: Zhang et al. showed that asymmetric outcomes (adjacent parallel breakpoints) are detectable in cancer genomes, demonstrating persistence through multiple cell divisions.

3. **Predictable strand bias**: The direction of asymmetry is determined by template strand identity, not random variation.

While Zhang et al. describe breakage-replication/fusion as a mechanism of genome rearrangement, we propose that cells have co-opted this inherent strand asymmetry for functional purposes: counting cellular divisions and guiding developmental fate decisions through repeat-guided segregation.

6.4 Threshold Enforcement via the PIDDosome

The ATR/CHK1 pathway directs segregation, but what enforces the threshold—ensuring that cells exceeding repeat limits are eliminated? We propose—as a hypothesis requiring direct experimental validation—that the PIDDosome may provide threshold enforcement for the repeat counter, potentially activated when repeat-induced replication stress exceeds cellular capacity for asymmetric segregation.

The PIDDosome is a multiprotein complex comprising PIDD1 (p53-induced death domain protein 1), RAIDD, and caspase-2 that triggers apoptosis in response to supernumerary centrosomes and other cellular stresses (Fava et al., 2017). Critically, PIDDosome activation is regulated by NPM1 (nucleophosmin), which sequesters PIDD1 in the nucleolus under normal conditions. ATR/CHK1 signaling phosphorylates NPM1, releasing PIDD1 to form the death-inducing complex (Hiregange et al., 2020). While this ATR-NPM1-PIDDosome connection is established for general DNA damage, its specific engagement by repeat-induced replication stress has not been directly demonstrated.

The threshold enforcement model gains molecular specificity from PIDD1's intrinsic two-stage autoprocessing (Tinel et al., 2007). The first cleavage at S446 generates PIDD-C (51 kDa), which recruits RIP1 and NEMO to activate NF-κB and promote survival. Only subsequent cleavage at S588 to PIDD-CC (37 kDa) enables RAIDD recruitment and caspase-2 activation. This sequential processing creates a built-in delay between damage sensing and death execution: moderate stress activates survival pathways through PIDD-C, while sustained or excessive stress tips the balance toward PIDD-CC accumulation and apoptosis. Importantly, Burigotto et al. (2021) demonstrated that local concentration of PIDD1 at centrosomes—achieved through clustering of damage-bearing structures—is required to exceed the activation threshold. Only the PIDD1 precursor localizes to centrosomes via ANKRD26 interaction; autoprocessing is constitutive, but PIDD-CC must accumulate above a concentration threshold for PIDDosome assembly. This provides a molecular precedent for the concentration-dependent threshold enforcement we propose operates at expanded repeats: just as clustered centrosomes locally concentrate PIDD-CC above the activation threshold, accumulated G4 structures at expanded repeats could similarly concentrate DDR proteins above a signaling threshold.

If this hypothesis is correct, the architecture would create a natural threshold mechanism for the repeat counter. At low repeat lengths, G4 structures would be transient and ATR activation minimal—NPM1 would remain unphosphorylated and sequester PIDD1. At intermediate lengths, sustained ATR/CHK1 signaling would enable asymmetric segregation to partition damage, but NPM1 phosphorylation would remain below the threshold for PIDD1 release. At pathogenic lengths, ATR/CHK1 capacity would be overwhelmed—the cell could not adequately segregate all the damage—and maximal NPM1 phosphorylation would release PIDD1 to trigger apoptosis via caspase-2 (Figure 2D).

Candidate threshold sensors include the p53-p21 axis, PIDD1 autoprocessing state, and mitotic checkpoint proteins, all of which can translate accumulated damage into binary cell fate decisions.

6.5 Histone Inheritance and Comparison to Known Mechanisms

The distinction between template and nascent strands is reinforced by asymmetric histone inheritance. Parental H3-H4 tetramers are assembled into nucleosomes onto both leading and lagging strands, with a slight preference for lagging strands (Yu et al., 2018). The machinery mediating this asymmetric distribution has been characterized: on the leading strand, DPB3/DPB4 transfer parental histones, while on the lagging strand, MCM2 performs this function. Mrc1 (CLASPIN in humans) coordinates this process, and Mrc1 mutants show disrupted segregation of parental histones along with clonal and asymmetric loss of H3K9me-mediated gene silencing (Charlton et al., 2024). The biological consequence is that sister chromatids differ not only in repeat length but in epigenetic age.

The repeat-guided segregation hypothesis complements rather than replaces known asymmetric division mechanisms. The adaptor protein Numb operates at the protein level during mitosis, while repeat-guided segregation operates at the DNA level during S-phase—the two mechanisms could function synergistically. Old and new centrosomes are preferentially inherited by stem cell and differentiating daughters respectively (Yamashita et al., 2007), sharing the feature of using pre-existing differences to guide segregation. In epithelial stem cells, division plane orientation relative to tissue architecture determines fate through apical-basal polarity. Repeat-guided segregation could provide a cell-autonomous mechanism that reinforces these polarity-based fate decisions. The key distinction is that repeat-guided segregation provides a molecular memory mechanism—the DNA itself carries information about division history that can influence subsequent fate decisions.

6.6 CENP-A Loading and Stem Cell Niche Variations

The distinction between sister chromatids is further amplified in G2 phase through asymmetric CENP-A loading. CENP-A is the centromeric histone variant that defines centromere identity and recruits kinetochore proteins for spindle attachment. Studies in *Drosophila* male germline stem cells have demonstrated spatiotemporally regulated asymmetric components that ensure biased sister chromatid segregation, with sister centromeres differentially enriched with proteins involved in centromere specification (Ranjan et al., 2019; Dattoli et al., 2020). Sister chromatids destined for the stem cell harbor more CENP-A, assemble more kinetochore proteins, and capture more spindle microtubules (Carty et al., 2021). Critically, CENP-A loading occurs in G2 phase—after replication is complete and repeat expansion has occurred. We

propose that the chromatid with shorter repeats and more parental histones presents a more favorable substrate for CENP-A deposition, resulting in asymmetric CENP-A levels that bias spindle capture and segregation.

Different stem cell populations may weight repeat-guided segregation differently. Hematopoietic stem cells rarely divide but must maintain lifelong self-renewal, creating strong selective pressure for asymmetric segregation to preserve low-count genomes. Intestinal stem cells divide frequently and experience high mutagenic exposure; the colon's vulnerability to MSI-driven cancer suggests the repeat counter is particularly active here. Neural stem cells transition from symmetric expansion to asymmetric neurogenic divisions during development, and this shift could be triggered by repeat counter thresholds. Muscle satellite cells are quiescent in adults but activated for repair; satellite cell exhaustion in muscular dystrophies could reflect accelerated counter accumulation during repeated activation cycles.

6.7 Integrated Model: The "Segregate and Sacrifice" Strategy

The complete model operates through four sequential phases. During S-phase, replication generates sister chromatid asymmetry: the template-strand chromatid accumulates shorter repeats and more parental histones, while the nascent-strand chromatid accumulates longer repeats, more newly synthesized histones, and G4-induced DDR protein recruitment. During G2 phase, CENP-A loading amplifies this asymmetry based on chromatin state. During M-phase, ATR/CHK1-dependent segregation partitions DNA by counter value, with the stem cell receiving the shorter-repeat chromatid and the differentiating daughter receiving the longer-repeat chromatid. When repeat length exceeds segregation capacity, the PIDDosome enforces threshold-dependent elimination via NPM1-PIDD1-caspase-2 activation.

This "Segregate and Sacrifice" strategy—our term for the protective logic described by Xing et al. (2020)—reveals that trinucleotide repeats serve dual functions: as a clock counting divisions through progressive expansion, and as a compass guiding segregation by creating sister chromatid asymmetry. The G4/i-motif structures provide the readout mechanism, ATR/CHK1 provides the segregation machinery, and the PIDDosome provides threshold enforcement. By partitioning damage asymmetrically and eliminating the damaged lineage, stem cells preserve genomic integrity across generations despite continuous repeat expansion. The specific testable predictions arising from this model are detailed in Section 9.

6.8 Epigenetic Asymmetry: 5-Hydroxymethylcytosine as a Parallel Mechanism

The repeat-based asymmetry described above operates in parallel with an epigenetic asymmetry system based on 5-hydroxymethylcytosine (5hmC). During zygotic development, TET3 oxidizes paternal 5-methylcytosine to 5hmC, creating an epigenetic distinction between parental genomes (Gu et al., 2011). Critically, after DNA replication, 5hmC is inherited by only ONE sister chromatid because DNMT1/UHRF1 cannot recognize hemi-hydroxymethylated CpG sites (Inoue & Zhang, 2011). This creates replication-dependent asymmetry: one daughter chromatid retains 5hmC marks while the other loses them through passive dilution.

The molecular reader connecting this epigenetic asymmetry to cellular machinery is MSH6, a mismatch repair protein already implicated in repeat instability. Proteomics screens identified MSH6 as preferentially binding 5hmC over 5mC (Iurlaro et al., 2013). This positions MMR machinery to read BOTH repeat-based AND epigenetic asymmetry through the same molecular

sensor. The convergence on MSH6 suggests that these are not independent systems but components of an integrated asymmetry-detection network.

A critical mechanistic link connects repeat-induced damage to epigenetic asymmetry. ATR kinase, activated by G4-induced replication stress at expanded repeats, directly phosphorylates TET3, enhancing its enzymatic activity (Jiang et al., 2017). This means repeat expansion (Counter 1) causally drives 5hmC generation: longer repeats → more ATR activation → more TET3 activity → more 5hmC asymmetry. The two asymmetry mechanisms are not merely parallel—they are causally connected through the ATR-TET3 axis.

6.9 Metabolic Integration: The Three-Hub Model

Sister chromatid asymmetry generation and interpretation is controlled by cellular metabolic state through three interconnected hubs operating on different temporal scales. The first hub centers on α-ketoglutarate (α-KG). TET enzymes and Jumonji C domain-containing histone demethylases (KDM4, KDM6 families) share cofactor requirements: α-KG, Fe(II), and molecular oxygen. This creates coordinated metabolic control where α-KG availability simultaneously regulates DNA demethylation (5mC→5hmC) and histone demethylation (H3K9me3, H3K27me3 removal). IDH1/2 mutations produce 2-hydroxyglutarate (2-HG), which competitively inhibits both enzyme classes (Xu et al., 2011). Furthermore, α-KG controls DNMT3A/3B expression through the LIF-Stat3-Otx2 pathway (Betto et al., 2021), meaning a single metabolite coordinates DNA demethylation, histone demethylation, AND de novo methyltransferase expression. Vitamin C enhances both TET and KDM activity by maintaining iron in the Fe(II) state (Blaschke et al., 2013).

The second hub centers on NAD+. PARP1 senses DNA damage by consuming NAD+ to generate poly(ADP-ribose) chains that recruit repair factors and the chromatin remodeler ALC1; serine ADP-ribosylation specifically marks nucleosomes for ALC1-dependent remodeling (Sellou et al., 2016; Mohapatra et al., 2021). Sirtuins (SIRT1, SIRT6, SIRT7) also require NAD+ for deacetylase activity, creating competition: high DNA damage → PARP1 activation → NAD+ depletion → Sirtuin inhibition. NAD+ levels decline with age in humans, correlating with accumulated DNA damage (Massudi et al., 2012). This positions NAD+ as a quantitative “damage counter” integrating lifetime DNA damage burden.

The third hub centers on succinyl-CoA. Nuclear α-ketoglutarate dehydrogenase generates succinyl-CoA for histone succinylation by KAT2A (Wang et al., 2017). The functional importance of this pathway is demonstrated by the finding that succinyl-CoA synthetase deficiency causes widespread hypersuccinylation and chromatin landscape changes (Lancaster et al., 2023). We propose that SIRT7, recruited to DNA damage sites in a PARP1-dependent manner, may serve as a bridge between the NAD+ hub and chromatin state through its H3K122 desuccinylase activity (Li et al., 2016).

6.10 Temporal Hierarchy of Counter Function

These metabolic hubs operate on different temporal scales, creating a hierarchical damage-integration system that filters transient fluctuations from sustained signals.

The fast response operates on a timescale of seconds to minutes. PARP1 immediately consumes NAD+ to generate PAR chains, ALC1 relaxes chromatin for repair access, and SIRT7 is recruited

for post-repair chromatin condensation. This fast hub responds to individual damage events and recovers quickly if damage is isolated. The intermediate response operates on a timescale of minutes to hours. Sirtuins modulate histone acetylation based on cumulative NAD+ status, while KAT2A adds succinylation marks based on TCA cycle flux. This intermediate hub integrates recent metabolic history and begins to register repeated stress events. The slow response operates on a timescale of hours to days. TET and KDM enzymes modify DNA and histone methylation based on sustained α-KG availability. These changes require multiple cell cycles to fully establish and create semi-stable epigenetic states that persist through division.

The stable memory system operates on a timescale of days to years. DNA methylation patterns maintained by DNMT1 provide long-term cellular memory, changed only through active TET-mediated demethylation or passive replication-dependent dilution during cell division.

The functional consequence of this temporal hierarchy is that single damage events activate only the fast hub and recover without engaging slower systems, while repeated or sustained damage progressively engages intermediate and slow hubs, eventually producing stable epigenetic changes that bias cell fate toward differentiation or senescence.

6.11 Critical Developmental Caveat: Counter Acquisition at Zygotic Genome Activation

The temporal hierarchy and counter mechanisms described above apply to checkpoint-competent cells. Early embryonic development operates under fundamentally different rules that must be explicitly acknowledged, because pre-implantation embryos have abbreviated cell cycles lacking conventional G1 and G2 gap phases, with checkpoint mechanisms that are either absent or severely attenuated. In mouse embryos, γH2AX—the canonical marker of DNA double-strand break signaling—is NOT detected in 1-cell or 2-cell stage embryos after ionizing radiation, despite the presence of ATM and DNA-PK proteins (Yukawa et al., 2007). The kinases are present, but the substrate phosphorylation that defines checkpoint activation does not occur. Consequently, G2/M checkpoint and DNA repair mechanisms have insufficient function in these early stages, causing hypersensitivity to genotoxic stress.

In zebrafish, the Chk1 arm of the DNA damage response is inactive before the mid-blastula transition (MBT), even though Chk2 can be activated by double-strand breaks (Zhang et al., 2014). Chk1 gain-of-function at MBT underlies the dramatic cell cycle remodeling that occurs at this transition. Zygotic genome activation (ZGA) is regulated independently by the nuclear:cytoplasmic ratio, and ZGA itself can trigger widespread chromosome damage through topoisomerase II activity (Butuči et al., 2015).

