

The Architecture of Cell Fate Determination: The ARC Model

Abstract

Cell fate decisions in renewing tissues require coordination among chromatin modification, organelle partitioning, and cell death machinery. This manuscript proposes the ARC (Asymmetry–Remodeling–Commitment) model, in which DDR-associated enzymes function as architects of the chromatin transitions that specify cell fate through write/erase chemistry at specific genomic addresses, with partition-mediated organelle inheritance coupling the nuclear program to the cytoplasmic outcome.

The model organizes cell fate progression into three molecular states orthogonal to the cell cycle: Asymmetry (initiating partition that creates molecularly divergent daughters), Remodeling (buffered chromatin changes at commitment-relevant addresses), and Commitment (irreversible resolution through compound sequence, epigenetic, and structural changes). The same enzymatic machinery that executes regulated cell death operates at sub-lethal intensity during terminal differentiation, with the partition-inherited organelle configuration determining which death-associated chemistry each daughter engages. Cancer is interpreted as a disruption of the regulatory coordination that governs these transitions, producing cells that retain functional execution machinery but have lost the partition-dependent programs that would direct it toward coherent outcomes.

Computational analysis across nine TCGA cancer types, seven CPTAC proteomic cohorts, single-cell data from hematopoietic differentiation, and a mouse cross-species atlas provides evidence that the coordination the model describes is present in normal tissue, varies across tissues in ways that track regenerative capacity, changes progressively along differentiation

trajectories, and dissolves selectively in cancer. Pairwise coupling between regulatory coordination components weakens systematically in tumor tissue while structurally obligate execution partnerships are preserved — a selective dissolution reproduced at protein level, at single-cell resolution, and across species. Five lineage-specific failure modes are identified and tested against representative cancers. The proposed mutagenic mechanism — that partition failure accelerates mutation accumulation at commitment-relevant regulatory addresses through missegregation and mislicensing of the write/erase chemistry — remains a hypothesis whose direct experimental test has not been performed.

Part 1: Introduction

Every cell division in a renewing tissue confronts the same fundamental problem: the dividing cell must produce daughters whose fates are appropriate for the tissue's current needs. A tissue that needs to expand its progenitor pool requires symmetric self-renewal — two equivalent daughters. A tissue that needs to replace a lost differentiated cell requires asymmetric division — one self-renewing daughter and one committed daughter. A tissue that needs to eliminate a damaged cell requires the damaged cell to die through a modality that signals the right information to its neighbors. The mechanisms by which individual cells execute each of these outcomes are increasingly well characterized at the molecular level. What is missing is an account of how the mechanisms are coordinated as a system — how the cell integrates the signals that specify the outcome, executes the molecular changes that produce it, and partitions the products to the correct daughter.

The Landscape of Existing Models

Three research fields have addressed parts of this problem, each illuminating a critical dimension while leaving the integrative question unanswered.

The asymmetric division field, pioneered in *Drosophila* neuroblasts and extended to mammalian systems over three decades, has characterized the molecular machinery that generates unequal daughters at mitosis: the PAR complex organizes cortical polarity, aPKC phosphorylation directs the asymmetric localization of fate determinants — NUMB, TRIM32, MIRANDA, PROSPERO — and their mammalian orthologs ensure that each daughter inherits a distinct complement of regulatory proteins (Knoblich, 2010; Gönczy, 2008; Betschinger & Knoblich, 2004). What this literature has not produced is an account of how the cortical partition coordinates with the chromatin-level changes that specify each daughter's gene expression program. The polarity

proteins are partitioned asymmetrically, and the chromatin marks that determine lineage identity are installed differentially — but the mechanism that couples the cortical partition to the chromatin outcome has not been described as an integrated system. The gap is not in the parts but in the architecture that connects them.

The cancer stem cell hypothesis, formalized by Bonnet and Dick (1997) and elaborated over two decades, proposes that tumors are maintained by a subpopulation of cells with stem-like self-renewal capacity. The hypothesis has generated productive research in leukemia, breast cancer, and glioblastoma, establishing that not all cancer cells contribute equally to tumor maintenance. But it raises a question it does not answer: what molecular mechanism gives the cancer stem cell its stem-like properties? If the cancer stem cell is defined by its ability to self-renew and generate heterogeneous progeny, then the mechanism that governs the self-renewal-versus-differentiation decision in these cells is the central explanatory target. The cancer stem cell phenotype is a description of what the cell does; the question is how it does it and what has gone wrong.

The Waddington landscape metaphor, formalized computationally by Huang et al. (2005) and by attractor-state models of gene regulatory networks, describes cell fate as movement through an abstract potential-energy surface whose topology determines the available stable states. The metaphor has been productive for understanding how cells transition between discrete identity states. But it does not specify the molecular mechanism that generates the landscape's topology — what determines the number, depth, and arrangement of the valleys that cells settle into. The landscape describes the space of possibilities; it does not explain what builds the space.

An Evolutionary Observation

A fourth consideration, less commonly discussed in the cell fate literature, bears on why an integrated architectural program for fate decisions might exist at all. Since King and Wilson (1975) demonstrated that regulatory changes rather than protein-coding changes drive morphological divergence between species, the field has recognized that the evolution of complex multicellular form depends on coordinated changes at cis-regulatory elements — promoters, enhancers, and other non-coding regulatory DNA (Carroll, 2005). Yet the mathematical constraints on such coordination are severe. The expected waiting time for two specific mutations to arise in a population scales as $1/(N \cdot \mu^2)$, where N is effective population size and μ is the per-site mutation rate; for three coordinated changes the waiting time becomes $1/(N \cdot \mu^3)$ (Durrett & Schmidt, 2008). At human effective population sizes and mutation rates, the waiting time for three coordinated regulatory mutations at random genomic positions exceeds the timescale of vertebrate evolutionary history by orders of magnitude.

Recent work has begun to challenge the assumption that mutation is random with respect to genomic position. Monroe et al. (2022) demonstrated in *Arabidopsis* that mutation frequency is reduced by half inside gene bodies and by two-thirds in essential genes, with epigenomic and physical features explaining over 90% of variance in the genome-wide mutation pattern. Schuster-Böckler and Lehner (2012) showed that chromatin organization is a major influence on regional mutation rates in human cancer, and Supek and Lehner (2015) identified differential mismatch repair as one mechanism underlying this variation. Mutation rates vary over 100-fold across the mammalian genome (Hodgkinson & Eyre-Walker, 2011), with the variation correlated to replication timing, chromatin state, and DNA secondary structure.

An observation that has received less attention connects these findings to the chemistry of differentiation. The two most common point mutation signatures in mammalian genomes —

C → T at methylated CpG dinucleotides and G → A at guanine-rich sequences — are precisely the chemical outputs of TET/TDG-mediated active demethylation at CpG sites and OGG1-mediated repair of 8-oxoguanine at G-quadruplex-forming sequences. These are not obscure or peripheral mutation types; CpG → TpG alone accounts for approximately one-quarter of all human point mutations (Alexandrov et al., 2013, in somatic tissue; Kong et al., 2012, in the germline) despite CpG dinucleotides comprising roughly one percent of the genome. The chemistry that produces the dominant mutational signatures in mammalian evolution is the same chemistry that the ARC model identifies as the write/erase program operating during cell fate commitment. Whether this correspondence is coincidental or reflects a deeper connection between developmental chromatin chemistry and evolutionary mutation patterns is a question the model raises but cannot resolve with the data presented here.

The model's central claim — that the DDR serves as developmental architect — may appear to contradict decades of research establishing the DDR as a reactive guardian system. However, the scientific community already accepts a major exception to this paradigm. Activation-induced cytidine deaminase (AID), discovered by Muramatsu et al. (2000), deliberately mutates immunoglobulin genes through targeted cytidine deamination in B lymphocytes, using base excision repair and mismatch repair as downstream effectors of a programmed mutagenesis event (Di Noia & Neuberger, 2007). AID-mediated somatic hypermutation is not damage — it is licensed mutagenesis serving an essential developmental function. Vitale et al. (2017) reviewed the broader evidence that stem cells across multiple lineages depend on specialized DNA damage response pathways whose disruption produces developmental abnormalities. The model proposes that AID represents a prototype for a broader principle: renewing tissues deploy DDR-associated chemistry — oxidation, excision, repair, translesion synthesis — as regulated,

address-specific tools for fate-specifying chromatin transitions, with the specific enzymes and genomic addresses varying by lineage.

The DDR-as-Architect Thesis

This manuscript proposes that the three gaps identified above — the coupling between cortical partition and chromatin outcome, the mechanism of the cancer stem cell phenotype, and the molecular generator of the fate landscape — are filled by a single integrated system. The system coordinates DNA damage response components, polarity proteins, mitochondrial dynamics, and cell death machinery into an architectural program that the manuscript terms the ARC (Asymmetry–Remodeling–Commitment) model.

The central claim is that DNA damage response components are not merely guardians of genome integrity. The model proposes that these components participate directly in executing the chromatin transitions required for differentiation — not as a secondary or co-opted function, but as a primary operating mode of the same enzymatic activities. This is not a metaphor. It is a mechanistic hypothesis about specific enzymes — TET2, OGG1, NEIL1, TDG, APEX1 — performing specific chemistry — oxidation, excision, repair — at specific genomic addresses — G-quadruplex structures, CpG islands, bivalent promoters — during specific temporal windows across the cell divisions that comprise the ARC cycle. In the ARC model, DDR-associated oxidation, excision, repair, and bypass chemistry constitutes a major enzymatic layer of the differentiation machinery, integrated with canonical transcription-factor, signaling, and polarity systems. The partition system that distributes these products at mitosis is proposed as a cell fate decision mechanism.

What This Model Adds

Terminology note: This manuscript uses several terms in specific technical senses that may differ from their colloquial use. "Partition" refers to the physical asymmetric distribution of proteins and organelles at mitosis; "marking" refers to covalent DNA and histone modifications at specific genomic addresses.

The novel contribution of this manuscript is not the identification of new molecular components — every gene in the model's panel has been studied extensively in isolation — but the proposal that the components described above operate as an integrated architectural system whose coupling patterns are computationally measurable and whose disruption in cancer follows a predictable, testable pattern.

The ARC model proposes that these three systems are not independent but form a single coordinated architecture — one that links nuclear marks to cytoplasmic organelle configurations at every cell division.

These claims are testable because the coupling patterns make specific, falsifiable predictions: which gene-pair correlations should differ across tissues, how coordination should change as cells differentiate, what should happen at lineage branch points, and which patterns cancer should disrupt while leaving others intact. Part 6 provides computational evidence from bulk transcriptomic, proteomic, single-cell, and cross-species data.

Part 2 defines the model's three phases, functional categories, and their relationship to the cell cycle. Part 3 describes the molecular implementation across five functional modules. Part 4 develops the death-as-commitment-endpoint hypothesis. Part 5 derives the cancer predictions. Part 6 provides computational evidence that the coordination the model describes is measurable, tissue-specific, developmentally structured, and selectively disrupted in cancer. Parts 7 and 8

discuss the implications and limitations, including what the model does not prove and what would falsify it. Part 9 provides methods for full reproducibility.

Part 2: The ARC Model

2.1 Three Phases Defined by Molecular State

The ARC model proposes that every dividing cell progressing through a fate decision occupies one of three molecular states — Asymmetry, Remodeling, or Commitment — and that the transitions between these states are governed by the partition architecture described in Part 3. The three states are defined by the cell's current molecular configuration rather than by position in a fixed developmental sequence, which can be assessed through the coupling relationships among the model's component genes. A cell's ARC phase determines what kind of division it will undergo, what chromatin chemistry is active, and what range of fates is available to its daughters (Figure 1).

Figure 1. The ARC Architecture of Cell Fate Three molecular states (ARC phases) operate orthogonally to the cell cycle to generate fate-appropriate outcomes.

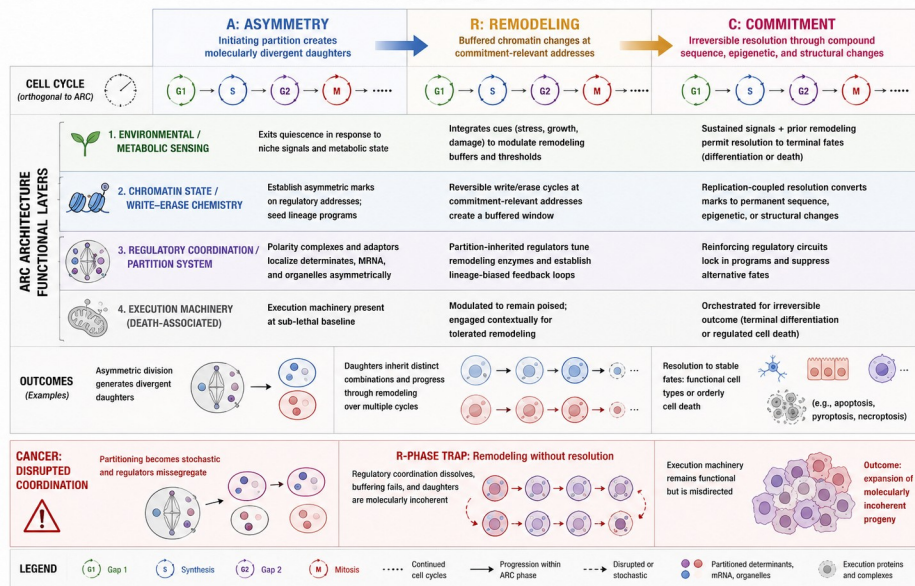


Figure 1. The ARC architecture of cell fate. Three molecular states — Asymmetry (A), Remodeling (R), and Commitment (C) — operate orthogonally to the cell cycle to govern fate-appropriate outcomes. Top row: ARC phases with defining features (A, initiating partition creates molecularly divergent daughters; R, buffered chromatin changes at commitment-relevant addresses over multiple cell cycles; C, irreversible resolution through compound sequence, epigenetic, and structural changes). Cell cycle progression (G1 → S → G2 → M) is shown within each phase to illustrate orthogonality. Middle rows: four functional layers — environmental/metabolic sensing, chromatin state/write-erase chemistry, regulatory coordination/partition system, and execution machinery (death-associated) — with phase-specific roles. Outcomes row: normal tissue produces coherent differentiation or orderly death; cancer produces molecularly incoherent progeny through R-phase trapping with dissolved regulatory coordination and preserved execution machinery.