This has critical implications for the counter hypothesis. In 1-2 cell mouse embryos, ATM is present but γH2AX is absent, indicating that counters are NOT active. During the pre-ZGA/pre-MBT phase, Chk1 remains inactive and counter sensing appears minimal. At ZGA/MBT, Chk1 function is acquired, cell cycles lengthen, and counter acquisition begins. Post-ZGA embryos have full checkpoint function with counters active, progressing to adult somatic cells with complete DDR and counter maintenance.

Counter acquisition occurs at ZGA/MBT, not at fertilization. The three-hub temporal model becomes operative only after checkpoint competence is established. Damage occurring before ZGA may not trigger counter mechanisms and could be “grandfathered in” to the cellular genome without appropriate checkpoint responses. This explains why early embryos are

hypersensitive to radiation—damage occurs but checkpoints do not arrest the cycle to allow repair.

This developmental window of checkpoint deficiency may have important implications for disease. Repeats that expanded during pre-ZGA divisions without triggering checkpoint responses could establish pathogenic alleles that would otherwise have been eliminated—potentially explaining why certain inherited repeat lengths cause disease while others do not. This represents a vulnerability in the counting framework that may itself be under evolutionary selection, with counter acquisition timing optimized to balance developmental plasticity against genomic integrity.

This developmental caveat also raises the question of whether cancer cells can revert to an "embryonic-like" checkpoint-deficient state. Loss of Chk1 function in some cancers may represent such a reversion, allowing tolerance of replication stress at the cost of genomic instability. Therapeutic strategies that induce replication stress (PARP inhibitors, ATR inhibitors) may be particularly effective against such cells precisely because they lack the checkpoint capacity to respond.

6.12 Evidence Classification and Testable Predictions

To facilitate critical evaluation, we classify the evidence supporting each component of the integrated model:

Tier 1—Established biochemical facts: TET3 converts paternal 5mC→5hmC in zygotes (Gu et al., 2011); 5hmC creates replication-dependent sister chromatid asymmetry (Inoue & Zhang, 2011); MSH6 preferentially binds 5hmC (Iurlaro et al., 2013); ATR phosphorylates TET3 during DNA damage response (Jiang et al., 2017); α-KG is required cofactor for TET and KDM enzymes; 2-HG inhibits TET and KDM enzymes (Xu et al., 2011); PARP1 consumes NAD+ at damage sites; SIRT7 is PARP1-dependent H3K122 desuccinylase (Li et al., 2016); NAD+ declines with age in humans (Massudi et al., 2012); Chk1 is inactive pre-MBT, acquired at MBT (Zhang et al., 2014); γH2AX not formed in 1-2 cell mouse embryos (Yukawa et al., 2007); strand-specific expansion bias depends on replication direction (Kang et al., 1995).

Tier 2—Evidence consistent with the counting hypothesis: 5hmC asymmetry biases sister chromatid segregation; α-KG hub coordinates multiple asymmetry mechanisms; NAD+/PARP constitutes parallel metabolic hub; counter exhaustion triggers senescence; ATR-TET3 axis links repeat expansion to epigenetic asymmetry. These observations are also consistent with alternative interpretations (see Section 10.3).

Tier 3—Evidence specifically supporting counting over alternatives: The two-threshold structure of repeat diseases (permissive vs. pathogenic) corresponds to biologically meaningful division numbers; ATR-TET3 phosphorylation provides causal integration between damage sensing and epigenetic modification; cross-species calibration of mutation rates to lifespan suggests selection on counter mechanisms. However, these observations remain consistent with alternative interpretations under appropriate auxiliary assumptions.

Tier 4—Novel predictions requiring testing: (1) ATR inhibition should reduce TET3-mediated 5hmC generation during replication stress; (2) MSH6 ChIP-seq should show co-localization at repeat loci AND 5hmC-enriched regions; (3) NAD+ supplementation should enhance

asymmetric division capacity in stem cells; (4) SIRT7 depletion should disrupt chromatin condensation timing after DNA repair; (5) Combined TET inhibition plus repeat expansion block should synergistically impair asymmetric division; (6) Pre-ZGA embryos should tolerate DNA damage without engaging counter mechanisms; (7) Sister chromatid segregation bias should function as a counter-reading mechanism; (8) Counter acquisition should occur specifically at ZGA.

7. TISSUE-SPECIFIC COUNTER RATES

7.1 Quantified Mutation Rates Across Tissues

Different tissues accumulate indels at dramatically different rates, reflecting differences in cell division frequency, mutational exposure, and repair capacity. Analysis of tissue-specific mutation patterns has identified cell types bearing unique mutation signatures characterized by enrichment in active chromatin, with somatic mutation landscapes that enhance tumor transformation risk in specific cell populations (Franco et al., 2019).

Quantified mutation rates from published studies reveal substantial tissue-specific variation. Using whole-genome sequencing of clonal organoid cultures, Blokzijl et al. (2016) found that colon, small intestine, and liver adult stem cells all accumulate approximately 40 single nucleotide variants (SNVs) per year—despite large differences in cancer incidence among these tissues. Subsequent studies have extended these measurements: hematopoietic stem cells accumulate approximately 10-17 SNVs per year (Lee-Six et al., 2018), while neurons accumulate approximately 15-17 SNVs per year despite their post-mitotic status (Lodato et al., 2018). These rates reflect SNVs rather than indels specifically, but indel rates show similar tissue-specific patterns with additional modulation by repeat content and MMR status.

7.2 Counter Dominance Patterns by Tissue

Different counters dominate in different tissues based on tissue characteristics. In the colon, Counter 2 (homopolymer frameshifts) dominates, as high division rate plus abundant poly-A tracts in tumor suppressors such as APC and TGFBR2 (Markowitz et al., 1995) creates vulnerability to MSI-driven carcinogenesis. In the brain, Counter 3 (L1 insertions) and Counter 4 (transcription indels) dominate, since post-mitotic neurons cannot accumulate division-coupled mutations but experience continuous transcription-associated and retrotransposon-mediated changes. In the blood, clonal selection amplifies counter effects, as clonal hematopoiesis of indeterminate potential (CHIP) demonstrates how driver mutations accelerate counter readout through clonal expansion. In the bladder, counter effects are modulated by selection rather than mutation rate; men have significantly more truncating driver mutations in RBM10, CDKN1A and ARID1A than women despite similar levels of non-protein-affecting mutations (Calvet et al., 2025). This demonstrates that the same counter rate can produce different cancer risks depending on selection pressure on counter outputs.

7.3 Cross-Species Constraint on Counter Values

Perhaps the most striking evidence for the counting framework comes from cross-species comparisons. Analysis of somatic mutation rates across 16 mammalian species revealed that despite widely different life histories—including variation of approximately 30-fold in lifespan and 40,000-fold in body mass—the somatic mutation burden at the end of lifespan varied only

by a factor of approximately 3 (Cagan et al., 2022). An important caveat is that this study measured single nucleotide variants (SNVs) in intestinal crypt stem cells rather than repeat expansions or indels specifically; whether indel accumulation shows the same cross-species constraint remains to be directly tested.

This remarkable constraint suggests that end-of-life mutation burden—the counter's terminal value—is under strong evolutionary selection. Species have tuned their mutation rates to match their lifespans, maintaining approximately constant counter endpoints regardless of how fast they age.

Alternative interpretations warrant consideration. The mutation rate scaling could reflect metabolic rate differences (higher metabolism → more oxidative damage → faster mutation accumulation), body temperature variation, or normalization to cell division rates rather than direct selection on counter mechanisms. Distinguishing these possibilities requires measuring repeat expansion rates specifically—not just SNVs—across species with controlled cell division rates. If repeat expansion rates scale inversely with lifespan independently of division rate, this would support direct selection on the counting mechanism.

The implication, if the counting interpretation is correct, is profound: the indel counters would not merely track age but determine it, with species lifespan set by the rate at which counters run and the thresholds at which they trigger terminal events.

7.4 Illustrative Quantitative Framework

To facilitate quantitative reasoning about the counter framework—and to address the challenge of formalizing inherently probabilistic biological processes—we present an illustrative mathematical structure. This framework is intended as a conceptual scaffold for generating testable predictions, not as a validated quantitative model. The specific parameter values are approximations derived from published data; rigorous parameterization awaits systematic experimental measurement.

The simplest formulation treats counter value as a function of cell divisions and tissue-specific parameters:

$$Counter(t) = \Sigma\ divisions \times base\_rate \times tissue\_factor \times (1 / repair\_efficiency)$$

Where base_rate represents the intrinsic mutation generation rate per division (approximately 1–2 mutations per cell division in normal human tissues, extrapolated from SNV accumulation rates reported by Blokzijl et al., 2016 and Werner et al., 2020); tissue_factor captures tissue-specific modulation reflecting proliferation rate, transcriptional activity, and chromatin accessibility (ranging from approximately 0.5 in quiescent tissues to approximately 2.0 in highly proliferative epithelia); and repair_efficiency reflects MMR and other repair pathway function (1.0 for normal function, decreasing to approximately 0.01 in complete MMR deficiency, yielding the observed ~100-fold acceleration in CMMRD).

This formulation generates specific, testable predictions. For colorectal epithelium with approximately 1 division per day and normal MMR: Counter(70 years) ≈ 25,550 divisions × 1.5 × 1.5 × 1.0 ≈ 57,000 potential counter increments. With complete MMR deficiency (repair_efficiency = 0.01): the same tissue would reach Counter ≈ 5,700,000—explaining why

CMMRD patients develop cancer in childhood rather than late adulthood. The ~100-fold difference in counter rate predicts ~100-fold acceleration of cancer onset, consistent with the observed median cancer age of approximately 9 years in CMMRD (Ercan et al., 2024) versus 65+ years in sporadic colorectal cancer.

For cross-species comparisons, the framework predicts that longer-lived species should show proportionally lower base_rates or higher repair_efficiency to maintain similar end-of-life counter values. The Cagan et al. (2022) finding that end-of-life mutation burden varies only ~3-fold across species with 30-fold lifespan variation is consistent with this prediction, suggesting evolutionary tuning of these parameters.

We emphasize that this framework is illustrative rather than definitive. The actual relationship between counter value and biological outcome is almost certainly more complex, involving threshold effects, tissue-specific readout mechanisms, and interactions between counter types. Nevertheless, even this simplified formulation demonstrates how the conceptual framework can be translated into quantitative predictions suitable for experimental test.

7.5 Why Different Tissues Get Different Cancers

The tissue-specificity of cancer can be understood through counter dominance patterns. Colorectal cancer arises in a tissue where high division rate drives rapid counter accumulation, and the MSI pathway exploits Counter 2; familial cases such as Lynch syndrome show accelerated Counter 2. Glioma arises from dividing glial cells rather than post-mitotic neurons, involving Counter 1 (repeat expansion) and selection on IDH mutations. Leukemia demonstrates the interaction between Counter 1 and clonal selection, as CHIP studies show that different drivers accelerate counters at different rates. Bladder cancer shows male predominance that reflects selection differences rather than mutation rate differences, with environmental exposure such as smoking accelerating Counter 2. Breast cancer is modulated by hormonal exposure that affects counter rates, with parity affecting Counter 1 through epithelial proliferation history.

The same fundamental counting mechanisms operate across tissues, but tissue-specific division rates, counter dominance, and selection pressures create distinct cancer vulnerability profiles.

8. CANCER AS COUNTER ESCAPE

8.1 The Multi-Counter Escape Requirement

If cells employ multiple parallel counters, cancer must escape all of them. A cell that crosses one counter threshold but remains subject to others will still face fate enforcement. Only by disabling multiple counting systems can a cell achieve unlimited proliferation. This multi-hit requirement is fundamental to cancer biology, first established by Knudson's two-hit model for retinoblastoma (Knudson, 1971) and later generalized to multi-step carcinogenesis in colorectal cancer (Fearon & Vogelstein, 1990).

This explains why cancer requires multiple driver events. Each driver may disable a different counter or counter-response pathway:

Table 1:

Counter Escape	Mechanism	Driver Example
Repeat ATR/ATM threshold (DDR)	DDR mutation	ATM loss
Repeat ATR/CHK1 segregation pathway disruption	ATR mutation	ETAA1 loss
Repeat PIDDosome threshold enforcement	Inactivation	PIDD1 mutation, caspase-2 loss
Frameshift tumor suppressor sensing	Frameshift	APC, TP53
Telomere limit	Telomerase activation or ALT	TERT promoter, ATRX/DAXX loss
Immune antigen surveillance	Antigen loss	B2M frameshift
L1 silencing	Global demethylation	DNMT mutations

Cancer cells often show evidence of multiple counter escapes: TERT promoter mutations for immortalization, TP53 frameshift for checkpoint bypass, B2M frameshift for immune evasion, and—we now propose—disruption of the ATR/CHK1-PIDDosome axis for repeat counter escape. The prediction that PIDD1 or caspase-2 mutations should be enriched in cancers arising from tissues with high repeat burden remains to be systematically tested.

8.2 Counter Escape and Telomere Maintenance Mechanism Selection

The telomere limit represents a critical counter that all cancers must escape to achieve unlimited proliferation. However, the mechanism by which this counter is disabled varies systematically with how other counters have been escaped, revealing an unexpected interdependence between counter pathways.

Tumors that escape through Counter 2 (MSI pathway) face a constraint: frameshift mutations in MRE11's poly(T)11 tract—occurring in the majority of MSI-H tumors as detected by targeted microsatellite analysis—disrupt the MRN complex essential for ALT (Giannini et al., 2002; Jiang et al., 2005). These tumors are effectively locked into telomerase-dependent immortalization through TERT promoter mutations or amplification. Conversely, tumors with intact MRN—including most sarcomas and neural tumors arising through chromatin dysfunction (ATRX/DAXX loss)—retain access to the ALT pathway, explaining the tissue-specific distribution of telomere maintenance mechanisms: carcinomas predominantly use telomerase (85-95%) while sarcomas frequently use ALT (25-60%) (Heaphy et al., 2011).

This pattern generates a testable prediction: the correlation between MSI status and telomerase (rather than ALT) dependence should be disrupted in tumors where MRE11 is protected from frameshift (e.g., engineered MSI tumors with synonymous mutations that shorten the poly(T)11 tract).