2.1.1 A-phase: Asymmetry

A-phase is the initiating asymmetric division — the event in which a cell exits quiescence and divides in a manner that creates divergent outcomes in its daughters through the differential sorting of chromatin marks, organelles, and mRNA. A-phase is named for what defines it: the generation of asymmetry. A cell entering A-phase has three molecular prerequisites: its partition system is intact (the polarity proteins — NUMB, PRKCI, PRKCZ, DLG1, PARD3 — are coupled in expression, reflecting the coordinated program that distributes them asymmetrically at mitosis), its chromatin maintenance machinery is active (DNMT1 copies the CpG methylation

pattern onto newly synthesized strands at each replication; Hermann et al., 2004), and a niche-derived signal has activated the cell to exit quiescence. When these conditions are met, the partition system executes an asymmetric mitosis, and the two daughters inherit different complements of chromatin-marked regions, organelles with different configurations (§3.2), and differentially localized mRNAs. This differential sorting is the molecular event that initiates the fate-decision process. The daughters differ in molecular configuration — one biased toward self-renewal, the other toward commitment — and their ultimate fate is resolved through competition for niche occupancy rather than being specified cell-autonomously at the moment of division. This distinction is critical because the model must be compatible with the population-level observation that Lgr5⁺ intestinal stem cell dynamics follow neutral drift arising from morphologically symmetric divisions with stochastic niche competition (Snippert et al., 2010). The model does not claim that these divisions are "asymmetric" in the classical *Drosophila* neuroblast sense of producing morphologically unequal daughters with predetermined fates. Multiple molecular components segregate asymmetrically at division even between daughters that are morphologically indistinguishable: histones (Tran et al., 2012; Wooten et al., 2019), organelles (Katajisto et al., 2015; Loeffler et al., 2019, 2022; Bunel et al., 2025), and DNA damage marks (Wang et al., 2026). The population-level neutral dynamics observed by Snippert and colleagues emerge from stochastic competition among daughters whose individual molecular configurations differ (§2.3).

Direct experimental evidence for DDR-mediated asymmetric segregation determining daughter cell fate has been provided by Wang et al. (2026), who used long-term live-cell imaging of mouse embryonic stem cells to demonstrate that cells entering a two-cell-embryo-like state undergo asymmetric division in which accumulated DNA damage — visualized as 53BP1 foci

— segregates preferentially to one daughter. The daughter inheriting more 53BP1 foci is eliminated, while the daughter inheriting fewer foci reverts to pluripotency with reduced damage and enhanced chimeric contribution ($n = 54$ division events, $p = 0.008$). Critically, inhibition of ATM, ATR, CHEK, or PARP each reduced both the frequency of asymmetric divisions and the degree of 53BP1 asymmetry between sisters, establishing that the DNA damage response pathway physically implements the asymmetric segregation that creates fate divergence.

The pattern Wang and colleagues identified in mouse ESCs extends a decade-long line of evidence from *Drosophila* male germline stem cells. Tran et al. (2012) showed that pre-existing canonical H3 (but not variant H3.3) is selectively segregated to the GSC while newly synthesized H3 is enriched in the differentiating daughter — an asymmetry that is lost when GSCs are forced to divide symmetrically. Xie et al. (2015) identified H3 threonine-3 phosphorylation (H3T3P) as the molecular mark that distinguishes pre-existing from newly synthesized H3, and showed that genetically disrupting this distinction (through H3T3A or H3T3D substitutions) produces both stem cell loss and germline tumors — establishing that asymmetric histone inheritance is causally required for normal stem cell behavior. Wooten et al. (2019) extended the asymmetric-inheritance finding to H4 (while showing that H2A and H2B inherit symmetrically — the asymmetry is specific to histones carrying the lineage-specifying PTMs) and identified an underlying replication-fork mechanism: old H3 preferentially incorporates on the leading strand while newly synthesized H3 enriches on the lagging strand, with biased fork movement in GSC chromatin coupling strand identity to histone age. The data across these studies are consistent with a model in which the DNA damage response and replication machinery together organize the asymmetric inheritance that determines which

daughter survives — an interpretation the present manuscript develops mechanistically in Parts 3–5 and tests computationally in Part 6.

A-phase requires active cell division. Quiescent cells — a long-term hematopoietic stem cell in the bone marrow niche, a satellite cell in resting muscle, a hepatocyte in homeostatic liver — have intact partition machinery and stable chromatin but are in a pre-ARC state, poised to enter A-phase when a niche signal activates them. A transit-amplifying cell that has already undergone the initiating asymmetric division is in R-phase. The distinction between a quiescent stem cell (outside ARC, poised for A-phase entry) and a transit-amplifying cell (in R-phase, actively remodeling) is central to the model's account of tissue homeostasis.

2.1.2 R-phase: Remodeling

R-phase — the Remodeling phase — begins when a niche-derived signal initiates the write/erase program at specific genomic addresses. The signal may be a cytokine, a cell-cell contact, a metabolic shift in the local microenvironment, or a change in extracellular matrix composition. The model does not specify the signal's identity because different tissues use different signals to initiate the same architectural transition. What the model specifies is the molecular consequence: TET2 and TET3 begin oxidizing 5-methylcytosine to 5-hydroxymethylcytosine (Tahiliani et al., 2009; Ito et al., 2010) at bivalent promoters — promoters carrying both activating H3K4me3 and repressive H3K27me3 marks (Bernstein et al., 2006) — that govern lineage-specific gene expression. The BER glycosylases TDG, OGG1, and NEIL1 begin excising the oxidized bases to complete the demethylation cycle, and the histone demethylases KDM1A (Shi et al., 2004), KDM5A, KDM6A, and KDM6B (Agger et al., 2007) begin removing repressive histone marks at the same addresses. The cell's chromatin is now actively changing — it is being remodeled from one stable configuration toward another. The critical property of R-phase is that these

changes are buffered and conditionally modifiable. If the niche signal withdraws before sufficient irreversible changes have accumulated at commitment-relevant addresses (§3.4), the cell can exit R-phase and return to quiescence: DNMT1 maintenance methylation at the next replication restores the unmethylated CpGs to their prior state, the histone marks are re-established by the writers (EZH2 for H3K27me3, Cao et al., 2002; the DNMT3 enzymes for de novo methylation), and the cell resumes its quiescent, pre-ARC configuration as though the R-phase episode had not occurred. R-phase is the model's account of the transit-amplifying compartment observed in most renewing tissues (Barker et al., 2007): cells that are cycling with active chromatin remodeling but have not yet committed to a terminal fate.

The model does not specify how many divisions a cell undergoes in R-phase, because this is tissue-specific and depends on the balance between the write/erase program's progression rate and the niche signal's persistence. Critically, R-phase divisions can be either asymmetric or symmetric. Symmetric R-phase divisions expand the progenitor pool — both daughters inherit comparable configurations and continue the write/erase program in parallel, with CLASPIN coordinating symmetric parental-histone inheritance at the replication fork (Charlton et al., 2024). Asymmetric R-phase divisions produce daughters with distinct configurations, one more advanced toward commitment and one retaining an earlier state, through differential engagement of the partition system (§3.2). The alternation between symmetric and asymmetric R-phase divisions is governed by the niche environment and the partition system's engagement: a progenitor in a niche that favors expansion divides symmetrically, while one in a niche that favors differentiation divides asymmetrically. In the intestinal epithelium, R-phase may last two to four divisions in the transit-amplifying zone. In the hematopoietic system, the multipotent progenitor compartment may sustain R-phase cycling for weeks. In neural development, the

intermediate progenitor may divide once or twice in R-phase before committing (Götz & Huttner, 2005). The model is agnostic to duration because it defines states by molecular configuration, not by time. What matters is not how long R-phase lasts but what molecular condition defines the boundary between R-phase and C-phase.

2.1.3 C-phase: Commitment

C-phase — the Commitment phase — is the final division in which commitment becomes irreversible. The C-phase division can produce symmetric outcomes (two daughters of the same committed type) or asymmetric outcomes (one differentiating daughter and one dying daughter, or two daughters of distinct committed types), depending on the tissue and the niche context. The model proposes that irreversibility arises from the compound accumulation of permanent changes at commitment-relevant addresses through multiple convergent mechanisms. Stable DNA methylation patterns maintained by DNMT1, heritable histone marks propagated through replication, and chromatin compartment transitions all contribute to durable epigenetic locking. The resolution architecture of §3.4 describes additional pathways — spanning S-phase fork-level events through mitotic and post-mitotic resolution — including fork-uncoupling, translesion synthesis, mitotic DNA synthesis (MiDAS, occurring during mitosis itself), and post-mitotic 53BP1 NB-mediated resolution of transmitted intermediates — whose relative contributions to normal commitment remain an open empirical question. The SBS1 mutational clock — dominated by CpG → TpG transitions — accumulates in all renewing tissues proportionally to division count, consistent with the C → T write/erase chemistry of §3.3.1. A complementary oxidative signature (SBS18, dominated by C → A transversions at G-rich sequences — the complement-strand readout of the G:C → T:A transversions produced by 8-oxoG bypass described in §3.3.2) provides a distinct but related division-correlated signal. Whether these

mutations concentrate at the commitment-relevant addresses the model predicts, or are distributed more broadly, is the key unresolved question.

These mechanisms are not independent: a single replication failure at a G4-containing commitment address can simultaneously produce a sequence change from low-fidelity TLS, strip epigenetic marks through fork-uncoupling from histone recycling, and generate underreplicated DNA whose post-mitotic resolution installs structural changes (§3.4). None of the sequence-level or structural changes can be reversed by demethylation or histone remodeling alone. Once sufficient changes have accumulated at lineage-commitment addresses, the cell's regulatory repertoire has been narrowed to a single coherent lineage program, and the cell's fate is determined.

Commitment encompasses a spectrum of endpoints, of which terminal differentiation and cell death are variations rather than separate outcomes. A neuron that has committed to its identity lives for decades after commitment. A cardiomyocyte functions for a lifetime without re-entering the cell cycle. An aged neutrophil undergoes programmed death. All three have completed the ARC process — their fate is irreversibly determined. The model proposes that death and differentiation share overlapping molecular mechanisms precisely because they are endpoints on the same commitment axis: the partition state and the chromatin configuration at the moment of commitment determine where on this axis the cell lands. In many lineages, the execution of the committed fate involves the activation of death-associated proteins at sub-lethal levels — caspase-3 for local dendrite and spine pruning (Ertürk, Wang & Sheng, 2014), lipid peroxidation for erythrocyte membrane remodeling (Stockwell et al., 2017), and MLKL-dependent regulation of adipocyte differentiation (Magusto et al., 2022). These are not aberrant activations of death machinery during differentiation; they are the same molecular process operating at different

intensities along the commitment spectrum. Part 4 develops this death-differentiation continuum in detail.

2.1.4 Phase Transitions and Cancer Interpretation

The transition architecture of the three phases has a strict directionality under normal conditions. Entry from quiescence into A-phase is initiated by niche signal, and the A-phase division generates daughters that enter R-phase. R-phase chromatin remodeling is conditionally modifiable while insufficient irreversible changes have accumulated — if the niche signal withdraws, DNMT1 restores methylation at the next replication, histone writers re-establish repressive marks, and the few sequence changes or 53BP1 NB-resolved addresses produced in early R-phase cycles are insufficient to lock the configuration. The cell divisions themselves cannot be undone, but the chromatin state can be restored to a pre-remodeling configuration. R-to-C is gated by the resolution architecture of §3.4 and is irreversible once a sufficient number of lineage-relevant regulatory addresses have accumulated convergent irreversible changes — sequence, epigenetic, and structural. There is no C-to-R or post-commitment reversal in normal tissue homeostasis. Wound healing and injury responses can engage dedifferentiation programs in specific tissues — notably hepatocyte dedifferentiation after liver injury and intestinal cell plasticity when crypt stem cells are depleted — but these are damage-triggered exceptions to homeostatic turnover rather than features of it. The forced reversal of C-phase back to an R-phase-like state — the essential feat of Yamanaka-factor reprogramming — is consistent with this architecture: reprogramming requires reversing compound irreversible changes (sequence, epigenetic, and structural) at commitment-relevant addresses, not merely reversing methylation state alone, and the low efficiency of reprogramming (~0.1–1% of transduced cells; Takahashi & Yamanaka, 2006) is consistent with the difficulty of reversing three independent classes of

irreversible change simultaneously. The model predicts that reprogramming efficiency should correlate with the total burden of irreversible changes at commitment-relevant addresses, and that tissues with fewer such changes (less terminally differentiated cells) should reprogram more efficiently — a prediction consistent with the observation that progenitor cells reprogram more readily than terminally differentiated cells (Eminli et al., 2009).

Cancer, in this model, is interpreted as a disruption of the R-to-C transition that can arise through diverse molecular routes (Part 5). One architectural consequence is an R-phase-trapped state: the write/erase program is active, the chromatin is being remodeled, but the resolution architecture at commitment-relevant addresses is disrupted — TLS operates without helicase-mediated fidelity, 53BP1 NB-gated post-mitotic resolution is abnormal, and the quality-control mechanisms that would normally eliminate incoherent daughters are not engaged. The cell cycles indefinitely in R-phase, producing daughters with actively remodeling chromatin that never resolves into a stable committed state. The partition system, which in normal tissue would coordinate the distribution of remodeling products between daughters to produce coherent outcomes (one more committed, one less committed, or both equivalently advanced), has been disrupted (Part 5). Both daughters receive incoherent combinations of partially remodeled marks and partially redistributed organelles, and neither daughter can complete commitment or exit the remodeling process. This is one cancer stem cell phenotype the model predicts: not a special cell type with intrinsic stemness properties, but a cell trapped in R-phase by architectural disruption, cycling with active chromatin remodeling that produces neither clean commitment nor coherent self-renewal. Other failure-mode configurations — progenitor lock-in, TF-target bifurcation, lineage-specific decoupling, and quiescent evasion — produce phenotypically distinct lesions through

the same underlying partition failure operating along different architectural axes, and are developed in §5.4.

2.2 Functional Categories of Model Components

The model's approximately fifty molecular components are organized into functional categories that reflect their position in the causal chain from environmental signal to physical cellular outcome. The functional classification is not merely a classification convenience; it generates specific predictions about which molecular relationships should be robust and which should be vulnerable to disruption, predictions that the computational evidence in Part 6 directly tests.