A striking third category of tumors—particularly common in neuroblastoma (approximately 50% of cases)—lack any detectable telomere maintenance mechanism (Koneru et al., 2020).

These TMM-negative tumors show dramatically superior survival and frequently undergo spontaneous regression. Similar TMM-negative subsets have been identified in osteosarcoma (approximately 18% of cases, with favorable survival; P=0.05 vs TMM-positive) and other pediatric tumors. The biological significance of this category is profound: it demonstrates that telomere maintenance is not merely correlated with malignancy but is mechanistically required for sustained tumor growth. TMM-negative tumors represent a natural experiment in incomplete counter escape.

This observation provides independent evidence for the counter framework: tumors that have escaped other counters but not yet escaped the telomere limit face an inevitable replicative barrier. TMM-negative status likely represents tumors caught between counter escapes—having acquired driver mutations but not yet achieving immortalization. The favorable prognosis reflects this incomplete escape: the telomere counter remains intact and enforces eventual proliferative exhaustion. Importantly, TMM-negative tumors in neuroblastoma show significantly higher overall survival relative to both TERT-high and ALT-positive tumors ($P < 0.01$ for both comparisons), suggesting that telomere counter escape—regardless of mechanism—confers aggressive phenotype.

The relationship between counter escape and TMM selection also explains therapeutic resistance patterns. Telomerase inhibitor treatment can select for rare clones (approximately 2.5×10^{-7} frequency) that switch to ALT, particularly in ATM-null backgrounds (Hu et al., 2012). This switching is possible only in tumors where MRN remains intact—a condition more likely in CIN-pathway carcinomas than in MSI-pathway tumors. The counter escape history thus constrains therapeutic options: MSI tumors may be particularly vulnerable to telomerase inhibition precisely because their MRE11 frameshifts have closed the ALT escape route.

8.3 Clonal Hematopoiesis: Watching the Counters in Real-Time

Clonal hematopoiesis of indeterminate potential (CHIP) provides a unique window into counter dynamics. In CHIP, hematopoietic stem cells acquire driver mutations that confer proliferative advantage, leading to clonal expansion detectable in peripheral blood.

Longitudinal studies have shown that 92.4% of clones expand at a stable exponential rate, with different mutations driving substantially different growth rates ranging from 5% per year for DNMT3A and TP53 to more than 50% per year for SRSF2 (Fabre et al., 2022). This 10-fold variation in expansion rate within the same tissue demonstrates that drivers differentially affect counter dynamics. SRSF2 mutations, affecting splicing, may accelerate Counter 4 (transcription-associated errors). DNMT3A mutations, affecting methylation, may affect Counter 3 (L1 silencing).

Positive selection, not drift, is the major force shaping clonal hematopoiesis (Watson et al., 2020). CHIP thus represents counter escape in progress—cells that have partially disabled counting systems and gained proliferative advantage, but have not yet achieved full malignant transformation.

8.4 MMR Deficiency as Maximum Counter Acceleration

MMR deficiency provides a natural experiment in counter acceleration. When MMR is disabled, Counter 2 (homopolymer frameshifts) runs at maximum speed. The clinical

consequences demonstrate the relationship between counter speed and cancer timing: with normal MMR, cancer typically arises at age 60-80 in sporadic cases; with Lynch syndrome (heterozygous MMR deficiency), cancer arises at age 40-60; with CMMRD (homozygous deficiency), cancer develops by age 18. The indel mutational signatures associated with MMR deficiency—characterized by small insertions and deletions at homopolymer tracts (Alexandrov et al., 2020)—serve as molecular speedometers directly measuring Counter 2 activity, and indel signature burden correlates with MMR deficiency severity.

8.5 Selection, Counter Limits, and the ~5,000 Mutation Ceiling

Not all cancer risk differences reflect counter rate differences. Men have 4-fold higher bladder cancer risk than women despite equivalent mutation rates (Calvet et al., 2025). The difference lies in selection on counter outputs: men show stronger positive selection on truncating driver mutations, meaning the counter runs at the same speed but the consequences differ. This suggests that cancer risk equals the product of counter rate, selection pressure, and tissue context—the counters provide raw material, and selection determines whether that material drives cancer.

Is there a maximum counter value beyond which cells cannot survive? Preliminary evidence from single-cell sequencing of MMR-deficient cells suggests that while SNVs can accumulate to high levels (30,000+ per cell), indel accumulation may reach a functional ceiling. Zhang et al. (2024) observed approximately 5,000 mutations per cell in Msh2-deficient mouse fibroblasts, noting this "is likely the maximum tolerated mutation load of this cell type in vivo." Whether this represents a hard ceiling or varies by cell type remains to be determined.

> **** The concept of a functional indel ceiling is based on single-cell sequencing data from mouse models that has not been comprehensively validated across cell types and species. The specific threshold likely varies by tissue context, ploidy, and the genomic distribution of indels. Systematic characterization across human tissues is needed.

If confirmed, such a plateau would suggest a functional limit: cells may tolerate indels only up to a certain burden before genomic function becomes impossible, which could represent the maximum counter value—the point at which counting becomes irrelevant because survival is impossible regardless of cellular decisions.

9. TESTABLE PREDICTIONS

The indel counter framework generates specific, falsifiable predictions organized by mechanism and priority (Figure 3). We designate two predictions as *high-priority* based on their foundational importance and immediate testability with existing technologies.

Table 2:

Prediction	Method	Expected Result	Falsification
HIGH-PRIORITY Sister chromatid repeat asymmetry	Strand-seq + long-read sequencing	1-2 unit difference; consistent asymmetry absent	No consistent asymmetry

Prediction	Method	Expected Result	Falsification
		in MMR-deficient cells	
HIGH-PRIORITY ATR/CHK1 requirement for segregation during mitosis	ATR/CHK1 inhibition during mitosis	Abolishes nonrandom segregation	No effect on segregation bias
NPM1-PIDDosome threshold with defined repeat lengths	Isogenic cell lines with defined repeat lengths	Sharp increase in NPM1 phosphorylation at pathogenic threshold	Linear relationship without threshold
G4 formation threshold across repeat lengths	BG4 antibody/G4-seq	Sharp increase in G4 signal above pathogenic length	Gradual increase without threshold
ETAA1 vs TopBP1 specificity	Knockdown during mitosis	ETAA1 KD abolishes segregation; TopBP1 does not	Both or neither KD affect segregation
DDR protein co-segregation of longer-repeat chromatid	Immunofluorescence RPA, FANCD2, γH2AX anaphase figures	DDR markers co-segregate with longer-repeat chromatid	Random distribution

The two high-priority predictions were selected because: (1) sister chromatid repeat asymmetry is foundational—if repeats do not differ between sister chromatids, the entire segregation mechanism requires fundamental revision (Figure 3A); and (2) ATR/CHK1 inhibitors are commercially available and can be applied during a defined mitotic window, enabling immediate testing in any laboratory with live-cell imaging capability (Figure 3B).

Figure 3: Experimental Predictions with Expected Outcomes

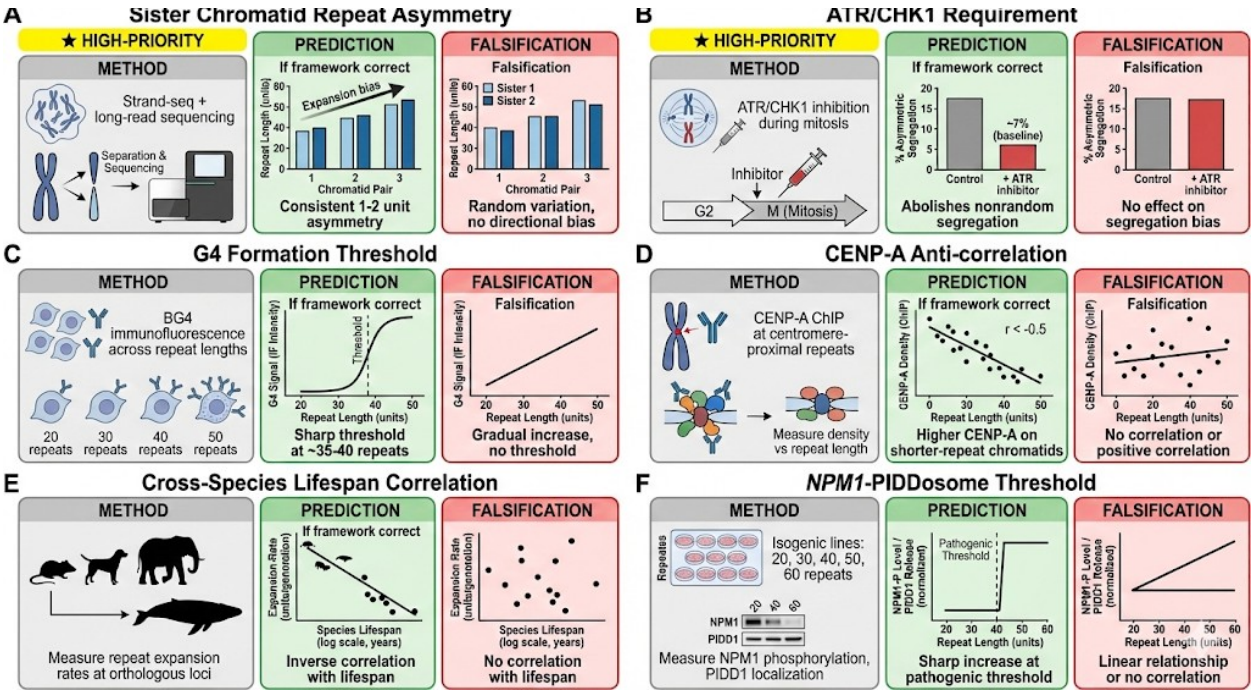

Figure 3. Key testable predictions derived from the indel counter framework. Each panel illustrates the experimental approach, the predicted outcome if the framework is correct, and the null outcome that would falsify the hypothesis. (A) Sister chromatid repeat asymmetry measured by long-read sequencing of metaphase chromosomes should reveal consistent 1-2 repeat unit differences favoring one sister; random variation would falsify this prediction. (B) ATR inhibition during mitosis should abolish nonrandom segregation, returning to baseline levels of approximately 7%; persistence of asymmetric segregation would falsify the ATR requirement. (C) G4 formation assessed by BG4 antibody staining should show sharp threshold behavior above 35-40 repeats rather than gradual linear increase. (D) CENP-A loading should anti-correlate with repeat length, with higher loading on shorter-repeat chromatids. (E) Cross-species analysis of repeat expansion rates at orthologous loci should reveal inverse correlation with species lifespan. (F) NPM1 phosphorylation and PIDD1 localization should increase sharply at pathogenic repeat lengths rather than showing linear or no correlation.

9.2 Threshold Enforcement and Chromatin Predictions

The PIDDosome threshold enforcement model generates several predictions. NPM1 phosphorylation at ATR/CHK1-dependent sites should increase with repeat length across a series of isogenic cell lines, with a sharp increase at the pathogenic threshold rather than a linear relationship (Figure 3F). PIDD1 knockout or caspase-2 knockout should rescue viability of cells with repeat lengths above the pathogenic threshold, at the cost of increased genomic instability and transformation potential. Cells with pathogenic-length repeats should show increased PIDD1 release from the nucleolus (measurable by immunofluorescence) and

increased caspase-2 activation (measurable by FLICA assays) compared to cells with sub-threshold repeats.

Chromatin-based predictions include that CENP-A density should anti-correlate with repeat length at centromere-proximal loci, with higher CENP-A loading on shorter-repeat chromatids (Figure 3D); this correlation should be disrupted by CENP-C knockdown. Forcing symmetric histone inheritance through MCM2 and DPB3 double mutation should disrupt asymmetric division; recent evidence is consistent with this prediction, demonstrating that loss of histone inheritance alters gene expression in embryonic stem cells and challenges differentiation programs (Kollenstart et al., 2025).

9.3 Disease Model and Evolutionary Predictions

Disease model predictions include that iPSCs derived from Huntington's disease, myotonic dystrophy, or Fragile X patients should show altered asymmetric division patterns compared to isogenic corrected controls, with increased symmetric divisions (both daughters receiving expanded repeats) in affected lines. MMR-deficient organoids should show altered stem cell pool dynamics due to disrupted asymmetric division, testable by measuring the ratio of symmetric to asymmetric divisions using time-lapse imaging of labeled stem cells.

Cross-species and cancer predictions relate counter rates to organismal outcomes. Repeat expansion rates at orthologous loci should inversely correlate with species lifespan when measured in cell lines from species with different lifespans (Figure 3E). Faster tissue-specific counters should predict earlier cancer onset after controlling for exposure, testable by correlating tissue-specific indel accumulation rates with cancer age-of-onset. Tumors arising in tissues with high repeat burden should show enrichment for mutations in ATR, CHK1, ETAA1, PIDD1, or CASP2 compared to tumors from tissues with lower repeat burden; this can be tested using TCGA data stratified by tissue repeat content.

Counter escape and telomere maintenance predictions test the relationship between MSI status and TMM selection. MSI tumors should show near-exclusive telomerase dependence rather than ALT, due to MRE11 frameshift blocking ALT access. TMM-negative tumors should show fewer total counter escapes (driver mutations) than TMM-positive tumors with equivalent pathological grade, reflecting incomplete escape. Engineered MSI cell lines with protected MRE11 (shortened poly(T) tract through synonymous mutations) should retain ALT capability and show reduced vulnerability to telomerase inhibition compared to natural MSI lines.

9.4 Experimental Technologies and Falsification Criteria

Testing these predictions requires several emerging technologies. Long-read sequencing platforms such as Nanopore and PacBio enable direct measurement of repeat lengths that are intractable for short-read methods (Giesselmann et al., 2019). Strand-seq enables assignment of DNA strands to specific sister chromatids (Falconer et al., 2012). Single-cell multiomics combines genomic measurement with chromatin state in individual cells. Organoid systems enable controlled perturbation of stem cell division in relevant tissue contexts.

The framework is designed to be falsifiable, with clear predictions that distinguish it from alternatives. If sister chromatids show no consistent repeat asymmetry, the segregation hypothesis requires modification. If ATR/CHK1 inhibition does not affect segregation bias, the

mechanism linking repeat length to segregation differs from our proposal. If NPM1 phosphorylation does not scale with repeat length, the PIDDosome threshold enforcement model requires revision. If PIDD1 knockout does not rescue supraphysiological expansion, alternative threshold mechanisms must be considered.