The environmental and metabolic inputs include the niche-derived signals (cytokines, Wnt ligands, Notch ligands, extracellular matrix composition), the metabolic parameters that gate chromatin chemistry (α -ketoglutarate availability for TET and Jumonji enzymes, NAD⁺ for sirtuin-mediated deacetylation, iron-sulfur cluster assembly for the Fe-S-dependent enzymes (Maio & Rouault, 2015) of the write/erase matrix), and the physicochemical parameters that affect DNA secondary structure (intracellular pH for i-motif stability (Zeraati et al., 2018), redox state for G-quadruplex oxidation sensitivity). These determine which chromatin addresses are accessible to modification and whether the enzymatic machinery has the cofactors required to operate.

The chromatin state itself (reviewed in Allis & Jenuwein, 2016) comprises DNA methylation at CpG dinucleotides, the oxidized intermediates of active demethylation (5hmC, 5fC, 5caC), histone modifications at specific residues (H3K4me1/2/3, H3K27me3, H3K9me3, H4K20me1/2/3), and the non-B DNA secondary structures (G-quadruplexes and i-motifs at

specific regulatory regions) together define the cell's epigenetic configuration. This is the information layer: it encodes the cell's identity and its current position in the A-R-C trajectory.

The regulatory coordination components include several sub-networks. The polarity proteins — exemplified by NUMB, PRKCI, PRKCZ, DLG1, PARD3, and SCRIB (reviewed in Knoblich, 2010; Gönczy, 2008) — govern the cortical asymmetry that directs fate-determinant segregation. The BER coordination network couples TET2 to TDG, OGG1 to MUTYH, and NEIL1 to XRCC1 (Beard et al., 2019). The write/erase coupling pairs DNMT1 with TET2 and EZH2 with KDM6A (Tahiliani et al., 2009; Cao et al., 2002). CDK6, beyond its canonical role in RB1 phosphorylation (Malumbres & Barbacid, 2009), has established kinase-independent transcriptional functions — it participates in transcription complexes at target promoters independently of its catalytic activity (Kollmann et al., 2013). This dual function places CDK6's regulatory relationships in a different functional category from the structural kinase-substrate coupling of CDK4 with cyclin D1. Several regulatory coordination members — TET2, EZH2, KDM6A — are themselves chromatin-modifying enzymes. Their classification reflects not the absence of catalytic activity but the regulatory coordination relationships between them that determine where and when that catalytic activity is deployed. This layer determines how the available chromatin-modifying activities are coordinated across genomic addresses and how their products are partitioned between daughters at mitosis.

The fork-protection complex (CLSPN with PCNA, ATR, CHEK1, PARP1), the Fanconi anemia ID complex (FANCI with FANCD2), the mitotic kinase partnership (PLK1 with AURKA; Lens, Voest & Medema, 2010), and the CDK4-cyclin D1 structural coupling (CDK4 with CCND1) are execution components. FANCD2 × BRCA1 is classified as regulatory coordination rather than execution because both proteins have extensive partner-independent functions and their

interaction is damage-induced rather than constitutive. BLM × WRN are excluded from the class analysis as paralogous helicases with non-overlapping substrate specificities and no demonstrated direct heterodimeric complex. These are structurally obligate partnerships — the proteins form physical complexes or act in obligate sequence at the replication fork (Cimprich & Cortez, 2008 for ATR-CHEK1 fork protection; Ceccaldi et al., 2016 for FANCI-FANCD2 ubiquitination) or in mitosis — and their co-regulation reflects this structural necessity rather than any regulatory program. These carry out the physical work of DNA replication, repair, and chromosome segregation regardless of the cell's ARC phase.

The functional assignments are made entirely on mechanistic grounds — whether the relationship between two gene products reflects a regulatory program or a structural obligation — without reference to any correlation data from cancer or normal tissue. The criteria for distinguishing regulatory from execution partnerships (the C1–C5 framework evaluated in §6.2 and scored in Table S3) are derived from established molecular biology — physical binding evidence, compartment co-localization, partner-independent functions — and can be independently verified by any reviewer. The prediction that cancer should selectively disrupt regulatory coordination while preserving execution follows from the model's architecture and is tested in Part 6. The four criteria that operationalize this prediction are stated in §5.5 and evaluated systematically in §6.2–6.3; readers seeking the empirical result before the full mechanism is developed may consult those sections. The conceptual structure presented here is independent of that empirical evaluation. Because pair classification and the correlation analyses were developed iteratively during the model's construction, the class-separation result should be treated as internally validated rather than prospectively blinded. The independent external-pair replication and the C1–C5 criteria's verifiability by any reviewer partially address this limitation.

2.3 Orthogonality with the Cell Cycle

A distinction that is essential for interpreting all of the model's predictions concerns the relationship between ARC phases and cell cycle phases. The two classification systems are orthogonal: a cell's position in the cell cycle (G1, S, G2, M) is independent of its ARC phase (A, R, C). A cell in S-phase can be in ARC-A (preparing for the asymmetric division that will initiate fate divergence), ARC-R (replicating while the write/erase program remodels its chromatin at commitment-relevant addresses), or ARC-C (undergoing its final replication before terminal commitment). The cell's ARC phase is determined by its chromatin configuration and partition status, not by where its replication machinery stands. Cell cycle markers — PCNA, MKI67, the cyclins, the CDKs — are shared infrastructure used by all three ARC phases. A cell that is MKI67-positive may be in any ARC phase; a cell that is MKI67-negative is either quiescent (outside ARC entirely) or terminally committed (post-C-phase, also outside ARC). The ARC model applies exclusively to dividing cells that are progressing through the fate-decision process. The terms A-phase, R-phase, and C-phase refer to architecture-defined states of this process, not to temporal subphases of a single cell cycle.

2.4 Walkthrough: Two Tissues

The three phases can be illustrated through two well-characterized systems that use different molecular gates for the same architectural transitions.

In the hematopoietic hierarchy, the long-term HSC resides quiescent in the endosteal niche with its partition system intact and its lineage-commitment promoters bivalently marked (Bernstein et al., 2006; Orkin & Zon, 2008). Niche-derived signals activate it into A-phase, where NUMB-aPKC polarization (Smith et al., 2007) produces an asymmetric division. The commitment-biased daughter enters R-phase, where TET2 oxidizes 5-methylcytosine at lineage promoters and

KDM6A removes repressive H3K27me3 — the multipotent progenitor compartment. If niche signals withdraw before irreversible changes accumulate, DNMT1 maintenance and histone writers can restore the pre-commitment state. The R-to-C transition occurs when the resolution architecture (§3.4) produces compound irreversible changes at sufficient commitment-relevant addresses. The endpoint — erythrocyte, neutrophil, monocyte, B lymphocyte — depends on which regulatory addresses were remodeled during R-phase, which depends on which niche signals and transcription factors were active. Terminal differentiation and programmed cell death are both C-phase outcomes on the same commitment axis. The model predicts that this trajectory should be visible as a progressive, selective change in coordination profiles — a prediction tested with single-cell resolution in §6.7.

In the intestinal epithelium, the same architecture uses a different gate. The Lgr5-positive crypt stem cell is maintained in a self-renewing state by Wnt ligands from adjacent Paneth cells (Sato et al., 2011). APC — functioning both as a Wnt pathway regulator and as a spindle-orientation hub (Quyn et al., 2010) — executes the A-phase division, positioning daughters at different positions along the crypt-villus axis where they encounter different niche signals. The hematopoietic R-phase gate is metabolic (α -ketoglutarate availability for TET enzymes); the intestinal R-phase gate is positional (APC-dependent crypt-villus placement). The architecture is the same — quiescence $\rightarrow$ A $\rightarrow$ R $\rightarrow$ C — but the molecular components that implement each gate differ between tissues. This tissue specificity is developed in Part 5, where APC loss is interpreted as the canonical partition-failure mutation in colorectal cancer.

Part 3: Molecular Implementation

The molecular implementation described in this Part uses specific, well-characterized proteins as exemplars of broader functional categories. NUMB is the exemplar cortical fate determinant because it has the most extensive published validation — direct Notch inhibition, p53 stabilization through Mdm2 binding, and a defined molecular chain linking cortical polarity to mitochondrial dynamics through MFN1-PKC ζ . The model's claims are about the functional categories, not the specific proteins. In tissues where NUMB is not the primary cortical fate determinant, the model predicts that another member of the same category serves the equivalent role. The established biochemical function, model-specific interpretation, and per-gene falsification test for every protein in the panel are detailed in Supplementary Table S1.

3.1 Upstream Signal Integration: From Niche to Genomic Address

The ARC model requires that niche-derived signals be converted into address-specific chromatin modifications at defined genomic loci. The niche signal — a cytokine, a cell-cell contact, a change in extracellular matrix composition, or a metabolic shift — must ultimately result in the activation of specific enzymes at specific promoters. The mechanism by which this conversion occurs operates through three physical parameters that link the extracellular environment to intracellular DNA chemistry: pH, redox state, and metal ion availability.

Intracellular pH governs the stability of non-B DNA secondary structures at regulatory regions. I-motif structures — tetrahelical DNA structures formed by C-rich sequences at promoter regions whose physiological existence has been demonstrated (Zeraati et al., 2018) and whose thermal stability is modification-dependent (Wright et al., 2020; Gehring et al., 1993) — are one example of pH-responsive genome regulatory elements. The model treats i-motif chemistry as a possible structural element of commitment irreversibility at specific promoters (§3.4) while

identifying the primary commitment mechanism as the resolution architecture at replication-problematic addresses.

The cellular redox state governs the chemistry of the write/erase program through multiple mechanisms. The TET enzymes that oxidize 5-methylcytosine require molecular oxygen and α -ketoglutarate (Iyer et al., 2009), both of which are sensitive to the cell's metabolic state. The iron center in TET2 cycles between Fe^{2+} (active) and Fe^{3+} (inactive) depending on cellular reducing capacity, and the cofactor ascorbate (vitamin C) directly enhances TET activity by maintaining the iron center in its reduced active state (Blaschke et al., 2013). The OGG1 glycosylase that excises 8-oxoguanine from G-quadruplex structures is itself redox-sensitive, with a critical cysteine residue whose oxidation state modulates substrate binding. The niche's oxygen tension, the mitochondrial electron transport chain's contribution to cytoplasmic redox buffering, and the availability of reducing equivalents (NADPH, glutathione) all feed into the redox parameter that governs which write/erase reactions can proceed. A stem cell in a hypoxic niche (the endosteal bone marrow niche, the intestinal crypt base) has a different redox environment from a transit-amplifying cell that has migrated away from the niche, and the redox difference directly affects which chromatin-modifying reactions are thermodynamically and kinetically accessible.

Metal ion availability — particularly iron, zinc, and magnesium — gates the enzymatic machinery of the write/erase program at the cofactor level. The iron-sulfur clusters required by the FANCD1 helicase, the RTEL1 helicase (White, 2009), and the DNA glycosylase NTHL1 (Aspinwall et al., 1997) must be assembled by the mitochondrial iron-sulfur cluster biogenesis pathway (ISCU, NFS1, LYRM4) before these enzymes can function. Tissue-specific differences in iron-sulfur cluster assembly gene expression are consistent with the tissue-specific temporal strategies described in §5.4. The model treats metal ion availability as an environmental gating

parameter: if the cofactor is not present, the enzyme cannot function, regardless of how strongly the regulatory coordination program is calling for its activity.

The three parameters — pH, redox, metal ions — converge at the DNA to determine which addresses are accessible to modification and which enzymes are competent to modify them. The niche signal does not specify "activate gene X" through a transcription-factor cascade. Instead, it alters the physical and chemical environment in which the write/erase program operates, and the program's output — which addresses are modified, in which direction, to what extent — is determined by the interaction between the environmental parameters and the sequence-dependent structural properties of each genomic address. In this framework, the same set of enzymes, operating in different niche contexts, produces different lineage outcomes: the enzymes are the same, but the addresses they can reach and the chemistry they can perform depend on the environment. The fidelity-control architecture developed in §3.4 — in which the model proposes that RecQ-family helicases provide a proofreading function at TLS bypass events — influences whether the chemistry installs permanent sequence change at any given address or restores the original base (Figure 2).

Figure 2. Molecular Implementation of ARC: Niche Gating, Dual Partition, and the Write/Erase Matrix

Niche-derived physicochemical signals specify address-level chromatin chemistry, while coupled partition systems deliver matched nuclear and organelle states to each daughter.

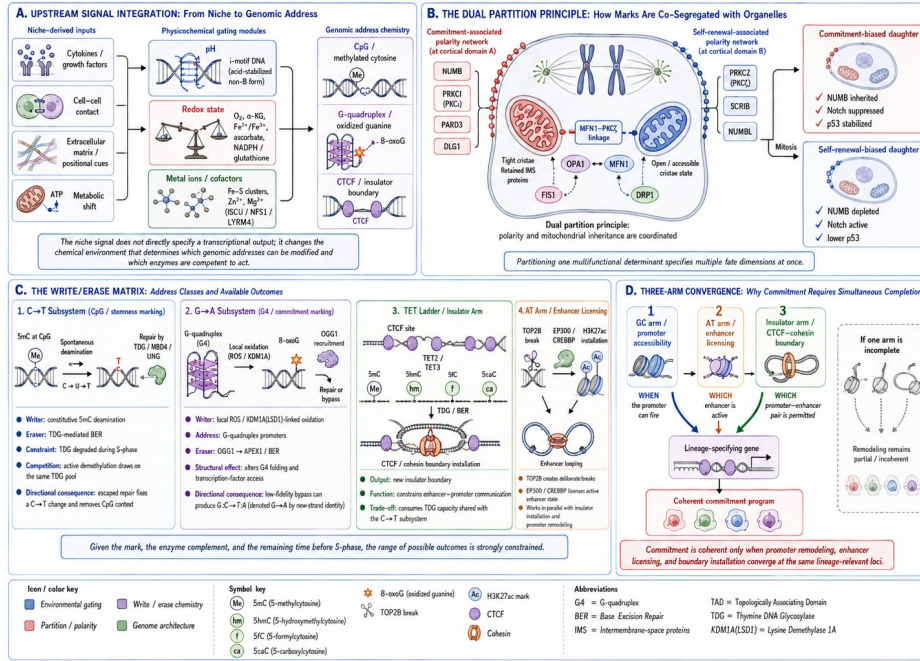


Figure 2. Molecular implementation of ARC: niche gating, dual partition, and the write/erase matrix. Niche-derived physicochemical signals specify address-level chromatin chemistry, while coupled partition systems deliver matched nuclear and organelle states to each daughter. (A) Upstream signal integration. Niche inputs (cytokines, cell-cell contact, extracellular matrix, metabolic shift) act through physicochemical gating modules — intracellular pH governing i-motif DNA stability (Zeraati et al., 2018), redox state governing G-quadruplex oxidation sensitivity (Perillo et al., 2008), and metal ion/cofactor availability governing Fe-S-dependent enzyme competence (Maio & Rouault, 2015) — to determine which genomic addresses (CpG, G4/8-oxoG, CTCF boundaries) are accessible to modification. (B) The dual partition principle. Cortical polarity proteins are organized into commitment-associated (NUMB, PRKCI, PARD3, DLG1) and self-renewal-associated (PRKCZ, SCRIB, NUMBL) networks (§3.2.1). The MFN1-PKCζ molecular chain (Wu et al., 2019) connects cortical polarity to mitochondrial outer-membrane fusion, which is prerequisite for OPA1-mediated inner-membrane fusion and cristae junction tightening. The commitment-biased daughter inherits NUMB (Notch suppressed, p53 stabilized) with tightly cristae'd mitochondria; the self-renewal-biased daughter inherits the reciprocal configuration. FIS1, which the model predicts opposes OPA1-mediated cristae junction tightening (§3.2.2), is shown on the outer mitochondrial membrane where it resides constitutively. (C) The write/erase matrix. Four address-modification subsystems operate during R-phase: (1) C → T subsystem at CpG addresses (5mC deamination, TDG/MBD4/UNG erasure; §3.3.1); (2) G → A subsystem at G4 addresses (KDM1A/LSD1-generated ROS, OGG1/APEX1 erasure; §3.3.2); (3) TET ladder/insulator arm (5mC → 5hmC → 5fC → 5caC oxidation, CTCF/cohesin boundary installation; §3.3.4); (4) AT arm/enhancer licensing (TOP2B deliberate breaks, EP300/CREBBP-mediated H3K27ac; §3.3.6). (D) Three-arm convergence. Commitment is coherent only when promoter remodeling (GC arm), enhancer licensing (AT arm), and insulator boundary installation (CTCF-cohesin arm) converge simultaneously at lineage-specifying genes (§3.3.7). If any arm is incomplete, remodeling remains partial or incoherent.

3.2 The Two Partition Systems: How Marks Are Co-Segregated with Organelles

The upstream signal (§3.1) directs where to mark. The write/erase matrix (§3.3) constrains each mark's outcome. But neither is sufficient to produce divergent fates unless the marked chromatin is delivered to the correct daughter together with the appropriate enzymatic machinery to process it. Mitochondria with different ages and cristae configurations are asymmetrically inherited at division (Katajisto et al., 2015), and this asymmetry causally determines daughter cell fate (Döhla et al., 2022; Bunel et al., 2025). A daughter receiving commitment-marked chromatin but stem-configured organelles (fused mitochondria, tight cristae, retained IMS proteins) would have the wrong enzymes for its chromatin state. A daughter receiving stem-marked chromatin but commitment-configured organelles would face the same incoherence. The biological problem is co-segregation: the cell must ensure that its nuclear program and its cytoplasmic program are matched in each daughter.

The model proposes that this co-segregation is achieved by two coupled partition systems — one at the cell cortex (the NUMB/aPKC/PAR system reviewed in Knoblich, 2010) and one at the inner mitochondrial membrane (the OPA1/MFN/DNM1L (DRP1) system whose asymmetric inheritance was demonstrated by Katajisto et al., 2015, and whose fate-determining function was shown by Döhla et al., 2022) — that are connected through a molecular chain linking outer membrane dynamics to inner membrane architecture.

3.2.1 The Commitment Polarity Network

At the cell cortex, asymmetric partition is organized by a distributed network of polarity proteins rather than by a single fate determinant. Classical studies in *Drosophila* neuroblasts identified the CDC42-PAR6-aPKC complex as the primary organizer of cortical asymmetry, with aPKC phosphorylation excluding fate determinants — NUMB, MIRANDA, PROSPERO, BRAT —

from the aPKC-enriched cortex (Knoblich, 2010; Gönczy, 2008; Betschinger & Knoblich, 2004). In mammalian cells, the orthologous machinery is elaborated into a larger network.

The model identifies at least ten commitment-associated proteins (including PRKCI, DLG1, NUMB, and PARD3, among others) and at least five self-renewal-associated proteins (including SCRIB, PRKCZ, and NUMBL, among others) based on their correlation profiles with the model's molecular axes. Several features of this network are notable. The two aPKC isoforms have opposite model alignments — PRKCZ is self-renewal-associated while PRKCI is commitment-associated — suggesting that the aPKC isoform ratio may itself function as a partition readout. The two NUMB paralogs similarly split: NUMB is commitment-associated while NUMBL is self-renewal-associated.

NUMB is presented as the exemplar partition protein because it has the most extensive published validation: direct Notch inhibition, p53 stabilization through Mdm2 binding (Colaluca et al., 2008), and a defined molecular chain linking mitochondrial dynamics to cortical fate determinant segregation through MFN1-PKC ζ (Wu et al., 2019). But the partition function is distributed across the network. Loss of any single member may be partially compensated by others — a design feature that makes the system robust to individual gene loss but vulnerable to network-level disruption — a vulnerability whose consequences the model develops in Part 5.

NUMB's dual molecular function is central to the model's account of how a single partition event can specify multiple fate dimensions simultaneously. Through its PTB domain, NUMB binds the Notch intracellular domain and promotes its ubiquitin-dependent degradation, suppressing Notch-mediated transcriptional programs in the daughter that inherits NUMB. The molecular mechanism — NUMB promoting Notch1 receptor ubiquitination and degradation of the

intracellular domain — was established in mammalian cells by McGill & McGlade (2003); the broader cell-fate-determinant context is reviewed in Knoblich (2010). Through a separate interaction surface, NUMB binds Mdm2 and prevents Mdm2-mediated ubiquitination and degradation of p53, thereby stabilizing p53 in the same daughter (Colaluca et al., 2008). The consequence is that asymmetric partition of a single protein — NUMB — simultaneously specifies the Notch axis and the p53 axis for both daughters: the NUMB-inheriting daughter has suppressed Notch and stabilized p53 (a commitment-compatible configuration), while the NUMB-depleted daughter has active Notch and lower p53 activity (a self-renewal-compatible configuration). The model's claim is that this multifunctionality is not a coincidence but a design principle. The partition system uses multifunctional proteins precisely because partitioning one molecule specifies multiple fate dimensions at once, producing the coherent daughter-cell identities that single-axis models cannot explain.

The model interprets the aPKC isoform switch as providing a mechanistic basis for directional control within the partition system. Atypical protein kinase C (aPKC) phosphorylates mammalian NUMB at conserved sites within its PTB domain, and this phosphorylation regulates NUMB's polarized membrane localization and asymmetric segregation at mitosis (Smith et al., 2007). The model's correlation analysis reveals that the two atypical PKC isoforms — PRKCZ and PRKCI — have opposite alignments with the maintenance-commitment axis: PRKCZ is self-renewal-associated (co-expressed with the genome-maintenance enzyme battery) while PRKCI is commitment-associated (co-expressed with the mark-installation arms). The model's prediction, building on the established aPKC–NUMB phosphorylation mechanism, is that the isoform ratio functions as a molecular switch: in a cell where PRKCZ activity predominates, NUMB is phosphorylated into a maintenance-compatible configuration (cytoplasmic, non-

polarized, distributed equally to both daughters at mitosis). In a cell where PRKCI activity predominates, NUMB is phosphorylated into a commitment-compatible configuration (cortical, polarized, asymmetrically distributed). The A-to-R transition described in Part 2 corresponds, at the molecular level within the partition system, to a shift in the PRKCZ-to-PRKCI activity ratio. The computational evidence in Part 6 tests this interpretation directly: if the aPKC isoform switch is real, then $\text{NUMB} \times \text{PRKCZ}$ and $\text{NUMB} \times \text{PRKCI}$ should show different coupling behaviors, and cancer — which disrupts the partition system — should differentially affect the two relationships.

3.2.2 OPA1 and the Dual Partition Principle

At the inner mitochondrial membrane, a second partition operates through cristae junction control. OPA1, a dynamin-related GTPase, oligomerizes at cristae junctions to control their width (Delettre et al., 2000; Frezza et al., 2006). When OPA1 oligomers are intact, cristae junctions are tight, trapping intermembrane space proteins within the cristae lumen and supporting efficient respiratory chain supercomplex assembly (Cogliati et al., 2013). When OPA1 oligomers disassemble, cristae contents become accessible for release through BAX/BAK pores in the outer membrane, and respiratory efficiency declines. The OPA1 partition therefore controls two things simultaneously: which IMS proteins are available to each daughter, and what metabolic capacity each daughter's mitochondria possess. Critically, this inner-membrane gate operates independently of BAX/BAK pore formation at the outer membrane — Yamaguchi et al. (2008) showed that BAX activation and MOMP can proceed while cytochrome c remains trapped within OPA1-protected cristae, establishing that the two membrane events are separable. IMS proteins differ in size and complex state, so cristae junction width determines which

proteins are released and which are retained, creating combinatorial differences between daughters depending on OPA1 status (Scorrano & Korsmeyer, 2003; Barbot et al., 2015).

The two partition systems — cortical polarity and cristae architecture — are connected through a single molecular chain. The cortical polarity network that distributes NUMB asymmetrically at mitosis simultaneously regulates mitochondrial outer-membrane fusion through MFN1 (Wu et al., 2019). MFN1-mediated outer-membrane fusion is the prerequisite for OPA1-mediated inner-membrane fusion, which tightens cristae junctions. The chain runs from cortical polarity through outer-membrane fusion to inner-membrane cristae state, producing what the model terms the dual partition principle: nuclear chromatin marks and cytoplasmic organelle configurations are co-partitioned at mitosis as a single coordinated event, not as two independent sorting processes. The daughter that receives commitment-marked chromatin also receives tightly cristae'd mitochondria with retained IMS proteins and high oxidative capacity. Cancer, which disrupts cortical polarity, is predicted to produce both chromatin incoherence and metabolic incoherence simultaneously.

Katajisto et al. (2015) demonstrated that old and young mitochondria are asymmetrically apportioned between stem cell daughters, with the daughter fated for differentiation preferentially receiving old mitochondria. Döhla et al. (2022) provided functional evidence that this asymmetry is causally upstream of fate divergence: through a RISP-dependent mechanism, the age of inherited mitochondria directly determines daughter cell fate. These findings are consistent with the co-partition principle — the asymmetric apportionment of mitochondria with different cristae architectures is the cytoplasmic half of the same partition that distributes chromatin marks. A distinction critical to the model separates cristae junction remodeling from organelle fission: FIS1 is constitutively anchored on the outer mitochondrial membrane and

recruits DRP1 for fission; the model predicts that FIS1 also opposes OPA1-mediated cristae junction tightening, making IMS proteins accessible, while DNM1L (DRP1) mediates physical organelle division without altering cristae architecture (Smirnova et al., 2001). The model predicts that FIS1 and DRP1 should show opposite relationships with the commitment axis, tested in Part 6.

3.3 The Write/Erase Matrix: How Each Mark Determines Available Outcomes

Three chemical subsystems modify DNA at regulatory addresses during R-phase: the C → T pathway, the G → A pathway, and the TET ladder. Two additional subsystems — the AT arm and the insulator arm — coordinate these modifications across the three-dimensional genome.

Before describing the matrix's three subsystems, a conceptual reframing is necessary. The chemistry described below involves deliberate oxidation, excision, and base replacement at specific genomic addresses — events conventionally classified as DNA damage. The model does not dispute the chemistry. It disputes the classification. The field already accepts that TET-mediated active demethylation — oxidation of 5-methylcytosine followed by TDG-mediated excision and base excision repair — is a deliberate, programmed process essential for development (Tahiliani et al., 2009; He et al., 2011; reviewed in Kohli & Zhang, 2013). This process deliberately creates abasic sites, single-strand breaks, and base-excision intermediates that, in any other context, would constitute damage. The distinction between damage and signal is not the chemistry but the context: whether the modification occurs at a random position or at a specific address within a regulated program. The endogenous rate of DNA lesion generation — approximately 40,000 per cell per day (Lindahl, 1993) — is conventionally attributed to unavoidable chemical decay, but a substantial fraction may represent the normal operation of the write/erase program in cells undergoing chromatin remodeling.

With the upstream signal (§3.1) directing where marks are installed and the partition systems (§3.2) delivering the marked chromatin to each daughter, the question becomes: once a mark is installed at a particular address, what happens next? In conventional models, the answer is treated as the output of a regulatory computation — transcription factors arrive, signaling cascades fire, and a context-dependent decision emerges. The model offers a more constrained answer. Each modification has a defined enzymatic clientele, a defined temporal window in which that clientele can act, and a defined fate when the window closes. The combination of mark identity, available enzymes, and remaining time before S-phase determines the outcome at every modified address. The model proposes that, given the inputs, the outputs are strongly constrained — not random, and not requiring additional regulatory computation to resolve.

Three subsystems are described below of the matrix that have been characterized in sufficient mechanistic detail to make specific predictions. Each subsystem has a writing reaction, an erasure pathway, a temporal constraint imposed by replication, a competition that biases the outcome, and a strand-specific consequence for what each daughter inherits. Other subsystems certainly exist; these three are presented because their participants and constraints are established well enough that the matrix's behavior at G4-containing and CpG-containing addresses can be tracked. Their integration into a single input-determined system is the central mechanistic claim of this manuscript.