9.5 Metabolic Intervention Points

The three-hub metabolic model identifies several intervention points for modulating counter function:

Table 3:

Intervention	Target	Prediction
NAD+ precursors (NMN, NR)	NAD+ hub	Enhance asymmetric division in aged stem cells
Vitamin C	TET/KDM enzymes (Fe^{2+} maintenance)	Increase 5hmC and histone demethylation asymmetry
IDH inhibitors (ivosidenib, enasidenib)	α-KG/2-HG balance	Restore asymmetric division in IDH-mutant cells
PARP inhibitors (olaparib, talazoparib)	NAD+ consumption	Complex effects—preserve NAD+ but impair damage signaling
α-Ketoglutarate supplementation	TET/KDM activity	Enhance asymmetric division in metabolically compromised cells

These metabolic interventions target the interpretation machinery (metabolic hubs) rather than the counting mechanisms themselves (repeat expansion, frameshifts), potentially offering therapeutic windows where normal counting continues but pathological consequences are ameliorated.

10. DISCUSSION

10.1 Integration with Existing Aging Models and Epigenetic Clocks

The indel counter framework complements rather than replaces existing models. Telomeres count total replicative capacity while indels count division history with tissue-specific resolution, and both operate as counters with different readouts and timescales. Replicative senescence represents one outcome when counters cross thresholds, and the indel framework specifies what is counted while senescence is one possible response to threshold crossing. The somatic mutation theory of cancer is compatible with indel counters: cancer requires accumulated mutations, and the framework specifies which mutation types serve counting functions and how escape from counting enables malignancy.

The Horvath clock and related epigenetic clocks predict chronological and biological age from DNA methylation patterns at specific CpG sites (Horvath, 2013; Horvath & Raj, 2018), and the relationship between indel counters and these established aging biomarkers deserves particular attention. Several potential mechanistic connections exist: repeat expansion affects local methylation through altered chromatin structure, frameshift mutations in DNMTs or TET enzymes directly alter methylation patterns, and L1 insertions carry their own methylation-sensitive promoters. The two systems may provide complementary information, since epigenetic clocks measure continuous analog signals in the form of methylation levels while indel counters provide both analog readouts (repeat length) and digital readouts (frameshift); combining both might improve age prediction and identify discordant samples. A possible hierarchy may exist between these systems, with indel accumulation potentially upstream and causing methylation changes that epigenetic clocks detect, or alternatively, epigenetic clocks and indel counters may measure independent aspects of aging. Testing whether indel burden predicts epigenetic age beyond chronological age could distinguish these models.

10.2 Clinical and Therapeutic Implications

The counting framework reframes repeat expansion diseases as premature aging of specific systems. Huntington's disease represents accelerated counting in striatal neurons—not a fundamentally different process from normal aging, but the same process running faster due to inherited expansion. This reframing suggests therapeutic approaches: slowing the counter by reducing expansion rate, raising the threshold by stabilizing secondary structures, or modifying the response by altering fate determination upon threshold crossing. Current therapeutic efforts targeting repeat expansion fit this framework, including antisense oligonucleotides reducing repeat-containing transcripts (Tabrizi et al., 2019) and small molecules stabilizing non-pathological conformations.

The indel counter framework also suggests broader clinical applications. Measuring counter values—repeat lengths at expandable loci, frameshift burden at MSI-sensitive genes—could stratify cancer risk beyond family history and known mutations, and individuals with accelerated counters might benefit from intensified surveillance. If counters serve functional roles, manipulating them could modify disease through counter deceleration by stabilizing repeats, enhancing MMR, or reducing replication stress; threshold elevation by preventing secondary structure formation or blocking DDR activation; and counter interpretation modification by altering cellular responses to threshold crossing. MSI-H status already guides immunotherapy selection, and since the frameshift counter directly generates the neoantigens that drive immunotherapy response, counter-based metrics might refine patient selection beyond current MSI testing. If counters determine rather than merely track aging, interventions that slow counters might extend healthspan, and reverse transcriptase inhibitors (to reduce L1 activity), MMR enhancement, or repeat-stabilizing compounds could be evaluated for lifespan effects.

10.3 Limitations and Alternative Interpretations

Several limitations of the indel counter framework warrant acknowledgment, alongside a thorough treatment of alternative interpretations.

The most parsimonious alternative to our framework holds that all observations can be explained by biophysics and selection against damage *without* invoking counting functions. Under this view: (1) Repeat instability reflects unavoidable biophysical properties of repetitive DNA—slippage, secondary structure formation, and replication fork stalling are consequences of sequence chemistry, not evolved counting mechanisms. (2) Cellular responses to repeat expansion represent generic DNA damage responses that happen to create threshold-like behaviors, not calibrated readouts of division history. (3) Cross-species mutation rate scaling reflects metabolic rate differences, body temperature variation, or repair capacity that happens to correlate with lifespan, rather than selection on counter mechanisms. (4) The multi-hit requirement for cancer reflects the need to disable multiple tumor suppressors, not escape from multiple counters. This alternative is fully consistent with the molecular observations we cite; our framework adds the interpretive claim that these mechanisms serve counting *functions*, which the alternative does not require.

We cannot definitively distinguish these interpretations with current evidence. However, several observations favor the counting interpretation: the two-threshold structure of repeat diseases (permissive vs. pathogenic) corresponds to biologically meaningful division numbers rather than arbitrary damage levels; the ATR-TET3 phosphorylation axis provides causal integration between damage sensing and epigenetic modification that seems designed rather than accidental; and the remarkable cross-species calibration of mutation rates to lifespan suggests selection on counter mechanisms. Nevertheless, we acknowledge that these observations are also consistent with the alternative interpretation under appropriate auxiliary assumptions.

10.3.1 Distinguishing Counting from Metabolic Scaling

The inverse correlation between somatic mutation rate and species lifespan (Cagan et al., 2022) admits multiple interpretations. The metabolic scaling hypothesis proposes that longer-lived species simply have lower metabolic rates, producing less oxidative damage—no evolved counting required. Several observations favor active calibration over passive scaling.

First, the correlation is tighter for indels than for SNVs, suggesting selective pressure specifically on length-changing mutations rather than general damage. If metabolic scaling alone explained the pattern, SNVs and indels should show equivalent scaling.

Second, tissue-specific mutation rates vary independently of metabolic rate. Colon epithelium and hematopoietic stem cells have similar metabolic demands but different mutation accumulation patterns, suggesting tissue-autonomous counting mechanisms.

Third, MMR expression levels vary across species in ways that correlate with lifespan independently of metabolic rate (Milholland et al., 2017). This active regulation of repair capacity suggests calibration rather than passive constraint.

Fourth, the lesion segregation phenomenon documented by Aitken et al. (2020) and elaborated by Panagopoulos et al. (2025) demonstrates that cells do not simply accumulate damage proportionally to metabolic output. Instead, lesion persistence and segregation dynamics actively shape mutation patterns, with differential cell fates depending on which daughter inherits damaged templates. This active sorting mechanism is inconsistent with purely passive metabolic scaling.

Fifth, Aitken et al. demonstrated that 67% of chemically-induced liver tumors in C3H mice developed from first-generation daughter cells after mutagenesis, indicating that threshold enforcement is most critical when damage burden is acutely elevated. This suggests cells have evolved mechanisms to actively manage damage inheritance, not merely tolerate passively accumulated damage.

Definitive distinction requires measuring repeat expansion rates in metabolically matched cell lines across species—for example, comparing fibroblasts cultured under identical conditions from mouse, dog, and human. If expansion rates still scale with species lifespan despite equivalent metabolic conditions, this would support evolved calibration over passive scaling.

Much of the evidence presented here is correlational—indels accumulate with age and correlate with cancer risk—but direct causal testing of counting functions remains limited. The framework could be partially or entirely epiphenomenal, with indels accumulating as damage without serving counting functions. While we propose that repeat length and frameshift accumulation cross thresholds that trigger fate decisions, the precise threshold values and their cellular readouts remain incompletely characterized, and different tissues may have different thresholds, complicating experimental validation.

The repeat-guided segregation hypothesis, while generating testable predictions, rests on several extrapolations that require validation: (1) Xing et al. (2020) demonstrated ATR/CHK1-dependent nonrandom segregation, but using aphidicolin and oncogene-induced stress, not repeat expansion specifically—whether repeat-induced G4 structures trigger the same pathway is untested. (2) The PIDDosome literature primarily concerns supernumerary centrosomes, not repeat-induced replication stress—the proposed NPM1-mediated threshold enforcement for repeat burden is hypothetical. (3) No published study has directly measured repeat length differences between sister chromatids in mammalian cells—this foundational claim awaits experimental validation.

We present four counters as parallel systems, but they may be more interconnected than this framing suggests. MMR deficiency accelerates multiple counters; transcription and replication share common stresses through R-loop formation; L1 activity is regulated by methylation patterns affected by DNA damage signaling. If counters are highly correlated, the framework's predictive power is reduced.

Most detailed evidence comes from model organisms (yeast, *Drosophila*, mice) or human cell culture, and whether the same counters operate identically across mammalian species remains to be established.

The testable predictions in Section 9 are designed specifically to discriminate between the counting interpretation and the damage interpretation. If repeat length differences between sister chromatids cannot be detected, or if ATR inhibition does not abolish repeat-guided segregation specifically, or if PIDDosome activation shows no dependence on repeat length, these failures would argue against our framework. We present this framework not as established fact but as a hypothesis-generating synthesis that organizes existing observations and generates new experimental directions.

A practical challenge for validating this framework is the extensive somatic mosaicism in counter values. Unlike germline mutations that affect every cell identically, counter values vary

cell-to-cell within a tissue, reflecting different division histories, stochastic frameshift events, and L1 insertion patterns. Bulk sequencing approaches mask this cell-to-cell variation, reporting only average counter values that may not reflect any individual cell's state. Single-cell and single-molecule approaches are essential for measuring the actual distribution of counter values within tissues, but current technologies have limited throughput and accuracy for detecting small indels. This technical limitation means that clinical monitoring of counter values—if validated as predictive of cancer risk—would require substantial methodological advances in single-cell long-read sequencing with sufficient coverage to detect repeat length and frameshift status at multiple loci simultaneously.

10.4 Germline Counter Reset and Intergenerational Transmission

A fundamental question for any counting system is how counters are reset between generations. If counters only accumulated, offspring would inherit aged genomes and lifespan would decrease across generations. Several mechanisms ensure partial or complete counter reset in the germline.

The first mechanism involves DNMT1 cytoplasmic exclusion. The oocyte isoform of DNMT1 (DNMT1o) is stored in the cytoplasm during early cleavage divisions rather than being localized to the nucleus (Ratnam et al., 2002). Nuclear trafficking of DNMT1 occurs only at the 8-cell stage in mouse. This cytoplasmic exclusion prevents maintenance of parental methylation patterns during early cleavages, enabling passive demethylation that partially resets the epigenetic counter. UHRF1, which is required for recruiting DNMT1 to replication forks (Sharif et al., 2007), shows similarly regulated expression.

The second mechanism involves a DNMT3B-to-DNMT3A developmental transition. De novo methyltransferase expression follows a stereotyped developmental pattern. DNMT3B is expressed in progenitor cells (neural progenitors, spermatogonia, ES cells), while DNMT3A is expressed in postmitotic and differentiating cells (Watanabe et al., 2002). This transition is conserved across hematopoiesis, spermatogenesis, and neurogenesis, suggesting DNMT3B establishes initial methylation patterns while DNMT3A refines tissue-specific marks. Notably, DNMT3L—an important cofactor in mouse—is not detected in human oocytes (Petrussa et al., 2014), indicating species-specific differences in germline reprogramming.

The third mechanism involves protection of imprinted regions. Not all methylation is reset. STELLA (DPPA3/PGC7) binds H3K9me2 to protect the maternal genome from TET3-mediated demethylation. ZFP57 and TRIM28 protect imprinting control regions through both waves of reprogramming (primordial germ cell and post-fertilization). These protection mechanisms create selectivity, allowing global counter reset while preserving parent-of-origin information at imprinted loci (Li et al., 2008; Wossidlo et al., 2011).

The fourth mechanism involves meiotic repeat dynamics. Repeat expansion in the germline occurs primarily in post-meiotic haploid cells through gap repair mechanisms (Kovtun & McMurray, 2001). Because these cells are haploid, expansion cannot involve recombination with a sister chromatid or homolog—it requires gap-filling synthesis that is inherently mutagenic. This explains why repeat instability shows parent-of-origin effects (paternal expansion in Huntington's disease, maternal expansion in Fragile X): the haploid phase differs in duration and transcriptional activity between spermatogenesis and oogenesis. The MSH2-

containing complex is required for expansion during gap-filling synthesis (Rajagopal et al., 2023), consistent with MMR's paradoxical role in promoting repeat expansion.

The incomplete reset of repeat counters explains anticipation—the phenomenon where repeat expansion diseases show earlier onset and increased severity in successive generations. Each generation inherits a counter that has already advanced partway, and somatic expansion during the individual's lifetime pushes it further toward the pathogenic threshold. The counter is not fully reset but partially reset, with the degree of reset depending on the specific repeat sequence and germline-specific expansion/contraction dynamics.

10.5 Future Directions and Evolutionary Perspective

The four indel counting systems described here generate signals that must be interpreted by cellular machinery to determine fate. A cell that crosses a structural threshold at a repeat locus, or accumulates sufficient frameshifts to lose a tumor suppressor, faces a decision: continue proliferating, differentiate, senesce, or die. The molecular architecture of this decision-making process—how counter signals are integrated with metabolic state, buffered across cell cycles, and ultimately resolved into irreversible fate commitment—represents a distinct layer of cellular logic. The counting mechanisms described here serve as inputs to this decision system, the architecture of which we address in subsequent work.

Beyond the four counters detailed here, other molecular systems may contribute to cellular counting. One speculative candidate involves wobble position mutations in tRNA genes. The wobble position (first anticodon position) determines codon recognition specificity through non-standard base pairing (Crick, 1966). Unlike Counters 1-4, such a system would be time-dependent rather than division-dependent, cumulative rather than threshold-based, and global rather than gene-specific—a wobble mutation would affect every mRNA using that codon. Post-mitotic cells (neurons, cardiomyocytes) would be particularly vulnerable because they cannot dilute damaged tRNAs through division. This hypothesis is motivated by established observations about translation fidelity decline with age (Llewellyn et al., 2024), but experimental demonstration of age-related wobble mutation accumulation and its functional consequences remains to be established. We do not include this among the core counters pending such validation, but note it as a direction warranting investigation.