3.3.1 The C → T Subsystem: Stemness Marking on the Template Strand

The cytosine subsystem is the most chemically simple. Hydrolytic deamination of 5-methylcytosine to thymine is a constitutive reaction that proceeds at approximately one hundred to five hundred events per cell per day across the genome, driven by water and temperature (Lindahl & Nyberg, 1974) rather than by any enzyme. The cell cannot regulate the deamination

rate directly; it can only modulate the substrate pool through the maintenance methyltransferase DNMT1 (Li et al., 1992) and the de novo methyltransferases DNMT3A and DNMT3B (Okano et al., 1999). More methylation produces more 5-methylcytosine, which produces more deamination, which produces more C → T transitions on the template strand at methylated CpG sites. Because methylation in mammalian genomes is concentrated at regulatory elements — promoters, enhancers, CpG islands (Deaton & Bird, 2011) — the deamination events are similarly concentrated at addresses that control transcription factor access, not distributed at random.

Erasure runs through thymine DNA glycosylase (TDG). At a G:T mismatch generated by deamination at a methylated CpG, TDG recognizes the unpaired thymine, excises it through a base-flipping mechanism (Maiti & Drohat, 2011) that distinguishes the mismatch from the surrounding Watson-Crick base pairs, and initiates the base excision repair pathway that restores the original cytosine. The methyl-CpG-binding domain protein MBD4 provides a redundant recognition pathway at methylated sites, while uracil DNA glycosylase (UNG) handles the analogous chemistry at unmethylated sites (Krokan & Bjørås, 2013) where C → U deamination has occurred. Together, these enzymes form a layered system that, when fully resourced, can keep up with the constitutive deamination rate.

The temporal constraint on this erasure is stringent. TDG is degraded via ubiquitination during S-phase (Hardeland et al., 2007), creating a narrow window in which the G:T mismatch can be recognized and repaired. Once replication generates a daughter strand carrying adenine opposite the deamination-derived thymine, the resulting T:A pair is a normal Watson-Crick base pair, indistinguishable from the millions of T:A pairs that occur naturally in the genome. No glycosylase can recognize it as damage. The biological significance of TDG-mediated active

demethylation is underscored by its embryonic lethality: TDG knockout mice die at E12.5 with defective gene regulation at developmental promoters (Cortellino et al., 2011), establishing TDG as essential for active demethylation during lineage commitment. The model emphasizes that the S-phase degradation window is not merely a passive enzymatic deadline — the cell's capacity to delay or forgo repair of a recognized lesion is itself a regulatory decision, and the decision not to repair at a specific address is what converts a reversible chemical modification into a permanent sequence change.

The competition that gives the C → T subsystem its directional consequence arises from TDG's dual role. The same enzyme that excises thymine from G:T mismatches also excises 5-formylcytosine and 5-carboxylcytosine produced by TET-mediated active demethylation (Ito et al., 2011) and excises both modified bases through the same base-excision repair mechanism (He et al., 2011). These substrates draw on the same active site. When TET activity is high — during active demethylation at lineage-specifying loci, driven by α -ketoglutarate availability from the metabolic environment — TDG is diverted toward 5fC and 5caC excision, leaving reduced capacity for G:T repair at the same regulatory regions where deamination is concentrated. The unrepaired thymine persists into S-phase and becomes permanent. This produces a counterintuitive feedback: precisely those signals that drive lineage-specifying demethylation also accelerate the accumulation of irreversible C → T transitions at the same addresses. The cell pays a deamination tax for each round of active demethylation it performs.

The result, traced across A and R phases, is strand-specific and directional. The deamination event occurs on the template strand, where the original 5-methylcytosine resided. After replication, the chromatid built from the deaminated template strand carries a T:A pair where a G:5mC pair had been — a permanent sequence change — and the loss of the CpG context means

DNMT1 cannot re-establish methylation at that position. The chromatid built from the complementary strand retains the original G:C sequence and the intact CpG context for normal maintenance methylation (the hemimethylated CpG recognition mechanism: Song et al., 2012). The two daughters thus differ at every successfully escaped deamination event: one has lost a methylation mark and gained a permanent sequence change; the other has lost the methylation mark and retained the original sequence on a new, unmodified strand. The accumulation of these events biases the daughter inheriting the template-rich chromatid toward a chromatin configuration progressively enriched in stemness-associated changes, while its sister begins each cycle with a more open, unmodified configuration. This is the molecular substrate of the asymmetric inheritance that A-phase divisions produce.

3.3.2 The G → A Subsystem: Commitment Marking at G-Quadruplex Addresses

The guanine subsystem is more spatially constrained. While cytosine deamination occurs throughout the genome, oxidative modification of guanine is concentrated at specific locations (Cadet & Wagner, 2013). The lysine-specific demethylase LSD1/KDM1A generates hydrogen peroxide as a byproduct of histone H3K9 demethylation, and the local hydrogen peroxide oxidizes nearby guanines to 8-oxoguanine. Perillo et al. (2008) demonstrated this sequence at estrogen-responsive promoters in breast cancer cells. Estrogen receptor-directed activation of LSD1 produces hydrogen peroxide at hormone-responsive promoters, the resulting 8-oxoG recruits 8-oxoguanine glycosylase (OGG1), and the entire sequence — receptor binding, LSD1 activation, local H₂O₂, 8-oxoG, OGG1 recruitment, abasic site processing — is required for transcription of the target genes. The oxidative event is the signal, not noise.

The spatial concentration arises because G-quadruplex structures, formed by guanine-rich sequences capable of folding into stacked guanine tetrads, are preferential substrates for both

reactive oxygen species and for the demethylases that generate them. G-quadruplexes are enriched at gene promoters, telomeres, immunoglobulin switch regions (Huppert & Balasubramanian, 2005), and other regulatory elements. The same guanine-rich sequences that fold into G4 structures are oxidized first when local hydrogen peroxide is produced. The combination — the writing enzyme (LSD1) and the structural substrate (G4) co-localized at regulatory addresses — concentrates the G → A subsystem at the loci that matter for transcriptional regulation.

Erasure follows the base excision repair pathway. OGG1 excises 8-oxoG and generates an abasic site. Apurinic/apyrimidinic endonuclease 1 (APE1/Ref-1), which functions both as a BER endonuclease and as a redox factor that regulates the DNA-binding activity of transcription factors including AP-1, NF-κB, and HIF-1α (Tell et al., 2009), processes the abasic site (Whitaker & Freudenthal, 2018), generating a single-strand break with 3'-OH and 5'-deoxyribose-phosphate ends. DNA polymerase β fills the resulting gap by inserting a normal guanine, and DNA ligase III seals the nick. If this pathway runs to completion before the replication fork arrives, the original guanine is restored and no permanent change occurs. If it does not, the consequences depend on which intermediate persists. A persistent 8-oxoG is bypassed by replicative polymerases that misinsert adenine opposite the oxidized guanine in approximately fifty percent of encounters (Markkanen, 2017); after a second round of replication, one daughter strand carries a permanent G:C → T:A transversion (denoted throughout this manuscript as "G → A" by new-strand base identity — the adenine that was misinserted on the new strand persists as the permanent sequence record at that position; see Figure 3 legend). A persistent abasic site causes more severe disruption: replicative polymerases stall, translesion synthesis polymerases are recruited, and the outcome is determined by which TLS polymerase

the cell recruits — a regulated event governed by PCNA ubiquitination state and chromatin context (the resolution architecture developed in §3.4). If neither helicase nor TLS resolves the obstacle before mitosis, the underreplicated DNA is transmitted to the daughter cell, where it is sequestered in 53BP1 nuclear bodies and resolved through a distinct post-mitotic pathway (§3.4).

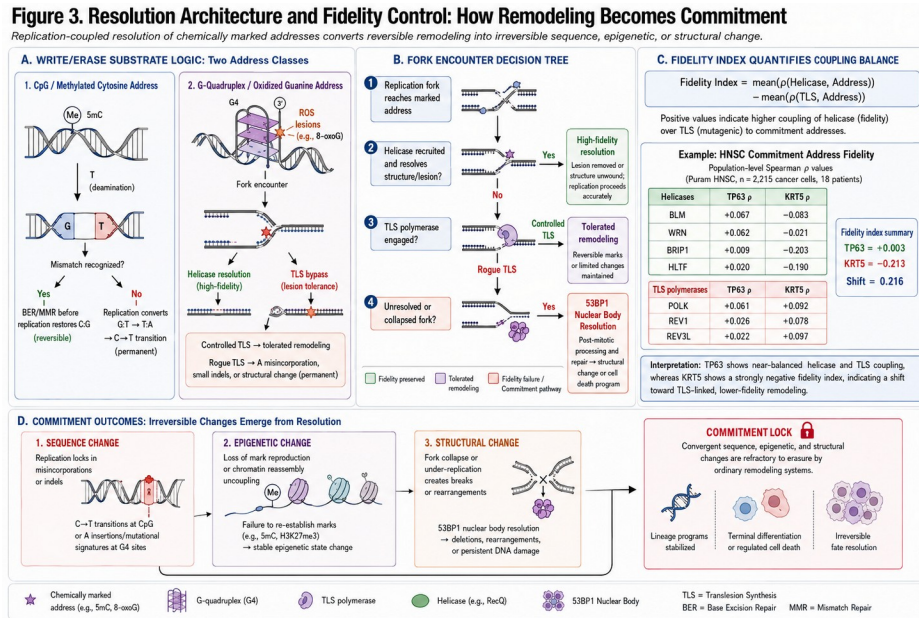


Figure 3. Resolution architecture and fidelity control: how remodeling becomes commitment. Replication-coupled resolution of chemically marked addresses converts reversible remodeling into irreversible sequence, epigenetic, or structural change. (A) Write/erase substrate logic at two address classes: CpG/methylated cytosine addresses (hydrolytic deamination of 5-methylcytosine produces C → T, SBS1; §3.3.1) and G-quadruplex/oxidized guanine addresses (8-oxoG bypass produces G:C → T:A, SBS18; §3.3.2; denoted G → A by new-strand base identity). (B) Fork encounter decision tree at each marked address: (1) fork reaches address; (2) helicase recruited (BLM, WRN, BRIP1) → high-fidelity resolution; (3) TLS polymerase engaged (REV1, POLK, REV3L) → tolerated remodeling or rogue bypass; (4) unresolved fork → 53BP1 nuclear body resolution. HLTF serves dual function as dsDNA translocase and PCNA polyubiquitination E3 ligase. (C) Fidelity index quantifies coupling balance. Fidelity index = $\text{mean}(\text{helicase } \rho) - \text{mean}(\text{TLS } \rho)$, computed from population-level Spearman correlations (Puram et al., 2017; n = 2,215 HNSC cancer cells, 18 patients; GSE103322). TP63+ (stem program): fidelity index = +0.003; KRT5+ (failed commit): fidelity index = -0.213; shift = 0.216. Individual gene values tabulated. (D) Commitment outcomes: convergent sequence change (C → T at CpG, G:C → T:A at G4), epigenetic change (failure to re-establish marks), and structural change (fork collapse, 53BP1 NB resolution) are collectively refractory to ordinary remodeling systems, producing stable terminal fate or regulated cell death.

The G4 architecture constrains both the writing and the erasure in ways that distinguish this subsystem from the C → T subsystem. OGG1's catalytic activity at folded G-quadruplexes is reduced relative to its activity on canonical duplex DNA (Fleming et al., 2017), meaning that 8-oxoG within an intact G4 is partially protected from erasure. The G4 must be unfolded — by the G4-resolving helicases described in §3.1 or by the local single-stranded DNA generated during transcription — before OGG1 can fully access its substrate. This adds a second temporal layer to the subsystem: the erasure window is constrained not only by the S-phase deadline but also by the kinetics of G4 unfolding at each address. Addresses where G4 unfolding is slow accumulate 8-oxoG and progress toward G → A transition; addresses where G4 unfolding is fast permit complete erasure.

A third asymmetry distinguishes the G → A subsystem from its cytosine counterpart. The adenine misincorporated opposite an unrepaired 8-oxoG can, in principle, be excised by MutY DNA glycosylase (MUTYH) — but the model predicts low MUTYH activity in stem and progenitor compartments, reflecting the framework's logic that the G → A commitment-marking subsystem requires permissive (unrepaired) 8-oxoG processing for the bypass-mediated installation step. Systematic characterization of MUTYH protein levels and activity across stem cell populations has not been published, so this prediction remains testable rather than confirmed. The asymmetry between the two subsystems is therefore directional: TDG-mediated repair of C → T transitions on the template strand operates with the cell's full enzymatic complement, while MUTYH-mediated repair of A misincorporated opposite 8-oxoG on the new strand is predicted by the model to be poorly resourced. C → T events accumulate on the template strand because erasure becomes invisible; G → A events accumulate on the new strand because erasure is predicted to be mechanistically absent. The two subsystems write different chemistries on different strands at

different addresses, and the model proposes that the daughters inherit the consequences differentially.

The functional consequence at G4-containing promoters is structural rather than purely informational. A $G \rightarrow A$ transition at a guanine that participates in a G-tetrad alters the geometry of that tetrad (the chemical basis is reviewed in Fleming & Burrows, 2017). Depending on which guanine is mutated and which alternative folding topology the sequence can adopt, the consequence ranges from mild (a slightly destabilized G4 that still folds) to severe (a G4 that can no longer fold at all, leaving the sequence in extended single-stranded or duplex form). G4 unfolding alters transcription factor accessibility, changes the binding of G4-resolving helicases, and reorganizes the local chromatin state. The progressive accumulation of $G \rightarrow A$ events at G4 addresses thus closes regulatory options progressively: each such transition represents a structural decision that cannot be reversed by demethylation, by histone modification, or by any other epigenetic intervention. The G4 either folds or it does not, and the sequence determines which.

The mutagenic vulnerability of G4 addresses is reinforced by a second, complementary mechanism. In vitro forward mutation assays demonstrate that G4 motifs from the human genome intrinsically reduce DNA polymerase fidelity, with polymerase-specific error signatures observed for pol δ , η , and κ during synthesis of parallel G4 structures (Stein et al., 2022) — errors occur within, immediately flanking, and encompassing the G4 motif, with each polymerase producing a characteristic error class (deletions and complex errors for pol δ ; flanking errors for pol η ; frameshifts within G-tracts for pol κ). The write/erase chemistry described above thus operates at addresses where polymerase errors are already enriched on biochemical grounds independent of the chemistry. Both mechanisms predict mutation

accumulation at the same addresses, and the observed enrichment of cancer driver mutations at G4-containing promoters is consistent with their combined contribution. The model does not claim that the write/erase chemistry is the sole source of G4-localized mutation.

3.3.3 Transcription Factor Binding Site Consequences of Write/Erase Chemistry

The conventional classification of translesion synthesis polymerases as error-prone emergency enzymes has been challenged by evidence that these polymerases serve developmental functions beyond damage tolerance. Sale et al. (2012) reviewed the expanding roles of Y-family polymerases in somatic hypermutation, immunoglobulin gene conversion, and replication through natural structural impediments including G-quadruplexes. Shah et al. (2024) demonstrated that REV1 is required for genome-wide DNA demethylation in primordial germ cells — REV1-deficient mice show complete failure of PGC demethylation and sterility, establishing that a translesion polymerase has an essential developmental function.

The model proposes that TLS polymerase chemistry at G4 addresses creates a constrained mapping between mutation chemistry and transcription factor binding site consequences. A concrete illustration comes from the TERT promoter, where the C228T and C250T mutations — both C → T transitions within a G4-forming sequence (Horn et al., 2013) — create de novo ETS/GABPA transcription factor binding sites while simultaneously disrupting CTCF repressor binding and altering the G4 folding landscape (Bell et al., 2015). At inflammatory promoters, OGG1 bound to 8-oxoG at intact G4 structures scaffolds NF-κB and Sp1 transcription factor recruitment without requiring any base change (Ba et al., 2014; Pan et al., 2016) — the oxidative intermediate itself is the recruitment signal. The model identifies four distinct TF binding consequences that can arise from write/erase chemistry at G4-containing commitment promoters: C → T transitions creating ETS-family motifs, G → A transitions creating p53 half-sites, G → T

transversions creating YY1/PRC2 motifs, and OGG1-8-oxoG scaffolding creating transient NF- κ B recruitment. The computational analysis supporting this four-arm logic — including a 30-promoter analysis against GC-matched null models showing significant enrichment of ETS-family motifs and depletion of NF- κ B proto-sites at G4 structures, with two of four arms confirmed as G4-context-specific — is presented in Supplementary Section S2, along with the temporal implications of progressive binding-site accumulation across the lifespan.

3.3.4 The TET Demethylation Ladder: Insulator Installation at CTCF Sites

The TET demethylation ladder operates in a different functional context from the TET-mediated oxidation described in §3.3.1. TET1, TET2, and TET3 sequentially oxidize 5-methylcytosine through three intermediates: 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxylcytosine. Each oxidation step requires α -ketoglutarate, ferrous iron, and molecular oxygen. Both 5-formylcytosine and 5-carboxylcytosine are excised by TDG (He et al., 2011; Maiti & Drohat, 2011), and the resulting abasic site is repaired through base excision repair to restore unmethylated cytosine. The full ladder converts methylated cytosine to unmethylated cytosine, completing what is conventionally called active demethylation.

The functional context that distinguishes this subsystem from the others is its operation at CTCF binding motifs. CTCF is the dominant insulator-binding protein in mammalian genomes (Phillips & Corces, 2009) and the principal architect of topologically associating domain boundaries. The genome-wide three-dimensional organization established by CTCF and cohesin was mapped at kilobase resolution by Rao et al. (2014), who identified approximately 10,000 chromatin loops linking promoters and enhancers, with loop anchors typically occurring at domain boundaries bound by CTCF in convergent orientation (>90% of cases) and the resulting loop structures conserved across cell types and species. Methylation of cytosines within a CTCF motif occludes

CTCF binding; demethylation through the TET ladder restores it (Wang et al., 2012). When CTCF binds, it recruits cohesin and establishes a chromatin boundary (Wendt et al., 2008) that constrains which enhancers can access which promoters within the resulting domain. The installation of a new CTCF boundary during differentiation is therefore not merely a local epigenetic change — it can reorganize the regulatory landscape over hundreds of kilobases by partitioning the genome into different neighborhoods, the architectural scale that the Rao et al. map identifies as the canonical unit of human chromatin organization.

The TET ladder intermediates also have consequences for i-motif DNA stability at the same CpG addresses: 5hmC is approximately structurally neutral, while 5fC and 5caC can destabilize i-motif folding in a context-dependent manner (Wright et al., 2020) — though the magnitude and direction of these effects vary with sequence context, ionic conditions, and the density of modifications at a given locus, and should not be treated as uniformly destabilizing. These modification-dependent stability effects are one element of the broader commitment chemistry developed in §3.4, where strand-specific sequence changes provide the primary irreversibility mechanism. The model proposes that the number of TET rounds required to advance any given address depends on both the initial methylation density and the available α -ketoglutarate flux.

The TET ladder competes with the $C \rightarrow T$ subsystem for TDG, and this competition has a directional consequence that connects the two subsystems mechanistically. Active TET-mediated demethylation at CTCF sites consumes TDG capacity. Reduced TDG capacity slows G:T mismatch repair at deaminated CpGs elsewhere in the genome. Each round of insulator installation therefore accelerates the accumulation of $C \rightarrow T$ transitions at unrelated regulatory addresses. The cell cannot simultaneously maximize insulator installation and minimize deamination-derived sequence change. The trade-off is resolved differently in different contexts:

stem cells with low TET activity accumulate fewer C → T transitions and install fewer new insulator boundaries; transit-amplifying cells with high TET activity install many new boundaries and accumulate more C → T transitions, narrowing their regulatory repertoire faster. The connection between these subsystems means that progressive insulator installation and progressive deamination-derived stemness marking are two faces of a single chromatin-remodeling program operating on a finite enzymatic budget.

3.3.5 Matrix Integration: Input-Constrained Outcomes from Specified Inputs

The three subsystems are not independent. They share enzymes (TDG appears in all three), they share metabolic inputs (α -ketoglutarate gates the cytosine oxidation chemistry; reactive oxygen species generated by demethylase activity feeds the guanine oxidation chemistry), and they share the temporal constraint imposed by S-phase. They also share their genomic targets: G4-containing promoters carry both the cytosine modifications that determine i-motif stability and the guanine modifications that determine G4 folding, and these modifications interact at the level of local DNA structure. A promoter with both 5fC at its CpG sites and 8-oxoG within its G4 is in a different structural state from a promoter with either modification alone, and the structural state determines transcription factor accessibility independently of any signaling input.

The model's central claim about this matrix is operational rather than mechanistic. The model does not propose new biochemistry; the individual reactions are established. It proposes that these reactions, taken together, constitute a system whose behavior at every modified address is strongly constrained by three observable inputs: which marks are present at that address, which enzymes are available in that cell, and how much time remains before S-phase. Cells with different enzyme complements — different because of the partition described in §3.2, different because of metabolic state, different because of cell-cycle position — produce different

outcomes from the same starting marks. But the outcome from a given combination of inputs is determined, not random. Two cells with identical mark configurations and identical enzyme complements at the same point in their cell cycles will, the model proposes, produce strongly similar outcomes at the addresses where this matrix operates — acknowledging that residual stochastic variance in enzyme kinetics and substrate accessibility introduces noise around a constrained central tendency. This is what the model means by input-constrained: outcomes are strongly constrained by measurable inputs, not determined by stochastic regulatory decisions, though the constraint is probabilistic rather than perfectly deterministic.

The connection to ARC phases is direct and gives the model its predictive power. During A-phase, the CLASPIN-mediated histone routing (whose evidence is reported in §6.4) creates strand-asymmetric starting conditions, and the partition systems (§3.2) deliver different enzymatic environments to each daughter. The matrix then operates differently in each daughter from the moment of division onward — same chromatin starting state in some respects, different enzymatic clientele, different outcomes. During R-phase, the matrix operates symmetrically on both daughters of each division, but the accumulation of permanent changes — $C \rightarrow T$ on the template strand, $G \rightarrow A$ at G4 addresses, structural collapse of i-motifs at insulator-installed sites — progressively narrows the regulatory options available to the cell. During C-phase, the cumulative changes have closed enough addresses that the remaining open repertoire can support only one lineage identity, and the resolution architecture described in §3.4 makes that identity irreversible.

This framing has a specific empirical implication that distinguishes the model from conventional epigenetic models. Conventional models treat chromatin marking as a regulated computation in which signaling inputs determine which marks are installed, and the marks then bias subsequent

transcription. The marks are inputs to a cellular decision; the decision is reached by integrating many inputs through pathways that converge on transcription factor activity. In the model proposed here, the marks are not inputs to a decision — they are the decision. Once installed, the matrix specifies what each mark can become, and the answer is not a function of further signaling. The cell's role is to determine where to install marks (the upstream signal of §3.1) and to ensure that the right enzymes are present in the right daughter (the partition systems of §3.2). After that, the chemistry runs.

3.3.6 The AT Arm: Enhancer Licensing Through Deliberate DNA Breaks

The three subsystems described above operate primarily at two classes of regulatory element: the $C \rightarrow T$ and $G \rightarrow A$ subsystems at promoters (CpG islands and G4-containing sequences), and the TET ladder at CTCF-bound insulator boundaries. A third class of regulatory element — the enhancer — is licensed through a distinct mechanism that completes the model's account of how commitment becomes coherent across the three-dimensional genome.

TOP2B (DNA topoisomerase II-beta) creates deliberate double-strand breaks at enhancers as a required step in enhancer-promoter loop formation. Madabhushi et al. (2015) demonstrated in neurons that activity-dependent transcription requires TOP2B-mediated DSBs at the promoters and enhancers of early-response genes, and that these breaks are not damage to be repaired but a required topological step in establishing the chromatin loops that bring enhancers into contact with their target promoters. EP300/CREBBP then deposits H3K27ac at the break-adjacent nucleosomes, establishing the active enhancer state.

The computational coupling data reported in Part 6 are consistent with the interpretation that TOP2B is commitment-associated: it correlates positively with commitment markers (EP300,

RIF1, CTCF, NUMB) and negatively with stem-state markers across multiple TCGA cohorts and at protein level in CPTAC data. The AT arm's coupling profile mirrors the insulator arm's: both are commitment-associated and both are anti-correlated with stem-state markers.

The functional consequence is that enhancer licensing and insulator installation are coordinated through shared upstream dependencies. TET2-mediated demethylation enables CTCF binding at boundaries (insulator arm) while the same commitment-associated chromatin remodeling enables TOP2B access at enhancers (AT arm). The two arms advance in parallel because they share the same metabolic permissions and the same temporal window within R-phase.

The model predicts that cancer should preserve the structural cohesin complex itself while weakening the regulatory coordination that directs where cohesin loads and what boundaries it establishes — the same selective dissolution of the decision layer with preservation of the machinery layer that the model predicts should be maintained across all model components.

3.3.7 Three-Arm Convergence: Why Commitment Requires Simultaneous Completion

The three arms of the write/erase matrix — the GC arm operating at G4-containing promoters, the AT arm licensing enhancers through TOP2B-mediated breaks, and the insulator arm establishing CTCF/cohesin boundaries — the model proposes, must complete simultaneously at each commitment-relevant gene for the commitment to be coherent. The insulator arm determines WHICH enhancer-promoter pair can interact by establishing the TAD boundary that constrains loop extrusion. The AT arm determines WHICH enhancer within the permitted TAD is active by licensing it through TOP2B breaks and EP300/CBP-mediated H3K27ac deposition (the canonical histone acetyltransferase activity of these transcriptional coactivators; Dancy & Cole, 2015). The GC arm determines WHEN the promoter fires by modifying the G4 structure

that controls transcription factor access. A gene whose insulator boundary is established but whose enhancer is not licensed will not activate. A gene whose enhancer is licensed but whose promoter G4 has not been modified will not respond. Commitment occurs when all three conditions are met simultaneously at a sufficient number of lineage-specifying genes — producing the coordinated activation of a lineage program rather than the piecemeal activation of individual genes.

The consequences of this convergence requirement for cancer — where partition failure disrupts the coordination of arm outputs between daughters — are developed in Part 5. The strand-specific predictions of §3.3.1–3.3.2 — that $C \rightarrow T$ events accumulate on the template strand while $G \rightarrow A$ events accumulate on the new strand, with daughters inheriting the consequences differentially — have not been tested at single-cell resolution in any mammalian system; the predictions are derived from established enzyme biochemistry and strand-inheritance logic but their operation as a coordinated strand-specific marking program remains a model proposal.

3.4 The Resolution Architecture: How Commitment Becomes Irreversible

The preceding sections describe how the write/erase matrix installs modifications at commitment-relevant addresses. These operations produce a chromatin landscape that is, in principle, reversible: oxidized bases can be excised, methylation can be restored, histone marks can be rewritten. The question this section addresses is how those reversible modifications become the permanent changes that define commitment (Figure 3). The answer lies in what happens when the replication fork encounters the modified addresses during S phase.

A cell in R-phase carries modified bases at commitment-relevant loci. When the replication fork arrives at one of these addresses, the outcome depends on a competition between resolution and

bypass. If one of the RecQ-family or FANCDJ-family helicases engages at the stalled fork, the obstacle is resolved at high fidelity. BLM dissolves replication-fork structures through the topoisomerase III α –RMI1/RMI2 dissolvasome (Manthei & Keck, 2013), WRN provides 3' $\rightarrow$ 5' exonuclease proofreading of any inserted bases (Brosh & Bohr, 2007; Croteau et al., 2014), and BRIP1 resolves G-quadruplex obstacles directly (London et al., 2008). When helicases engage successfully, the original sequence and chromatin state are preserved on both daughter chromatids. No irreversible change occurs at that address. But helicase engagement is not guaranteed. The model proposes that the cell's helicase capacity is finite, and that as R-phase deepens and more addresses carry modifications, more encounters exceed that capacity. When they do, the fork stalls and the outcome shifts to bypass.

The bypass decision is governed by the ubiquitination state of PCNA at the stalled fork. RAD6/RAD18-mediated monoubiquitination recruits Y-family TLS polymerases for direct bypass, while UBC13/HLTF-mediated polyubiquitination promotes error-free template switching using the sister chromatid (Hoege et al., 2002; Masuda et al., 2018). HLTF is particularly important at G4-containing addresses because it serves a dual function: it is a dsDNA translocase that can act on replication-blocking structures and the E3 ubiquitin ligase that promotes PCNA polyubiquitination and template switching (Masuda et al., 2018). When HLTF is present, it channels bypass toward the error-free template-switching pathway. When HLTF is absent, PCNA remains monoubiquitinated, channeling bypass toward mutagenic TLS. The finding that the HLTF-to-PRIMPOL expression ratio reverses in all examined cancer types relative to normal tissue is consistent with a systematic shift from error-free resolution to mutagenic bypass at G4-containing regulatory addresses.