10.5.1 Tissue Specificity Within Repeat Diseases

A related puzzle concerns tissue specificity within diseases. Huntington's disease preferentially affects medium spiny neurons of the striatum despite ubiquitous HTT expression and somatic repeat expansion occurring in multiple tissues. The counting framework explains *that* cells track repeat burden but not *why* striatal neurons are selectively vulnerable.

Potential explanations within the framework include: (1) **Differential threshold stringency**: Striatal neurons may have lower PIDDosome activation thresholds or more sensitive G4-detection mechanisms, triggering cell death at repeat lengths tolerated elsewhere. (2) **Metabolic hub concentrations**: The striatum has distinctive metabolic characteristics (high energy demands, dopaminergic innervation) that may modulate counter sensitivity through altered α-KG, NAD+, or succinyl-CoA levels. (3) **Post-mitotic vulnerability**: Unlike dividing cells that can segregate damaged DNA to sacrificial daughters, post-mitotic neurons must retain all accumulated damage. The striatum's specific neuronal populations may have higher

baseline repeat lengths or faster somatic expansion rates. (4) **Downstream effector expression**: Even with equivalent threshold crossing, different tissues may respond differently based on caspase-2 expression levels, p53 pathway activity, or alternative fate-determination circuits.

Resolving striatal specificity remains an important challenge. The framework provides a molecular basis for threshold-dependent toxicity but requires integration with tissue-specific cell biology to explain selective vulnerability.

Why would evolution create multiple parallel counters rather than a single unified system? This reflects a general principle in biological systems (Kirschner & Gerhart, 1998). Several advantages emerge. Multiple counters provide robustness: disabling one counter through mutation or environmental manipulation leaves others intact, and cancer must escape all counters, not just one. Different counters dominate in different tissues, enabling tissue-appropriate aging trajectories; rapidly dividing epithelia use division-coupled counters while post-mitotic neurons use time-dependent counters. Multiple counters enable fine-tuning of lifespan through evolutionary flexibility, as species can adjust individual counter rates rather than redesigning an entire counting system. The multi-counter requirement for cancer represents an evolutionary adaptation that suppresses malignancy, since each additional counter that must be escaped reduces cancer probability, providing selective advantage to organisms with diverse counting mechanisms.

If repeat expansion serves counting functions, why hasn't selection eliminated expandable repeats from essential genes? The persistence of CAG repeats in huntingtin—which causes devastating Huntington's disease when expanded—seems maladaptive if these repeats evolved for counting. Several considerations address this apparent paradox. First, most repeat expansion diseases manifest post-reproductively: Huntington's disease typically presents in the 30s-50s, after most reproduction has occurred, limiting selection pressure against disease-causing alleles. However, this generalization has important exceptions: Friedreich's ataxia typically presents in childhood or adolescence, juvenile-onset Huntington's disease (caused by very large expansions) affects children, and myotonic dystrophy type 1 can cause congenital disease with severe neonatal manifestations. These childhood-onset cases would experience strong negative selection, making the persistence of disease-causing repeats more puzzling. The resolution may involve balancing selection (heterozygote advantage), de novo expansion rates that replenish disease alleles faster than selection removes them, or the relative rarity of very large expansions compared to intermediate-length alleles. Second, normal-length repeats may provide net fitness benefits that outweigh the costs of rare expansions—huntingtin's polyglutamine tract, for example, may serve protein function or regulatory roles at normal lengths (The Huntington's Disease Collaborative Research Group, 1993). Third, the counting framework does not require that repeats *evolved* for counting—consistent with Interpretation B (Section 1.4), repeats may have arisen for other reasons (protein function, gene regulation, genomic architecture) and cells subsequently co-opted their expansion dynamics for informational purposes. The disease phenotypes would then represent the cost of exploiting an imperfect counting mechanism rather than evidence against counting function. Finally, genetic modifiers that delay disease onset in Huntington's disease families (Genetic Modifiers of Huntington's Disease Consortium, 2019) suggest ongoing selection to mitigate expansion-related pathology while presumably preserving whatever functions the repeats serve.

10.6 Conclusion

We have proposed that insertions and deletions—traditionally viewed as pathological errors—constitute a sophisticated system for counting cellular divisions and coordinating developmental timing. Four parallel counters operate through distinct molecular mechanisms: repeat expansion creating structural thresholds, homopolymer frameshifts providing digital gene inactivation, retrotransposon insertions generating somatic diversity, and transcription-associated errors diversifying proteomes.

The repeat-guided segregation hypothesis provides a mechanism linking counter values to asymmetric stem cell division, potentially explaining how stem cells maintain "younger" genomes while directing "older" DNA to differentiating progeny. We have identified a specific molecular pathway: G-quadruplex and i-motif structures formed by expanded repeats create strand-specific signals; ATR/CHK1, operating through a mitosis-specific ETAA1-dependent activation pathway, reads these structural signals and directs nonrandom segregation of damaged chromatids; and the PIDDosome (NPM1-PIDD1-caspase-2 axis) provides threshold enforcement when repeat burden exceeds the capacity for asymmetric segregation. This "Segregate and Sacrifice" strategy protects stem cell lineages from accumulated repeat damage by partitioning that damage to daughter cells destined for elimination or differentiation.

Cancer emerges, in part, as counter escape—the systematic disabling of counting mechanisms that normally constrain proliferation. The tissue-specificity of cancer reflects tissue-specific counter dominance. The approximately 100-fold acceleration of cancer in constitutional MMR deficiency reflects counter speed determining cancer timing. We now add that escape from the ATR/CHK1-PIDDosome axis may represent a critical but underappreciated step in carcinogenesis for tissues with high repeat burden.

Cross-species conservation of end-of-life mutation burden suggests evolution has tuned counter rates to match species lifespan, implying that the counters do not merely track aging but determine it.

The framework generates testable predictions spanning molecular biology (sister chromatid asymmetry, G4 formation thresholds), cell biology (CENP-A loading, ATR/CHK1 segregation effects), stem cell biology (division symmetry, NPM1 phosphorylation), and evolutionary biology (cross-species rates). Experimental tests will refine or refute specific mechanistic proposals while evaluating the broader claim that indels serve counting rather than merely damaging functions.

If correct, this framework transforms our understanding of genome dynamics—from viewing indels as noise to be suppressed, to recognizing them as signals encoding cellular history and guiding developmental fate.

ACKNOWLEDGMENTS

The author also acknowledges the broader scientific community whose published work made this synthesis possible, particularly researchers in repeat expansion biology, DNA damage response, asymmetric cell division, and cancer genomics whose empirical findings provided the foundation for this theoretical framework.

Figures were generated using Google's Nano Banana Pro. The author thanks the developers of PubMed and related databases for enabling comprehensive literature verification.

No external funding supported this work.

AUTHOR CONTRIBUTIONS

Dustin Lane: Conceptualization (lead); hypothesis generation (lead); framework architecture (lead); critical analysis and interpretation (lead); review and editing (lead); final approval and accountability for all content (sole).

COMPETING INTERESTS

The author declares no competing interests.

REFERENCES

A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. (1993). *Cell*, *72*(6), 971–983. https://doi.org/10.1016/0092-8674(93)90585-e

Al-Khalaf, M. H., Blake, L. E., Larsen, B. D., Bell, R. A., Brunette, S., Parks, R. J., Rudnicki, M. A., McKinnon, P. J., Jeffrey Dilworth, F., & Megeney, L. A. (2016). Temporal activation of XRCC1-mediated DNA repair is essential for muscle differentiation. *Cell discovery*, *2*, 15041. https://doi.org/10.1038/celldisc.2015.41

Alexandrov, L. B., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., Wu, Y., Boot, A., Covington, K. R., Gordenin, D. A., Bergstrom, E. N., Islam, S. M. A., Lopez-Bigas, N., Klimczak, L. J., McPherson, J. R., Morganella, S., Sabarinathan, R., Wheeler, D. A., Mustonen, V., PCAWG Mutational Signatures Working Group, Getz, G., … PCAWG Consortium (2020). The repertoire of mutational signatures in human cancer. *Nature*, *578*(7793), 94–101. https://doi.org/10.1038/s41586-020-1943-3

Aitken, S. J., Anderson, C. J., Connor, F., Pich, O., Sundaram, V., Feig, C., Rayner, T. F., Lukk, M., Aitken, S., Luft, J., Kentepozidou, E., Arnedo-Pac, C., Beentjes, S. V., Davies, S. E., Drews, R. M., Ewing, A., Kaiser, V. B., Khamseh, A., López-Arribillaga, E., Redmond, A. M., … Taylor, M. S. (2020). Pervasive lesion segregation shapes cancer genome evolution. *Nature*, *583*(7815), 265–270. https://doi.org/10.1038/s41586-020-2435-1

Baptiste, B. A., Ananda, G., Strubczewski, N., Lutzkanin, A., Khoo, S. J., Srikanth, A., Kim, N., Makova, K. D., Krasilnikova, M. M., & Eckert, K. A. (2013). Mature microsatellites: mechanisms underlying dinucleotide microsatellite mutational biases in human cells. *G3 (Bethesda, Md.)*, *3*(3), 451–463. https://doi.org/10.1534/g3.112.005173

Bass, T. E., Luzwick, J. W., Kavanaugh, G., Carroll, C., Dungrawala, H., Glick, G. G., Feldkamp, M. D., Putney, R., Chazin, W. J., & Cortez, D. (2016). ETAA1 acts at stalled replication forks to maintain genome integrity. *Nature cell biology*, *18*(11), 1185–1195. https://doi.org/10.1038/ncb3415

Betto, R. M., Diamante, L., Perrera, V., Audano, M., Rapelli, S., Lauria, A., Incarnato, D., Arboit, M., Pedretti, S., Rigoni, G., Guerineau, V., Touboul, D., Stirparo, G. G., Lohoff, T., Boroviak, T., Grumati, P., Soriano, M. E., Nichols, J., Mitro, N., Oliviero, S., … Martello, G. (2021). Metabolic control of DNA methylation in naive pluripotent cells. *Nature genetics*, *53*(2), 215–229. https://doi.org/10.1038/s41588-020-00770-2

Biffi, G., Tannahill, D., McCafferty, J., & Balasubramanian, S. (2013). Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature chemistry*, *5*(3), 182–186. https://doi.org/10.1038/nchem.1548

Blaschke, K., Ebata, K. T., Karimi, M. M., Zepeda-Martínez, J. A., Goyal, P., Mahapatra, S., Tam, A., Laird, D. J., Hirst, M., Rao, A., Lorincz, M. C., & Ramalho-Santos, M. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature*, *500*(7461), 222–226. https://doi.org/10.1038/nature12362

Blokzijl, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kuijk, E., Prins, P., Nijman, I. J., Martincorena, I., Mokry, M., Wiegerinck, C. L., Middendorp, S., Sato, T., Schwank, G., Nieuwenhuis, E. E., Verstegen, M. M., van der Laan, L. J., … van Boxtel, R. (2016). Tissue-specific mutation accumulation in human adult stem cells during life. *Nature*, *538*(7624), 260–264. https://doi.org/10.1038/nature19768

Bodea, G. O., Botto, J. M., Ferreiro, M. E., Sanchez-Luque, F. J., de Los Rios Barreda, J., Rasmussen, J., Rahman, M. A., Fenlon, L. R., Jansz, N., Gubert, C., Gerdes, P., Bodea, L. G., Ajjikuttira, P., Da Costa Guevara, D. J., Cumner, L., Bell, C. C., Kozulin, P., Billon, V., Morell, S., Kempen, M. H. C., … Faulkner, G. J. (2024). LINE-1 retrotransposons contribute to mouse PV interneuron development. *Nature neuroscience*, *27*(7), 1274–1284. https://doi.org/10.1038/s41593-024-01650-2

Bodea, G. O., McKelvey, E. G. Z., & Faulkner, G. J. (2018). Retrotransposon-induced mosaicism in the neural genome. *Open biology*, *8*(7), 180074. https://doi.org/10.1098/rsob.180074

Boyer, J. C., Yamada, N. A., Roques, C. N., Hatch, S. B., Riess, K., & Farber, R. A. (2002). Sequence dependent instability of mononucleotide microsatellites in cultured mismatch repair proficient and deficient mammalian cells. *Human molecular genetics*, *11*(6), 707–713. https://doi.org/10.1093/hmg/11.6.707

Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., & Rolf, B. (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American journal of human genetics*, *62*(6), 1408–1415. https://doi.org/10.1086/301869

Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., & Hudson, T. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, *68*(4), 799–808. https://doi.org/10.1016/0092-8674(92)90154-5

Budczies, J., Bockmayr, M., Klauschen, F., Endris, V., Fröhling, S., Schirmacher, P., Denkert, C., & Stenzinger, A. (2017). Mutation patterns in genes encoding interferon signaling and antigen presentation: A pan-cancer survey with implications for the use of immune checkpoint inhibitors. *Genes, chromosomes & cancer*, *56*(8), 651–659. https://doi.org/10.1002/gcc.22468

Burigotto, M., Mattivi, A., Migliorati, D., Magnani, G., Valentini, C., Roccuzzo, M., Offterdinger, M., Pizzato, M., Schmidt, A., Villunger, A., Maffini, S., & Fava, L. L. (2021). Centriolar distal appendages activate the centrosome-PIDDosome-p53 signalling axis via ANKRD26. *The EMBO journal*, *40*(4), e104844. https://doi.org/10.15252/embj.2020104844

Butuči, M., Williams, A. B., Wong, M. M., Kramer, B., & Michael, W. M. (2015). Zygotic Genome Activation Triggers Chromosome Damage and Checkpoint Signaling in C. elegans Primordial Germ Cells. *Developmental cell*, *34*(1), 85–95. https://doi.org/10.1016/j.devcel.2015.04.019

Cagan, A., Baez-Ortega, A., Brzozowska, N., Abascal, F., Coorens, T. H. H., Sanders, M. A., Lawson, A. R. J., Harvey, L. M. R., Bhosle, S., Jones, D., Alcantara, R. E., Butler, T. M., Hooks, Y., Roberts, K., Anderson, E., Lunn, S., Flach, E., Spiro, S., Januszczak, I., Wrigglesworth, E., … Martincorena, I. (2022). Somatic mutation rates scale with lifespan across mammals. *Nature*, *604*(7906), 517–524. https://doi.org/10.1038/s41586-022-04618-z

Calabrese, P., & Durrett, R. (2003). Dinucleotide repeats in the Drosophila and human genomes have complex, length-dependent mutation processes. *Molecular biology and evolution*, *20*(5), 715–725. https://doi.org/10.1093/molbev/msg084

Calvet, F., Blanco Martinez-Illescas, R., Muiños, F., Tretiakova, M., Latorre-Esteves, E. S., Fredrickson, J., Andrianova, M., Pellegrini, S., Huber, A. R., Ramis-Zaldivar, J. E., An, S. C., Thieme, E., Kohrn, B. F., Grau, M. L., Gonzalez-Perez, A., Lopez-Bigas, N., & Risques, R. A. (2025). Sex and smoking bias in the selection of somatic mutations in human bladder. *Nature*, *647*(8089), 436–444. https://doi.org/10.1038/s41586-025-09521-x

Cairns J. (1975). Mutation selection and the natural history of cancer. *Nature*, *255*(5505), 197–200. https://doi.org/10.1038/255197a0

Cairns J. (2006). Cancer and the immortal strand hypothesis. *Genetics*, *174*(3), 1069–1072. https://doi.org/10.1534/genetics.104.66886

Campuzano, V., Montermini, L., Moltò, M. D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Cañizares, J., Koutnikova, H., Bidichandani, S. I., Gellera, C., Brice, A., Trouillas, P., De Michele, G., Filla, A., De Frutos, R., … Pandolfo, M. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science (New York, N.Y.)*, *271*(5254), 1423–1427. https://doi.org/10.1126/science.271.5254.1423

Carty, B. L., Dattoli, A. A., & Dunleavy, E. M. (2021). CENP-C functions in centromere assembly, the maintenance of CENP-A asymmetry and epigenetic age in Drosophila germline stem cells. *PLoS genetics*, *17*(5), e1009247. https://doi.org/10.1371/journal.pgen.1009247

Charlton, S. J., Flury, V., Kanoh, Y., Genzor, A. V., Kollenstart, L., Ao, W., Brøgger, P., Weisser, M. B., Adamus, M., Alcaraz, N., Delvaux de Fenffe, C. M., Mattiroli, F., Montoya, G., Masai, H., Groth, A., & Thon, G. (2024). The fork protection complex promotes parental histone recycling and epigenetic memory. *Cell*, *187*(18), 5029–5047.e21. https://doi.org/10.1016/j.cell.2024.07.017

Cimprich, K. A., & Cortez, D. (2008). ATR: an essential regulator of genome integrity. *Nature reviews. Molecular cell biology*, *9*(8), 616–627. https://doi.org/10.1038/nrm2450

Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., Morell, M., O'Shea, K. S., Moran, J. V., & Gage, F. H. (2009). L1 retrotransposition in human neural progenitor cells. *Nature*, *460*(7259), 1127–1131. https://doi.org/10.1038/nature08248

Crick F. H. (1966). Codon--anticodon pairing: the wobble hypothesis. *Journal of molecular biology*, *19*(2), 548–555. https://doi.org/10.1016/s0022-2836(66)80022-0

Dattoli, A. A., Carty, B. L., Kochendoerfer, A. M., Morgan, C., Walshe, A. E., & Dunleavy, E. M. (2020). Asymmetric assembly of centromeres epigenetically regulates stem cell fate. *The Journal of cell biology*, *219*(4), e201910084. https://doi.org/10.1083/jcb.201910084

de Cecco, M., Ito, T., Petrashen, A. P., Elias, A. E., Skvir, N. J., Criscione, S. W., Caligiana, A., Brocculi, G., Adney, E. M., Boeke, J. D., Le, O., Beauséjour, C., Ambati, J., Ambati, K., Simon, M., Seluanov, A., Gorbunova, V., Slagboom, P. E., Helfand, S. L., Neretti, N., … Sedivy, J. M. (2019). L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature*, *566*(7742), 73–78. https://doi.org/10.1038/s41586-018-0784-9

Dieringer, D., & Schlötterer, C. (2003). Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome research*, *13*(10), 2242–2251. https://doi.org/10.1101/gr.1416703

DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G. Y., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. H., Wszolek, Z. K., … Rademakers, R. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, *72*(2), 245–256. https://doi.org/10.1016/j.neuron.2011.09.011

Donaldson, J., Powell, S., Rickards, N., Holmans, P., & Jones, L. (2021). What is the Pathogenic CAG Expansion Length in Huntington's Disease?. *Journal of Huntington's disease*, *10*(1), 175–202. https://doi.org/10.3233/JHD-200445

Du, J., Campau, E., Soragni, E., Jespersen, C., & Gottesfeld, J. M. (2013). Length-dependent CTG·CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells. *Human molecular genetics*, *22*(25), 5276–5287. https://doi.org/10.1093/hmg/ddt386

Ellegren H. (2004). Microsatellites: simple sequences with complex evolution. *Nature reviews. Genetics*, *5*(6), 435–445. https://doi.org/10.1038/nrg1348

Ercan, A. B., Aronson, M., Fernandez, N. R., Chang, Y., Levine, A., Liu, Z. A., Negm, L., Edwards, M., Bianchi, V., Stengs, L., Chung, J., Al-Battashi, A., Reschke, A., Lion, A., Ahmad, A., Lassaletta, A., Reddy, A. T., Al-Darraji, A. F., Shah, A. C., Van Damme, A., … Tabori, U. (2024). Clinical and biological landscape of constitutional mismatch-repair deficiency syndrome: an International Replication Repair Deficiency Consortium cohort study. *The Lancet. Oncology*, *25*(5), 668–682. https://doi.org/10.1016/S1470-2045(24)00026-3

Evrony, G. D., Cai, X., Lee, E., Hills, L. B., Elhosary, P. C., Lehmann, H. S., Parker, J. J., Atabay, K. D., Gilmore, E. C., Poduri, A., Park, P. J., & Walsh, C. A. (2012). Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell*, *151*(3), 483–496. https://doi.org/10.1016/j.cell.2012.09.035

Fabre, M. A., de Almeida, J. G., Fiorillo, E., Mitchell, E., Damaskou, A., Rak, J., Orrù, V., Marongiu, M., Chapman, M. S., Vijayabaskar, M. S., Baxter, J., Hardy, C., Abascal, F., Williams, N., Nangalia, J., Martincorena, I., Campbell, P. J., McKinney, E. F., Cucca, F., Gerstung, M., … Vassiliou, G. S. (2022). The longitudinal dynamics and natural history of clonal haematopoiesis. *Nature*, *606*(7913), 335–342. https://doi.org/10.1038/s41586-022-04785-z

Falconer, E., Hills, M., Naumann, U., Poon, S. S., Chavez, E. A., Sanders, A. D., Zhao, Y., Hirst, M., & Lansdorp, P. M. (2012). DNA template strand sequencing of single-cells maps genomic rearrangements at high resolution. *Nature methods*, *9*(11), 1107–1112. https://doi.org/10.1038/nmeth.2206

Faulkner, G. J., & Garcia-Perez, J. L. (2017). L1 Mosaicism in Mammals: Extent, Effects, and Evolution. *Trends in genetics : TIG*, *33*(11), 802–816. https://doi.org/10.1016/j.tig.2017.07.004

Fava, L. L., Schuler, F., Sladky, V., Haschka, M. D., Soratroi, C., Eiterer, L., Demetz, E., Weiss, G., Geley, S., Nigg, E. A., & Villunger, A. (2017). The PIDDosome activates p53 in response to supernumerary centrosomes. *Genes & development*, *31*(1), 34–45. https://doi.org/10.1101/gad.289728.116

Fearon, E. R., & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, *61*(5), 759–767. https://doi.org/10.1016/0092-8674(90)90186-i

Ferrand, J., Dabin, J., Chevallier, O., Kane-Charvin, M., Kupai, A., Hrit, J., Rothbart, S. B., & Polo, S. E. (2025). Mitotic chromatin marking governs the segregation of DNA damage. *Nature communications*, *16*(1), 746. https://doi.org/10.1038/s41467-025-56090-8

Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M., & Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell*, *75*(5), 1027–1038. https://doi.org/10.1016/0092-8674(93)90546-3

Franco, I., Helgadottir, H. T., Moggio, A., Larsson, M., Vrtačnik, P., Johansson, A., Norgren, N., Lundin, P., Mas-Ponte, D., Nordström, J., Lundgren, T., Stenvinkel, P., Wennberg, L., Supek, F., & Eriksson, M. (2019). Whole genome DNA sequencing provides an atlas of somatic mutagenesis in healthy human cells and identifies a tumor-prone cell type. *Genome biology*, *20*(1), 285. https://doi.org/10.1186/s13059-019-1892-z

Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. Electronic address: gusella@helix.mgh.harvard.edu, & Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium (2019). CAG Repeat Not Polyglutamine Length Determines Timing of Huntington's Disease Onset. *Cell*, *178*(4), 887–900.e14. https://doi.org/10.1016/j.cell.2019.06.036

Giannini, G., Ristori, E., Cerignoli, F., Rinaldi, C., Zani, M., Viel, A., Ottini, L., Crescenzi, M., Martinotti, S., Bignami, M., Frati, L., Screpanti, I., & Gulino, A. (2002). Human MRE11 is inactivated in mismatch repair-deficient cancers. *EMBO reports*, *3*(3), 248–254. https://doi.org/10.1093/embo-reports/kvf044

Giesselmann, P., Brändl, B., Raimondeau, E., Bowen, R., Rohrandt, C., Tandon, R., Kretzmer, H., Assum, G., Galonska, C., Siebert, R., Ammerpohl, O., Heron, A., Schneider, S. A., Ladewig, J., Koch, P., Schuldt, B. M., Graham, J. E., Meissner, A., & Müller, F. J. (2019). Analysis of short tandem

repeat expansions and their methylation state with nanopore sequencing. *Nature biotechnology*, *37*(12), 1478–1481. https://doi.org/10.1038/s41587-019-0293-x

Gold, M. A., Whalen, J. M., Freon, K., Hong, Z., Iraqui, I., Lambert, S. A. E., & Freudenreich, C. H. (2021). Restarted replication forks are error-prone and cause CAG repeat expansions and contractions. *PLoS genetics*, *17*(10), e1009863. https://doi.org/10.1371/journal.pgen.1009863

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., Xie, Z. G., Shi, L., He, X., Jin, S. G., Iqbal, K., Shi, Y. G., Deng, Z., Szabó, P. E., Pfeifer, G. P., Li, J., & Xu, G. L. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, *477*(7366), 606–610. https://doi.org/10.1038/nature10443

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Harley, C. B., Futcher, A. B., & Greider, C. W. (1990). Telomeres shorten during ageing of human fibroblasts. *Nature*, *345*(6274), 458–460. https://doi.org/10.1038/345458a0

HAYFLICK, L., & MOORHEAD, P. S. (1961). The serial cultivation of human diploid cell strains. *Experimental cell research*, *25*, 585–621. https://doi.org/10.1016/0014-4827(61)90192-6

Heaphy, C. M., Subhawong, A. P., Hong, S. M., Goggins, M. G., Montgomery, E. A., Gabrielson, E., Netto, G. J., Epstein, J. I., Lotan, T. L., Westra, W. H., Shih, I.eM., Iacobuzio-Donahue, C. A., Maitra, A., Li, Q. K., Eberhart, C. G., Taube, J. M., Rakheja, D., Kurman, R. J., Wu, T. C., Roden, R. B., … Meeker, A. K. (2011). Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. *The American journal of pathology*, *179*(4), 1608–1615. https://doi.org/10.1016/j.ajpath.2011.06.018

Hiregange, D., Naick, H., & Rao, B. J. (2020). ATR signalling mediates the prosurvival function of phospho-NPM against PIDDosome mediated cell death. *Cellular signalling*, *71*, 109602. https://doi.org/10.1016/j.cellsig.2020.109602

Horvath S. (2013). DNA methylation age of human tissues and cell types. *Genome biology*, *14*(10), R115. https://doi.org/10.1186/gb-2013-14-10-r115

Horvath, S., & Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature reviews. Genetics*, *19*(6), 371–384. https://doi.org/10.1038/s41576-018-0004-3

Horvath, S., Oshima, J., Martin, G. M., Lu, A. T., Quach, A., Cohen, H., Felton, S., Matsuyama, M., Lowe, D., Kabacik, S., Wilson, J. G., Reiner, A. P., Maierhofer, A., Flunkert, J., Aviv, A., Hou, L., Baccarelli, A. A., Li, Y., Stewart, J. D., Whitsel, E. A., … Raj, K. (2018). Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and *ex vivo* studies. *Aging*, *10*(7), 1758–1775. https://doi.org/10.18632/aging.101508

Hubaud, A., & Pourquié, O. (2014). Signalling dynamics in vertebrate segmentation. *Nature reviews. Molecular cell biology*, *15*(11), 709–721. https://doi.org/10.1038/nrm3891

Hu, J., Hwang, S. S., Liesa, M., Gan, B., Sahin, E., Jaskelioff, M., Ding, Z., Ying, H., Boutin, A. T., Zhang, H., Johnson, S., Ivanova, E., Kost-Alimova, M., Protopopov, A., Wang, Y. A., Shirihai, O. S.,

Chin, L., & DePinho, R. A. (2012). Antitelomerase therapy provokes ALT and mitochondrial adaptive mechanisms in cancer. *Cell*, *148*(4), 651–663. https://doi.org/10.1016/j.cell.2011.12.028

Inoue, A., & Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science (New York, N.Y.)*, *334*(6053), 194. https://doi.org/10.1126/science.1212483

Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D., & Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*, *363*(6429), 558–561. https://doi.org/10.1038/363558a0

Iurlaro, M., Ficz, G., Oxley, D., Raiber, E. A., Bachman, M., Booth, M. J., Andrews, S., Balasubramanian, S., & Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome biology*, *14*(10), R119. https://doi.org/10.1186/gb-2013-14-10-r119

Jiang, D., Wei, S., Chen, F., Zhang, Y., & Li, J. (2017). TET3-mediated DNA oxidation promotes ATR-dependent DNA damage response. *EMBO reports*, *18*(5), 781–796. https://doi.org/10.15252/embr.201643179

Jiang, W. Q., Zhong, Z. H., Henson, J. D., Neumann, A. A., Chang, A. C., & Reddel, R. R. (2005). Suppression of alternative lengthening of telomeres by Sp100-mediated sequestration of the MRE11/RAD50/NBS1 complex. *Molecular and cellular biology*, *25*(7), 2708–2721. https://doi.org/10.1128/MCB.25.7.2708-2721.2005