The PCNA modification state at each stalled fork is not determined solely by these ubiquitin ligases. PIDD1, a self-processing protein first identified as a p53 transcriptional target (Lin et al., 2000), directly interacts with PCNA at stalled replication forks, where it modulates p21 displacement, promotes PCNA monoubiquitination, and facilitates pol η recruitment for translesion synthesis (Logette et al., 2011). PIDD1 also recruits DNA-PKcs to stalled forks independently of the Ku heterodimer, and this recruitment is required for ATR checkpoint activation in the replication-stress context — disrupting the PIDD1–DNA-PKcs interaction compromises ATR signaling, the intra-S-phase checkpoint, and cellular resistance to replication stress (Lin et al., 2018). The same protein localizes to the mother centriole through the distal-appendage adapter ANKRD26 (Evans et al., 2021; Burigotto et al., 2021), which means that when the two centrioles segregate to opposite spindle poles and are inherited by different daughters, the daughter receiving the mother centriole inherits the fork-proximal regulator of PCNA modification and ATR checkpoint strength — a physical mechanism by which centrosome identity could determine resolution-architecture capacity at the same genomic addresses in the two daughters.

When TLS does engage without helicase-mediated fidelity control, the consequences are strand-specific and predictable. REV1 scaffolds the TLS complex and inserts dCMP (Nelson et al., 1996; Sale, 2013). POLK bypasses 8-oxoguanine with error-prone fidelity (Markkanen, 2017). POLH handles UV photoproducts at near-replicative fidelity but is error-prone at oxidized bases (Masutani et al., 2000). REV3L extends past mismatches, locking them into the permanent sequence record (Lawrence, 2004). Because the two daughter chromatids carry complementary strands, the strand orientation of the obstacle relative to fork direction determines which daughter inherits the change. G4 structures on the leading-strand template impede the CMG helicase and

produce single-stranded gaps through fork uncoupling or PRIMPOL-mediated repriming, while those on the lagging-strand template inhibit Pol δ and impair Okazaki fragment maturation (Sethi et al., 2025). The asymmetry is structural and predictable from genomic architecture: template-strand changes are inherited by the daughter receiving the template chromatid, new-strand changes by the daughter receiving the new chromatid (Haradhvala et al., 2016; Ginno et al., 2012). The model proposes that this strand-specific inheritance is the mechanism by which the C $\rightarrow$ T stemness marking of §3.3.1 and the G $\rightarrow$ A commitment marking of §3.3.2 become permanently allocated to specific daughters.

Sarkies et al. (2010) demonstrated a critical additional consequence of replication failure at G4 addresses that bridges the sequence-level and epigenetic effects of bypass. Processive replication through G4 DNA requires REV1-mediated TLS to maintain coupling between DNA synthesis and parental histone recycling. When REV1 fails at a G4 address, the fork uncouples DNA synthesis from histone recycling. Parental histones carrying repressive marks are not recycled to both daughter strands; newly synthesized unmarked histones are incorporated instead, producing localized loss of repressive chromatin and transcriptional derepression. A single replication failure at a commitment-relevant G4 promoter therefore produces compound irreversibility on one chromatid: a possible sequence change from low-fidelity TLS, loss of epigenetic silencing from histone recycling failure, and underreplicated DNA if TLS fails entirely. The daughter inheriting that chromatid receives all three changes. The daughter inheriting the other chromatid, where the fork replicated normally with intact histone recycling, retains the original sequence, the original histone marks, and complete replication at that address. No single reversal chemistry can undo the compound change. The Sarkies finding was demonstrated in avian DT40 cells; the conservation of REV1, G4 structures, and histone recycling mechanisms across vertebrates

(Sarkies et al., 2012) makes the extension to mammalian transit-amplifying cells plausible, though it has not been directly demonstrated.

Obstacles that escape the successive resolution attempts spanning S-phase through mitosis — helicase engagement, translesion synthesis, mitotic DNA synthesis (MiDAS) at the metaphase-anaphase transition, and anaphase bridge resolution — are transmitted to the daughter cell as underreplicated DNA (Lukas et al., 2011). In the daughter's G1, these lesions are sequestered in 53BP1 nuclear bodies, large chromatin domains enriched in 53BP1, ATM, γ H2AX, MDC1, CTCF, and 5-hydroxymethylcytosine but lacking the HR factors CtIP and RAD51 (Fernandez-Vidal et al., 2017). 53BP1 NBs form preferentially at common fragile sites in late-replicating, initiation-poor regions with DNA secondary structures. TET depletion is associated with reduced 53BP1 NB formation through a mechanism involving replication-stress-related chromosome missegregation (Fernandez-Vidal et al., 2017), and the model interprets this as consistent with the proposal that TET-mediated modifications at commitment addresses create the replication obstacles whose transmission produces 53BP1 NBs. Direct demonstration that specific TET-modified bases at named commitment loci produce the underreplicated DNA that becomes 53BP1 NBs remains to be carried out. The asymmetric inheritance of damage-associated marks operates at two scales, both dependent on ATR. At the chromosome level, replication stress induces ATR/CHK1-dependent nonrandom segregation of damaged chromatids, concentrating the damage burden in one daughter while the other inherits the old template DNA with its established marks (Xing et al., 2020). The damaged daughter subsequently undergoes cell-cycle arrest and death. At the base level, ATR phosphorylates TET3 at stalled forks, driving 5-hydroxymethylcytosine accumulation at those specific addresses (Jiang et al., 2017), and because 5hmC is not maintained by DNMT1/UHRF1, it undergoes passive replication-dependent dilution

at each subsequent division (Inoue et al., 2011; Onodera et al., 2021). The two scales of asymmetry are complementary: the chromosome-level sorting concentrates damaged chromatids in the daughter slated for elimination, while the base-level dilution creates site-specific epigenetic differences between the chromatids themselves.

53BP1 NBs have consequences for cell fate that operate through several converging mechanisms. They activate p53, which increases p21, which suppresses CDK2 activity, producing an extended G1 (Fernandez-Vidal et al., 2017). This chain is consistent with the p21/CDK2 bifurcation at mitotic exit demonstrated in fibroblast systems (Spencer et al., 2013) and with the finding that G1 lengthening is necessary and sufficient for differentiation in neural progenitors (Lange, Huttner & Calegari, 2009). 53BP1 NBs also adjust replication timing of the affected loci to late S phase (Spies et al., 2019), likely through the 53BP1-RIF1 axis that controls replication timing genome-wide (Foti et al., 2016), enabling RAD52-mediated repair of the underreplicated DNA. When 53BP1 capacity is adequate, it limits end-resection and promotes high-fidelity RAD51-dependent resolution. At classical DSBs, exhausting 53BP1's chromatin-binding capacity switches resolution to mutagenic RAD52-dependent single-strand annealing (Ochs et al., 2016). Whether this capacity-dependent switch operates identically at 53BP1 NBs has not been directly tested, but the model proposes that the number of 53BP1 NBs a daughter inherits determines both the depth of G1 arrest through the p21/CDK2 axis and the mutagenic load of resolution through 53BP1 capacity saturation. 53BP1 NBs can also be resolved in G1 by NHEJ or non-canonical end-joining pathways, and potentially by POLQ-mediated microhomology-mediated end joining (Wood & Doublié, 2016), with the specific repair products depending on the nature of the lesion and the cell's repair capacity during the extended G1.

The progressive accumulation of these irreversible changes across successive R-phase divisions is what converts the buffered remodeling of R-phase into the irreversibility that defines C-phase. Each cycle transmits more unresolved intermediates and installs more permanent changes at commitment-relevant addresses. This is the memory mechanism: the daughter of an R-phase division inherits a chromatin landscape that is closer to commitment than her mother's was, because more addresses have undergone irreversible change. The rate at which this accumulation proceeds depends on the balance between resolution capacity and obstacle burden — a balance set by the metabolic environment (§3.1), the helicase complement, and the number of addresses the write/erase program has modified. The scientific community has accepted DDR chemistry as developmental decision-making in several established contexts. ATM is required for hematopoietic stem cell self-renewal through regulation of oxidative stress (Ito et al., 2004). RAG1/RAG2 create deliberate double-strand breaks at antigen receptor loci whose NHEJ resolution produces the irreversible sequence diversity that specifies lymphocyte identity (Schatz & Ji, 2011). SPO11 initiates meiotic recombination through programmed DSBs whose controlled location, timing, and number are regulated by the chromatin environment (Lam & Keeney, 2015). The model proposes that the resolution architecture described here serves a comparable developmental function in somatic tissues — a function whose disruption has direct consequences for cancer, developed in Part 5, and whose measurable signature is tested in Part 6. Each link in the chain from modified base through fork stall, helicase/TLS competition, 53BP1 NB formation, and p21/CDK2-mediated G1 extension has been characterized in independent experimental systems; the proposal that these links operate as an integrated sequence within the same cell during the same R-phase division has not been demonstrated in a single experiment

and remains the model's most important mechanistic claim. The IDH1 mutation provides a particularly clear test of this architecture, developed in §5.3.

3.5 Execution Chemistry: How Commitment Becomes Physical

Terminal differentiation in each characterized lineage engages a specific death-pathway chemistry at sub-lethal levels — lipid peroxidation in erythropoiesis (Yang et al., 2014; Stockwell et al., 2017), caspase cleavage in neurogenesis (Fernando et al., 2005; Rosa et al., 2024), MLKL in adipogenesis (Magusto et al., 2022). These are not independent from cell death; they share molecular machinery with it, and the cell's partition state determines which regulatory programs deploy these shared activities.

The resolution architecture described in §3.4 closes regulatory options through compound irreversible changes at commitment-relevant addresses, but it does not, by itself, transform the cell physically. A cell that has crossed the commitment threshold has lost the capacity to revert to its prior chromatin configuration, but it still has the same morphology, the same cytoskeleton, the same membrane composition, the same nuclear architecture as before commitment. The extended G1 that 53BP1 NBs create through the p21/CDK2 pathway (§3.4) is the same G1 that provides time for the execution chemistry described below to translate the locked configuration into physical change. To become a neuron, an erythrocyte, a keratinocyte, or any other terminal cell type, the cell must undergo physical changes that consolidate its new identity into a form that cannot revert under homeostatic conditions. These physical changes are mediated, the model proposes, by the same molecular machinery that executes regulated cell death — operating at sub-lethal intensity in regulatory contexts that constrain the chemistry to differentiation-relevant outcomes rather than complete cellular destruction. The model's claim is that there is no separate

"differentiation chemistry" distinct from "death chemistry"; there is one chemistry, deployed at different magnitudes in different regulatory contexts.

The principle that death and differentiation engage the same molecular machinery at different magnitudes is supported by direct biochemical evidence in multiple lineages. In neural and muscle systems, sub-lethal caspase-3 activation is required for terminal differentiation. Fernando et al. (2002) demonstrated that pharmacological inhibition of caspase-3 blocks myogenic differentiation in skeletal muscle progenitors, and that the requirement is for a low level of caspase-3 activity rather than for the high-level apoptotic activation that would kill the cell. The same requirement applies in neural stem cell differentiation (Fernando et al., 2005). Rosa et al. (2024) showed that non-apoptotic caspase-3/7 activation directly drives neural stem cell-to-neuron conversion in the adult zebrafish pallium, with the transcription factor ATF3 mediating the fate bias toward lineage termination and neuron formation. In osteoclastogenesis, caspase-8 activation promotes scramblase-mediated phosphatidylserine exposure and fusion of osteoclast precursors through a non-apoptotic cascade — pharmacological blockage or genetic deletion of caspase-8 interferes with osteoclast fusion and bone resorption (Veselá et al., 2022, 2024). Larsen et al. (2010) showed that caspase-3 activates the caspase-activated DNase (CAD/DFFB), which generates DNA strand breaks at the p21 (CDKN1A) promoter during differentiation, and that these breaks are required for the chromatin reorganization that accompanies terminal commitment. CAD is a nuclease with established roles in chromatin architectural maintenance at insulator-zone boundaries. The strand breaks CAD generates are a form of regulated genomic damage that is repaired in the differentiating cell but that, in the apoptotic context with much higher caspase activity, would lead to the systematic chromatin fragmentation characteristic of programmed cell death.

This principle generalizes across lineages, with each lineage engaging a different specific death-associated chemistry as its execution chemistry. In the erythroid lineage, controlled lipid peroxidation — the same iron-dependent chemistry that drives ferroptotic cell death (Yang et al., 2014) — produces the membrane changes required for reticulocyte maturation. In adipogenesis, MLKL-mediated membrane permeabilization participates in lipid droplet maturation (Magusto et al., 2022). In keratinocytes, the lineage-specific caspase-14 processes cornification substrates with no apoptotic function (Denecker et al., 2007). The nucleases implicated in chromatin architectural maintenance — CAD, EndoG, and AIF — are also established components of regulated cell death pathways (Larsen et al., 2010; Dahal et al., 2022; Susin et al., 1999). The model interprets this molecular overlap as reflecting a continuous system in which the same enzymatic repertoire is deployed at sub-lethal magnitude for chromatin maintenance and at lethal magnitude for cell elimination, with the regulatory context determining outcome. Direct demonstration that individual nucleases shift from maintenance to execution function within the same cell across different ARC phases has not been performed.

Whether the same pattern holds in lineages where it has not yet been measured — hepatocytes, pancreatic acinar cells, intestinal enterocytes — is a generalization the model offers as a testable extension rather than an established finding. The directly supported observation is that each characterized lineage engages a specific death-pathway component as a sub-lethal commitment effector. The broader claim that execution chemistry is required across all lineages would require systematic profiling across additional lineages with their corresponding death-machinery pharmacology.

The architectural claim has a direct falsification test. If terminal differentiation can be achieved in multiple lineages while all execution chemistries are simultaneously blocked — pan-caspase

inhibition (blocking the caspase pathway), GPX4 overexpression to its highest physiologically tolerable levels (blocking the lipid peroxidation pathway), and MLKL inhibition (blocking the membrane permeabilization pathway) — then the architectural constant claim is wrong.

Differentiation would be possible without any death-pathway chemistry, and the parallel between the two would be coincidental rather than mechanistic. The model predicts that such combined blockade will not permit terminal differentiation in at least three lineages — a kill test whose positive result would require fundamental revision.