Kang, S., Jaworski, A., Ohshima, K., & Wells, R. D. (1995). Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. *Nature genetics*, *10*(2), 213–218. https://doi.org/10.1038/ng0695-213

Kano, H., Godoy, I., Courtney, C., Vetter, M. R., Gerton, G. L., Ostertag, E. M., & Kazazian, H. H., Jr (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes & development*, *23*(11), 1303–1312. https://doi.org/10.1101/gad.1803909

Kelkar, Y. D., Tyekucheva, S., Chiaromonte, F., & Makova, K. D. (2008). The genome-wide determinants of human and chimpanzee microsatellite evolution. *Genome research*, *18*(1), 30–38. https://doi.org/10.1101/gr.7113408

Kirschner, M., & Gerhart, J. (1998). Evolvability. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(15), 8420–8427. https://doi.org/10.1073/pnas.95.15.8420

Knoblich J. A. (2008). Mechanisms of asymmetric stem cell division. *Cell*, *132*(4), 583–597. https://doi.org/10.1016/j.cell.2008.02.007

Knudson A. G., Jr (1971). Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America*, *68*(4), 820–823. https://doi.org/10.1073/pnas.68.4.820

Koneru, B., Lopez, G., Farooqi, A., Conkrite, K. L., Nguyen, T. H., Macha, S. J., Modi, A., Rokita, J. L., Urias, E., Hindle, A., Davidson, H., Mccoy, K., Nance, J., Yazdani, V., Irwin, M. S., Yang, S., Wheeler, D. A., Maris, J. M., Diskin, S. J., & Reynolds, C. P. (2020). Telomere Maintenance

Mechanisms Define Clinical Outcome in High-Risk Neuroblastoma. *Cancer research*, *80*(12), 2663–2675. https://doi.org/10.1158/0008-5472.CAN-19-3068

Kollenstart, L., Biran, A., Alcaraz, N., Reverón-Gómez, N., Solis-Mezarino, V., Völker-Albert, M., Jenkinson, F., Flury, V., & Groth, A. (2025). Disabling leading and lagging strand histone transmission results in parental histones loss and reduced cell plasticity and viability. *Science advances*, *11*(8), eadr1453. https://doi.org/10.1126/sciadv.adr1453

Kovtun, I. V., & McMurray, C. T. (2001). Trinucleotide expansion in haploid germ cells by gap repair. *Nature genetics*, *27*(4), 407–411. https://doi.org/10.1038/86906

Kruglyak, S., Durrett, R. T., Schug, M. D., & Aquadro, C. F. (1998). Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(18), 10774–10778. https://doi.org/10.1073/pnas.95.18.10774

La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E., & Fischbeck, K. H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*, *352*(6330), 77–79. https://doi.org/10.1038/352077a0

Lai, Y., & Sun, F. (2003). The relationship between microsatellite slippage mutation rate and the number of repeat units. *Molecular biology and evolution*, *20*(12), 2123–2131. https://doi.org/10.1093/molbev/msg228

Lam, I., & Keeney, S. (2015). Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. *Science (New York, N.Y.)*, *350*(6263), 932–937. https://doi.org/10.1126/science.aad0814

Lancaster, M. S., Kim, B., Doud, E. H., Tate, M. D., Sharify, A. D., Gao, H., Chen, D., Simpson, E., Gillespie, P., Chu, X., Miller, M. J., Wang, Y., Liu, Y., Mosley, A. L., Kim, J., & Graham, B. H. (2023). Loss of succinyl-CoA synthetase in mouse forebrain results in hypersuccinylation with perturbed neuronal transcription and metabolism. *Cell reports*, *42*(10), 113241. https://doi.org/10.1016/j.celrep.2023.113241

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., … International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Laverde, E. E., Lai, Y., Leng, F., Balakrishnan, L., Freudenreich, C. H., & Liu, Y. (2020). R-loops promote trinucleotide repeat deletion through DNA base excision repair enzymatic activities. *The Journal of biological chemistry*, *295*(40), 13902–13913. https://doi.org/10.1074/jbc.RA120.014161

Le, D. T., Durham, J. N., Smith, K. N., Wang, H., Bartlett, B. R., Aulakh, L. K., Lu, S., Kemberling, H., Wilt, C., Luber, B. S., Wong, F., Azad, N. S., Rucki, A. A., Laheru, D., Donehower, R., Zaheer, A., Fisher, G. A., Crocenzi, T. S., Lee, J. J., Greten, T. F., … Diaz, L. A., Jr (2017). Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science (New York, N.Y.)*, *357*(6349), 409–413. https://doi.org/10.1126/science.aan6733

Le, D. T., Uram, J. N., Wang, H., Bartlett, B. R., Kemberling, H., Eyring, A. D., Skora, A. D., Luber, B. S., Azad, N. S., Laheru, D., Biedrzycki, B., Donehower, R. C., Zaheer, A., Fisher, G. A., Crocenzi, T. S., Lee, J. J., Duffy, S. M., Goldberg, R. M., de la Chapelle, A., Koshiji, M., … Diaz, L. A., Jr (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *The New England journal of medicine*, *372*(26), 2509–2520. https://doi.org/10.1056/NEJMoa1500596

Leach, F. S., Nicolaides, N. C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomäki, P., Sistonen, P., Aaltonen, L. A., & Nyström-Lahti, M. (1993). Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell*, *75*(6), 1215–1225. https://doi.org/10.1016/0092-8674(93)90330-s

Leclercq, S., Rivals, E., & Jarne, P. (2010). DNA slippage occurs at microsatellite loci without minimal threshold length in humans: a comparative genomic approach. *Genome biology and evolution*, *2*, 325–335. https://doi.org/10.1093/gbe/evq023

Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J., 3rd, Lohr, J. G., Harris, C. C., Ding, L., Wilson, R. K., Wheeler, D. A., Gibbs, R. A., Kucherlapati, R., Lee, C., Kharchenko, P. V., Park, P. J., & Cancer Genome Atlas Research Network (2012). Landscape of somatic retrotransposition in human cancers. *Science (New York, N.Y.)*, *337*(6097), 967–971. https://doi.org/10.1126/science.1222077

Lee-Six, H., Øbro, N. F., Shepherd, M. S., Grossmann, S., Dawson, K., Belmonte, M., Osborne, R. J., Huntly, B. J. P., Martincorena, I., Anderson, E., O'Neill, L., Stratton, M. R., Laurenti, E., Green, A. R., Kent, D. G., & Campbell, P. J. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature*, *561*(7724), 473–478. https://doi.org/10.1038/s41586-018-0497-0

Leopoldino, A. M., & Pena, S. D. (2003). The mutational spectrum of human autosomal tetranucleotide microsatellites. *Human mutation*, *21*(1), 71–79. https://doi.org/10.1002/humu.10153

Levinson, G., & Gutman, G. A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular biology and evolution*, *4*(3), 203–221. https://doi.org/10.1093/oxfordjournals.molbev.a040442

Li, L., Shi, L., Yang, S., Yan, R., Zhang, D., Yang, J., He, L., Li, W., Yi, X., Sun, L., Liang, J., Cheng, Z., Shi, L., Shang, Y., & Yu, W. (2016). SIRT7 is a histone desuccinylase that functionally links to chromatin compaction and genome stability. *Nature communications*, *7*, 12235. https://doi.org/10.1038/ncomms12235

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., & Ferguson-Smith, A. C. (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Developmental cell*, *15*(4), 547–557. https://doi.org/10.1016/j.devcel.2008.08.014

Llewellyn, J., Hubbard, S. J., & Swift, J. (2024). Translation is an emerging constraint on protein homeostasis in ageing. *Trends in cell biology*, *34*(8), 646–656. https://doi.org/10.1016/j.tcb.2024.02.001

Lodato, M. A., Rodin, R. E., Bohrson, C. L., Coulter, M. E., Barton, A. R., Kwon, M., Sherman, M. A., Vitzthum, C. M., Luquette, L. J., Yandava, C. N., Yang, P., Chittenden, T. W., Hatem, N. E., Ryu, S.

C., Woodworth, M. B., Park, P. J., & Walsh, C. A. (2018). Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science (New York, N.Y.)*, *359*(6375), 555–559. https://doi.org/10.1126/science.aao4426

Lopes, J., Piazza, A., Bermejo, R., Kriegsman, B., Colosio, A., Teulade-Fichou, M. P., Foiani, M., & Nicolas, A. (2011). G-quadruplex-induced instability during leading-strand replication. *The EMBO journal*, *30*(19), 4033–4046. https://doi.org/10.1038/emboj.2011.316

Lykke-Andersen, S., & Jensen, T. H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nature reviews. Molecular cell biology*, *16*(11), 665–677. https://doi.org/10.1038/nrm4063

Mamely, I., van Vugt, M. A., Smits, V. A., Semple, J. I., Lemmens, B., Perrakis, A., Medema, R. H., & Freire, R. (2006). Polo-like kinase-1 controls proteasome-dependent degradation of Claspin during checkpoint recovery. *Current biology : CB*, *16*(19), 1950–1955. https://doi.org/10.1016/j.cub.2006.08.026

Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R. S., Zborowska, E., Kinzler, K. W., & Vogelstein, B. (1995). Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science (New York, N.Y.)*, *268*(5215), 1336–1338. https://doi.org/10.1126/science.7761852

Massudi, H., Grant, R., Braidy, N., Guest, J., Farnsworth, B., & Guillemin, G. J. (2012). Age-associated changes in oxidative stress and NAD+ metabolism in human tissue. *PloS one*, *7*(7), e42357. https://doi.org/10.1371/journal.pone.0042357

Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y., & Vijg, J. (2017). Differences between germline and somatic mutation rates in humans and mice. *Nature communications*, *8*, 15183. https://doi.org/10.1038/ncomms15183

Miller, K. N., Victorelli, S. G., Salmonowicz, H., Dasgupta, N., Liu, T., Passos, J. F., & Adams, P. D. (2021). Cytoplasmic DNA: sources, sensing, and role in aging and disease. *Cell*, *184*(22), 5506–5526. https://doi.org/10.1016/j.cell.2021.09.034

Miquel, C., Jacob, S., Grandjouan, S., Aimé, A., Viguier, J., Sabourin, J. C., Sarasin, A., Duval, A., & Praz, F. (2007). Frequent alteration of DNA damage signalling and repair pathways in human colorectal cancers with microsatellite instability. *Oncogene*, *26*(40), 5919–5926. https://doi.org/10.1038/sj.onc.1210419

Miret, J. J., Pessoa-Brandão, L., & Lahue, R. S. (1997). Instability of CAG and CTG trinucleotide repeats in Saccharomyces cerevisiae. *Molecular and cellular biology*, *17*(6), 3382–3387. https://doi.org/10.1128/MCB.17.6.3382

Mirkin S. M. (2007). Expandable DNA repeats and human disease. *Nature*, *447*(7147), 932–940. https://doi.org/10.1038/nature05977

Mohapatra, J., Tashiro, K., Beckner, R. L., Sierra, J., Kilgore, J. A., Williams, N. S., & Liszczak, G. (2021). Serine ADP-ribosylation marks nucleosomes for ALC1-dependent chromatin remodeling. *eLife*, *10*, e71502. https://doi.org/10.7554/eLife.71502

Møller, P., Seppälä, T., Dowty, J. G., Haupt, S., Dominguez-Valentin, M., Sunde, L., Bernstein, I., Engel, C., Aretz, S., Nielsen, M., Capella, G., Evans, D. G., Burn, J., Holinski-Feder, E., Bertario, L., Bonanni, B., Lindblom, A., Levi, Z., Macrae, F., Winship, I., … European Hereditary Tumour Group (EHTG) and the International Mismatch Repair Consortium (IMRC) (2022). Colorectal cancer incidences in Lynch syndrome: a comparison of results from the prospective lynch syndrome database and the international mismatch repair consortium. *Hereditary cancer in clinical practice*, *20*(1), 36. https://doi.org/10.1186/s13053-022-00241-1

Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D., & Kazazian, H. H., Jr (1996). High frequency retrotransposition in cultured mammalian cells. *Cell*, *87*(5), 917–927. https://doi.org/10.1016/s0092-8674(00)81998-4

Mullaney, J. M., Mills, R. E., Pittard, W. S., & Devine, S. E. (2010). Small insertions and deletions (INDELs) in human genomes. *Human molecular genetics*, *19*(R2), R131–R136. https://doi.org/10.1093/hmg/ddq400

Muotri, A. R., Chu, V. T., Marchetto, M. C., Deng, W., Moran, J. V., & Gage, F. H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature*, *435*(7044), 903–910. https://doi.org/10.1038/nature03663

Oberlé, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boué, J., Bertheas, M. F., & Mandel, J. L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science (New York, N.Y.)*, *252*(5009), 1097–1102. https://doi.org/10.1126/science.252.5009.1097

Panagopoulos, A., Stout, M., Kilic, S., Leary, P., Vornberger, J., Pasti, V., Galarreta, A., Lezaja, A., Kirschenbühler, K., Imhof, R., Rehrauer, H., Ziegler, U., & Altmeyer, M. (2025). Multigenerational cell tracking of DNA replication and heritable DNA damage. *Nature*, *642*(8068), 785–795. https://doi.org/10.1038/s41586-025-08986-0

Panigrahi, G. B., Cleary, J. D., & Pearson, C. E. (2002). In vitro (CTG)*(CAG) expansions and deletions by human cell extracts. *The Journal of biological chemistry*, *277*(16), 13926–13934. https://doi.org/10.1074/jbc.M109761200

Papadopoulos, N., Nicolaides, N. C., Wei, Y. F., Ruben, S. M., Carter, K. C., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., & Adams, M. D. (1994). Mutation of a mutL homolog in hereditary colon cancer. *Science (New York, N.Y.)*, *263*(5153), 1625–1629. https://doi.org/10.1126/science.8128251

Park, S., Mali, N. M., Kim, R., Choi, J. W., Lee, J., Lim, J., Park, J. M., Park, J. W., Kim, D., Kim, T., Yi, K., Choi, J. H., Kwon, S. G., Hong, J. H., Youk, J., An, Y., Kim, S. Y., Oh, S. A., Kwon, Y., Hong, D., … Ju, Y. S. (2021). Clonal dynamics in early human embryogenesis inferred from somatic mutation. *Nature*, *597*(7876), 393–397. https://doi.org/10.1038/s41586-021-03786-8

Parsons, R., Li, G. M., Longley, M. J., Fang, W. H., Papadopoulos, N., Jen, J., de la Chapelle, A., Kinzler, K. W., Vogelstein, B., & Modrich, P. (1993). Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell*, *75*(6), 1227–1236. https://doi.org/10.1016/0092-8674(93)90331-j

Parsons R, Myeroff LL, Liu B, Willson JK, Markowitz SD, Kinzler KW, Vogelstein B. (1995). Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. Cancer Research, 55(23), 5548-5550.

Paulson H. (2018). Repeat expansion diseases. *Handbook of clinical neurology*, *147*, 105–123. https://doi.org/10.1016/B978-0-444-63233-3.00009-9

Pearson, C. E., Nichol Edamura, K., & Cleary, J. D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nature reviews. Genetics*, *6*(10), 729–742. https://doi.org/10.1038/nrg1689

Petrussa, L., Van de Velde, H., & De Rycke, M. (2014). Dynamic regulation of DNA methyltransferases in human oocytes and preimplantation embryos after assisted reproductive technologies. *Molecular human reproduction*, *20*(9), 861–874. https://doi.org/10.1093/molehr/gau049

Rajagopal, S., Donaldson, J., Flower, M., Hensman Moss, D. J., & Tabrizi, S. J. (2023). Genetic modifiers of repeat expansion disorders. *Emerging topics in life sciences*, *7*(3), 325–337. https://doi.org/10.1042/ETLS20230015

Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J. C., & Perucho, M. (1997). Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science (New York, N.Y.)*, *275*(5302), 967–969. https://doi.org/10.1126/science.275.5302.967

Ranjan, R., Snedeker, J., & Chen, X. (2019). Asymmetric Centromeres Differentially Coordinate with Mitotic Machinery to Ensure Biased Sister Chromatid Segregation in Germline Stem Cells. *Cell stem cell*, *25*(5), 666–681.e5. https://doi.org/10.1016/j.stem.2019.08.014

Ratnam, S., Mertineit, C., Ding, F., Howell, C. Y., Clarke, H. J., Bestor, T. H., Chaillet, J. R., & Trasler, J. M. (2002). Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development. *Developmental biology*, *245*(2), 304–314. https://doi.org/10.1006/dbio.2002.0628

Rhodes, D., & Lipps, H. J. (2015). G-quadruplexes and their regulatory roles in biology. *Nucleic acids research*, *43*(18), 8627–8637. https://doi.org/10.1093/nar/gkv862

Rose, O., & Falush, D. (1998). A threshold size for microsatellite expansion. *Molecular biology and evolution*, *15*(5), 613–615. https://doi.org/10.1093/oxfordjournals.molbev.a025964

Roudko, V., Cimen Bozkus, C., Greenbaum, B., Lucas, A., Samstein, R., & Bhardwaj, N. (2021). Lynch Syndrome and MSI-H Cancers: From Mechanisms to "Off-The-Shelf" Cancer Vaccines. *Frontiers in immunology*, *12*, 757804. https://doi.org/10.3389/fimmu.2021.757804

Samadashwily, G. M., Raca, G., & Mirkin, S. M. (2001). Replication and expansion of trinucleotide repeats in yeast. Molecular and Cellular Biology, 21(1), 297-308. https://doi.org/10.1128/MCB.21.1.297-308.2001

Schatz, D. G., & Ji, Y. (2011). Recombination centres and the orchestration of V(D)J recombination. *Nature reviews. Immunology*, *11*(4), 251–263. https://doi.org/10.1038/nri2941

Schmidt, K. H., Abbott, C. M., & Leach, D. R. (2000). Two opposing effects of mismatch repair on CTG repeat instability in Escherichia coli. *Molecular microbiology*, *35*(2), 463–471. https://doi.org/10.1046/j.1365-2958.2000.01727.x

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., & Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, *473*(7347), 337–342. https://doi.org/10.1038/nature10098

Sellou, H., Lebeaupin, T., Chapuis, C., Smith, R., Hegele, A., Singh, H. R., Kozlowski, M., Bultmann, S., Ladurner, A. G., Timinszky, G., & Huet, S. (2016). The poly(ADP-ribose)-dependent chromatin remodeler Alc1 induces local chromatin relaxation upon DNA damage. *Molecular biology of the cell*, *27*(24), 3791–3799. https://doi.org/10.1091/mbc.E16-05-0269

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M., & Koseki, H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, *450*(7171), 908–912. https://doi.org/10.1038/nature06397

Shastri, N., Tsai, Y. C., Hile, S., Jordan, D., Powell, B., Chen, J., Maloney, D., Dose, M., Lo, Y., Anastassiadis, T., Rivera, O., Kim, T., Shah, S., Borole, P., Asija, K., Wang, X., Smith, K. D., Finn, D., Schug, J., Casellas, R., … Brown, E. J. (2018). Genome-wide Identification of Structure-Forming Repeats as Principal Sites of Fork Collapse upon ATR Inhibition. *Molecular cell*, *72*(2), 222–238.e11. https://doi.org/10.1016/j.molcel.2018.08.047

Shay, J. W., & Wright, W. E. (2019). Telomeres and telomerase: three decades of progress. *Nature reviews. Genetics*, *20*(5), 299–309. https://doi.org/10.1038/s41576-019-0099-1

Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., & Inouye, M. (1966). Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. *Cold Spring Harbor symposia on quantitative biology*, *31*, 77–84. https://doi.org/10.1101/sqb.1966.031.01.014

Suarez, N. A., Macia, A., & Muotri, A. R. (2018). LINE-1 retrotransposons in healthy and diseased human brain. *Developmental neurobiology*, *78*(5), 434–455. https://doi.org/10.1002/dneu.22567

Tabrizi, S. J., Leavitt, B. R., Landwehrmeyer, G. B., Wild, E. J., Saft, C., Barker, R. A., Blair, N. F., Craufurd, D., Priller, J., Rickards, H., Rosser, A., Kordasiewicz, H. B., Czech, C., Swayze, E. E., Norris, D. A., Baumann, T., Gerlach, I., Schobel, S. A., Paz, E., Smith, A. V., … Phase 1–2a IONIS-HTTRx Study Site Teams (2019). Targeting Huntingtin Expression in Patients with Huntington's Disease. *The New England journal of medicine*, *380*(24), 2307–2316. https://doi.org/10.1056/NEJMoa1900907

Takahashi J. S. (2017). Transcriptional architecture of the mammalian circadian clock. *Nature reviews. Genetics*, *18*(3), 164–179. https://doi.org/10.1038/nrg.2016.150

Thibodeau, S. N., Bren, G., & Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. *Science (New York, N.Y.)*, *260*(5109), 816–819. https://doi.org/10.1126/science.8484122

Tinel, A., Janssens, S., Lippens, S., Cuenin, S., Logette, E., Jaccard, B., Quadroni, M., & Tschopp, J. (2007). Autoproteolysis of PIDD marks the bifurcation between pro-death caspase-2 and pro-survival NF-kappaB pathway. *The EMBO journal*, *26*(1), 197–208. https://doi.org/10.1038/sj.emboj.7601473

Traverse, C. C., & Ochman, H. (2017). Genome-Wide Spectra of Transcription Insertions and Deletions Reveal That Slippage Depends on RNA:DNA Hybrid Complementarity. *mBio*, *8*(4), e01230-17. https://doi.org/10.1128/mBio.01230-17

Upton, K. R., Gerhardt, D. J., Jesuadian, J. S., Richardson, S. R., Sánchez-Luque, F. J., Bodea, G. O., Ewing, A. D., Salvador-Palomeque, C., van der Knaap, M. S., Brennan, P. M., Vanderver, A., & Faulkner, G. J. (2015). Ubiquitous L1 mosaicism in hippocampal neurons. *Cell*, *161*(2), 228–239. https://doi.org/10.1016/j.cell.2015.03.026

Usdin, K., & Woodford, K. J. (1995). CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucleic acids research*, *23*(20), 4202–4209. https://doi.org/10.1093/nar/23.20.4202

Usdin, K., House, N. C., & Freudenreich, C. H. (2015). Repeat instability during DNA repair: Insights from model systems. *Critical reviews in biochemistry and molecular biology*, *50*(2), 142–167. https://doi.org/10.3109/10409238.2014.999192

Verdin E. (2015). NAD^+ in aging, metabolism, and neurodegeneration. *Science (New York, N.Y.)*, *350*(6265), 1208–1213. https://doi.org/10.1126/science.aac4854

Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., & Zhang, F. P. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, *65*(5), 905–914. https://doi.org/10.1016/0092-8674(91)90397-h

Viterbo, D., Michoud, G., Mosbach, V., Dujon, B., & Richard, G. F. (2016). Replication stalling and heteroduplex formation within CAG/CTG trinucleotide repeats by mismatch repair. *DNA repair*, *42*, 94–106. https://doi.org/10.1016/j.dnarep.2016.03.002

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr, & Kinzler, K. W. (2013). Cancer genome landscapes. *Science (New York, N.Y.)*, *339*(6127), 1546–1558. https://doi.org/10.1126/science.1235122

Wang, Y., Guo, Y. R., Liu, K., Yin, Z., Liu, R., Xia, Y., Tan, L., Yang, P., Lee, J. H., Li, X. J., Hawke, D., Zheng, Y., Qian, X., Lyu, J., He, J., Xing, D., Tao, Y. J., & Lu, Z. (2017). KAT2A coupled with the α-KGDH complex acts as a histone H3 succinyltransferase. *Nature*, *552*(7684), 273–277. https://doi.org/10.1038/nature25003

Watanabe, D., Suetake, I., Tada, T., & Tajima, S. (2002). Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mechanisms of development*, *118*(1-2), 187–190. https://doi.org/10.1016/s0925-4773(02)00242-3

Watson, C. J., Papula, A. L., Poon, G. Y. P., Wong, W. H., Young, A. L., Druley, T. E., Fisher, D. S., & Blundell, J. R. (2020). The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science (New York, N.Y.)*, *367*(6485), 1449–1454. https://doi.org/10.1126/science.aay9333

Werner, B., Case, J., Williams, M. J., Chkhaidze, K., Temko, D., Fernández-Mateos, J., Cresswell, G. D., Nichol, D., Cross, W., Spiteri, I., Huang, W., Tomlinson, I. P. M., Barnes, C. P., Graham, T. A., & Sottoriva, A. (2020). Measuring single cell divisions in human tissues from multi-region sequencing data. *Nature communications*, *11*(1), 1035. https://doi.org/10.1038/s41467-020-14844-6

Winquist, R. J., Hall, A. B., Eustace, B. K., & Furey, B. F. (2014). Evaluating the immortal strand hypothesis in cancer stem cells: symmetric/self-renewal as the relevant surrogate marker of tumorigenicity. *Biochemical pharmacology*, *91*(2), 129–134. https://doi.org/10.1016/j.bcp.2014.06.007

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., & Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature communications*, *2*, 241. https://doi.org/10.1038/ncomms1240

Xing, M., Zhang, F., Liao, H., Chen, S., Che, L., Wang, X., Bao, Z., Ji, F., Chen, G., Zhang, H., Li, W., Chen, Z., Liu, Y., Hickson, I. D., Shen, H., & Ying, S. (2020). Replication Stress Induces ATR/CHK1-Dependent Nonrandom Segregation of Damaged Chromosomes. *Molecular cell*, *78*(4), 714–724.e5. https://doi.org/10.1016/j.molcel.2020.04.005

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S. H., Ito, S., Yang, C., Wang, P., Xiao, M. T., Liu, L. X., Jiang, W. Q., Liu, J., Zhang, J. Y., Wang, B., Frye, S., Zhang, Y., Xu, Y. H., Lei, Q. Y., Guan, K. L., … Xiong, Y. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer cell*, *19*(1), 17–30. https://doi.org/10.1016/j.ccr.2010.12.014

Yamamoto, H., & Imai, K. (2015). Microsatellite instability: an update. *Archives of toxicology*, *89*(6), 899–921. https://doi.org/10.1007/s00204-015-1474-0

Yamashita, Y. M., Mahowald, A. P., Perlin, J. R., & Fuller, M. T. (2007). Asymmetric inheritance of mother versus daughter centrosome in stem cell division. *Science (New York, N.Y.)*, *315*(5811), 518–521. https://doi.org/10.1126/science.1134910

Yu, C., Gan, H., Serra-Cardona, A., Zhang, L., Gan, S., Sharma, S., Johansson, E., Chabes, A., Xu, R. M., & Zhang, Z. (2018). A mechanism for preventing asymmetric histone segregation onto replicating DNA strands. *Science (New York, N.Y.)*, *361*(6409), 1386–1389. https://doi.org/10.1126/science.aat8849

Yukawa, M., Oda, S., Mitani, H., Nagata, M., & Aoki, F. (2007). Deficiency in the response to DNA double-strand breaks in mouse early preimplantation embryos. *Biochemical and biophysical research communications*, *358*(2), 578–584. https://doi.org/10.1016/j.bbrc.2007.04.162

Zeraati, M., Langley, D. B., Schofield, P., Moye, A. L., Rouet, R., Hughes, W. E., Bryan, T. M., Dinger, M. E., & Christ, D. (2018). I-motif DNA structures are formed in the nuclei of human cells. *Nature chemistry*, *10*(6), 631–637. https://doi.org/10.1038/s41557-018-0046-3

Zhang, CZ., Mendez-Dorantes, C., Burns, K.H. *et al.* A breakage–replication/fusion process explains complex rearrangements and segmental DNA amplification. *Nat Genet* (2026). https://doi.org/10.1038/s41588-025-02434-5

Zhang, L., Lee, M., Hao, X., Ehlert, J., Chi, Z., Jin, B., Maslov, A. Y., Barabási, A. L., Hoeijmakers, J. H. J., Edelmann, W., Vijg, J., & Dong, X. (2024). Negative Selection Allows DNA Mismatch Repair-Deficient Mouse Fibroblasts *In Vitro* to Tolerate High Levels of Somatic Mutations. *bioRxiv : the preprint server for biology*, 2024.05.04.592535. https://doi.org/10.1101/2024.05.04.592535

Zhang, M., Kothari, P., Mullins, M., & Lampson, M. A. (2014). Regulation of zygotic genome activation and DNA damage checkpoint acquisition at the mid-blastula transition. *Cell cycle (Georgetown, Tex.)*, *13*(24), 3828–3838. https://doi.org/10.4161/15384101.2014.967066