The model proposes that the BCL-2 family membrane homeostasis system provides the physical mechanism for why these execution chemistries are lineage-specific. Each lineage's BCL-2 family configuration — set by its niche signal (Wnt → BCL-XL, Hedgehog → BCL2, HSC niche → MCL1) — determines which organelle configuration the cell maintains and therefore which membrane-failure mode it engages during commitment. A cell configured by BCL-XL (high Ca^{2+} transfer, efficient ATP synthase) engages lipid peroxidation chemistry when BCL-XL's membrane-stabilizing function is titrated down during commitment. A cell configured by BCL2 (dampened ER Ca^{2+}) engages caspase-mediated remodeling when BCL2's IP3R-inhibiting function is withdrawn. A minority of mitochondria can undergo MOMP in a stress-regulated manner without killing the cell, producing limited caspase activation that leads to DNA damage and chromatin remodeling rather than death (Ichim et al., 2015) — this "minority MOMP" is the sublethal engagement of the same machinery that, at full magnitude, would execute the cell. The execution chemistry for each lineage is not a separate chemistry deployed alongside the BCL-2 system; it is the specific mode of controlled membrane reorganization that occurs when the relevant BCL-2 family member's maintenance function is withdrawn during commitment. The individual biochemical functions cited here are established; the integration into a lineage-specific

execution system — the claim that BCL-2 member identity determines execution chemistry type — is the model's interpretive contribution.

The execution chemistry connects to the partition systems described in §3.2 through the BCL-2 family configuration each daughter inherits. The chemistry available in each daughter at the moment of the C-phase mitosis biases which execution chemistry engages in that daughter. A daughter whose inherited BCL-2 configuration maintains high organelle quality — MCL1-stabilized cristae, BCL-XL-mediated ATP synthase efficiency, low net OMM-resident BAX — has the metabolic capacity for continued self-renewal. A daughter whose inherited configuration has lower organelle quality — reduced MCL1 matrix levels, BCL-XL partially withdrawn from ATP synthase and fission sites, higher BAX OMM residence — has the substrates available for the lineage-specific execution chemistry: sub-lethal caspase activity at chromatin-remodeling loci, controlled lipid peroxidation for membrane reorganization, or MLKL-dependent changes, depending on the lineage. A daughter inheriting an incoherent combination — commitment-marked chromatin requiring execution chemistry that the inherited organelle configuration cannot support — faces the chemical incompatibility that the BCL-2 quality-control mechanism would normally detect and eliminate. In cancer, where that quality-control mechanism has been disrupted (Part 5), the incoherent daughter survives.

The execution chemistry completes the ARC architecture by translating commitment — the compound irreversible changes at lineage-relevant regulatory addresses described in §3.4 — into physical cellular change at the organelle, membrane, and cytoskeletal levels. The execution chemistry does not make the commitment decision; it implements it. The model therefore predicts that execution-layer gene pairs should be preserved in cancer (the structural execution partnerships FANCI×FANCD2, PLK1×AURKA, CLSPN×PCNA, CDK4×CCND1) — the

execution machinery functions regardless of whether the upstream regulatory architecture that should control it is intact. Two pairs — MFN1 $\times$ OPA1 and FANCD2 $\times$ BRCA1 — are classified as regulatory (Class I) rather than execution because both have partner-independent functions and their transcriptomic co-expression reflects shared regulatory programming rather than stoichiometric complex assembly. Cancer cells retain functional caspases, functional lipid peroxidation chemistry, and functional cytoskeletal remodeling machinery. What they lack is the partition-controlled regulatory system that would direct these tools toward coherent commitment outcomes, and the selective purging mechanism (developed in §4.3) that would eliminate cells whose partition-inherited configuration is incoherent.

Part 4: Death as a Commitment Endpoint

4.1 The Death-Differentiation Overlap

The execution chemistry developed in §3.5 established that death and differentiation engage the same molecular machinery at different intensities, with each lineage's BCL-2 family configuration determining which execution chemistry is deployed. The question Part 4 addresses is how the partition state at mitosis determines which deployment each daughter engages. The model proposes that daughters of an asymmetric division inherit different organelle portfolios and chromatin configurations through the molecular asymmetry described in §2.3, and that the BCL-2 family configuration each daughter inherits determines which death-associated programs are accessible. A daughter whose inherited configuration has collapsed — MCL1 matrix loss producing cristae opening, BCL-XL withdrawal permitting BAX accumulation at the OMM — engages the intrinsic mitochondrial pathway, with the specific death modality reflecting which BCL-2 member's function failed (§3.5). Daughters receiving aged mitochondria with looser cristae architecture inherit organelles with higher ROS production capacity, influencing intrinsic

apoptotic competency independently of BCL-2 family expression levels (Katajisto et al., 2015). The model treats this as a regulatory bias rather than an absolute constraint: pharmacological blockade of one death pathway can redirect cells to another (Galluzzi et al., 2018), but under physiological conditions the partition-inherited configuration biases the cell toward a specific death modality.

4.2 Organelle Inheritance as a Determinant of Death-Pathway Accessibility

Direct evidence that organelle inheritance at division determines daughter cell fate has expanded substantially in recent years. Loeffler et al. (2019) demonstrated in mouse haematopoietic stem cells that lysosomes, autophagosomes, mitophagosomes, and the protein NUMB are asymmetrically inherited at division, and that asymmetric lysosomal inheritance predicts the future metabolic and translational activation of HSC daughters and their offspring. Loeffler et al. (2022) extended this to human blood stem cells, showing that lysosomes and active mitochondria are asymmetrically inherited in a coordinated, nonrandom process — multiple additional organelles including autolysosomes and recycling endosomes co-segregate with lysosomes — and that the asymmetric lysosomal inheritance predicts daughter cell-cycle length, differentiation, and stem cell marker expression while asymmetric mitochondrial inheritance correlates with daughter metabolic activity. Candelas et al. (2024) showed that specific heterotypic interactions with osteoblast and endothelial cells promote asymmetric division of human HSPCs: upon niche interaction, cells polarize their centrosome, Golgi apparatus, and lysosomes toward the contact site and orient their spindle perpendicular to it, producing daughters with unequal amounts of lysosomes and of the differentiation marker CD34. The organelle asymmetry is not a passive consequence of imperfect segregation — it is an actively organized process that the niche regulates.

The functional cargo that the centrosome carries extends beyond microtubule-organizing capacity. Because PIDD1 localizes to the mother centriole's distal appendages through ANKRD26 (§3.4), centrosome inheritance is simultaneously an inheritance of resolution-architecture capacity — the daughter receiving the mother centriole inherits the fork-proximal adapter that controls PCNA modification dynamics and ATR checkpoint strength, while the daughter receiving the newly assembled centriole does not. Centrosome asymmetry therefore delivers both structural polarity (spindle orientation, organelle distribution) and molecular resolution capacity to each daughter in a single co-inherited package.

The causal relationship between organelle inheritance and fate has been established most directly in the neural lineage. Bunel et al. (2025) used live imaging in the chick embryonic neuroepithelium to demonstrate that the frequency of unequal mitochondrial distribution increases in parallel with the rate of asymmetric divisions during development, and that a cell inheriting fewer mitochondria than its sister consistently differentiates into a neuron. Critically, they used a chemogenetic approach to experimentally displace mitochondria during mitosis, forcing unequal inheritance, and found that this manipulation drives premature neuronal differentiation — establishing a direct causal relationship between unequal mitochondrial inheritance and asymmetric cell fate *in vivo*.

The convergence of these findings with the lineage-specific execution chemistry described in §3.5 generates the model's proposal about death-modality determination. If the inherited organelle complement differs between daughters — one receiving lysosome-enriched, aged-mitochondria configurations and the other receiving lysosome-depleted, young-mitochondria configurations — then the death-pathway machinery available to each daughter differs correspondingly. Lysosomes contain cathepsins whose release during lysosomal membrane

permeabilization can activate the NLRP3 inflammasome and engage the pyroptotic cascade; mitochondria with looser cristae produce higher ROS, promoting lipid peroxidation and ferroptotic susceptibility; and the MLKL-dependent membrane permeabilization that participates in adipocyte differentiation (Magusto et al., 2022) operates through a necroptotic machinery whose availability depends on the inherited membrane configuration. Each of these connections has been established individually in the literature cited here and in §3.5. The model's interpretive contribution is the proposal that the organelle complement inherited at a single asymmetric division configures which of these death pathways is accessible in each daughter — producing lineage-appropriate death modality through the same partition event that specifies lineage-appropriate chromatin configuration.

The lineage-specific observations assembled across §3.5 and the evidence cited above suggest a pattern: mesoderm-derived lineages show particular engagement with necroptotic machinery during both differentiation (MLKL in adipogenesis; Magusto et al., 2022) and aging (RIPK1-RIPK3-MLKL in HSC decline; He et al., 2024), ectoderm-derived lineages show ferroptosis susceptibility tied to lipid-rich membrane composition and GPX4 dependence, and sub-lethal caspase activation appears across multiple lineages as a differentiation mechanism. Whether this pattern reflects a systematic germ-layer-level organization of death-pathway susceptibility or represents independent lineage-specific adaptations is an open question that the model raises but cannot resolve with the data presented. The testable prediction is that organelle profiling at division — quantifying the lysosomal, mitochondrial, and ER complement each daughter inherits — should predict which death modality that daughter engages if subsequently forced toward commitment.

PIDD1 bridges the three temporal layers described in the preceding sections: the fidelity of S-phase marking (§3.4), the regulated partitioning of marked components at mitosis (§3.2), and the fate outcomes in the following G1. The transition from fork-level resolution to cell-fate outcome operates across a temporal hierarchy separated by mitosis, and PIDD1 connects the damage-responsive and polarity-directed components of this hierarchy at each step. During S-phase, PIDD1 at the fork facilitates ATR activation, which drives both the nonrandom chromosome segregation that concentrates damage in one daughter (Xing et al., 2020) and the 5hmC marking described in §3.4. After mitosis, PIDD1 — now inherited with the mother centriole — can activate the PIDDosome through an entirely distinct, centrosome-triggered route (Burigotto et al., 2021), and the centrosome itself physically couples this axis to the polarity system by nucleating the astral microtubules that align the NUMB crescent (Giansanti et al., 2001). PIDD1 undergoes intein-like autoprocessing whose depth determines output: an initial cleavage generates a survival-competent fragment, while a second cleavage liberates the death-domain fragment that activates caspase-2, the most evolutionarily ancient caspase (Tinel et al., 2007; Krasovec et al., 2023). The death decision therefore executes in the daughter after mitosis has already distributed both the damage-associated marks and the polarity determinants. The model proposes that these two partition arms converge on a single target: NUMB prevents MDM2-mediated p53 degradation by complex formation (Colaluca et al., 2008), while caspase-2 destroys MDM2 by direct cleavage (Fava et al., 2017; Rizzotto et al., 2021), so that the daughter inheriting both the NUMB crescent and the PIDDosome-competent centrosome would receive dual-input p53 stabilization whose dissolution in cancer would disable both quality-control arms simultaneously. This integrated bridging function — connecting fork-derived damage sorting to

polarity-directed identity assignment through a single centrosome-localized hub — has not been demonstrated as an integrated sequence.

4.3 Partition State and Selective Purging

The biology of selective purging is well established in multiple developmental contexts. In thymic selection, T-cell precursors whose TCR signaling falls below a survival threshold are eliminated by BIM-dependent apoptosis, while those with appropriate signaling receive BCL-2-mediated survival (Strasser, 2005). In neuronal development, neurons compete for limiting trophic factors, with those failing to establish appropriate connections eliminated through BAX-dependent death (Oppenheim, 1991). In epithelial tissues, cell competition eliminates cells with lower fitness relative to their neighbors through mechanisms involving the BCL-2 family and mechanical signaling (Merino et al., 2016). In each case, the cell's internal state — its signaling capacity, metabolic fitness, or competitive position — determines whether it survives or is purged. No external evaluator is required; the selective pressure emerges from the cell's own configuration interacting with its environment.

The model proposes that the same principle operates on partition-inherited states. A daughter whose inherited organelle configuration and chromatin marks are coherent — commitment-marked chromatin paired with commitment-configured organelles — deploys the shared death/differentiation chemistry at the sub-lethal magnitude that Larsen et al. (2010) identified as productive for chromatin reorganization. A daughter whose inherited configuration is incoherent encounters sustained DDR signaling at mismarked addresses, which escalates the same chemistry to lethal magnitude. The outcome is determined by the partition-inherited configuration, not by a dedicated evaluation event. Daughters with coherent partitions differentiate; daughters with incoherent partitions are purged through the same BCL-2 family-

dependent mechanisms that operate in thymic selection and neuronal competition, applied to the partition context.

This account generates a specific prediction about cancer: the coupling between apoptotic effectors and the DDR-repair chemistry should weaken in cancer, reflecting the loss of the partition-dependent fitness selection that would normally connect the two. The coupling between stem-program markers and the purging machinery should be preserved, while the coupling between failed-commit markers and the same machinery should dissolve. Part 6 tests both predictions.

Part 5 develops the cancer application of this architecture, in which partition failure simultaneously disrupts both coherent differentiation and coherent death — the dual failure that the model proposes distinguishes cancer cells from normal proliferating progenitors.

Part 5: Cancer as a Specific Failure of the Partition System

5.1 Reframing the Cancer Question

The model presented in Parts 2 through 4 describes a system in which cell fate is determined by the combination of marks and enzymatic machinery each daughter inherits at mitosis, with the partition systems of §3.2 ensuring that the combinations are coherent, the execution chemistry of §3.5 translating commitment into physical change, and the death modalities of Part 4 selecting which terminal outcome the cell engages. The coordination this model predicts is quantifiable: partition coupling varies systematically with regenerative capacity across normal tissues, and execution partnerships are selectively reconfigured during differentiation while commitment-address coordination is comparatively preserved (Figure 4). The model proposes that a cell with an intact partition can differentiate coherently, and a cell with an intact partition can die

appropriately to its niche context. A cell whose partition has failed loses access to coherent versions of both — though the severity of this loss varies with the tissue context and the nature of the partition disruption. Within this framing, cancer may not be primarily a disease of uncontrolled proliferation or of accumulated driver mutations alone (Vogelstein et al., 2013), nor a disease in which signaling pathways have been hijacked to promote inappropriate growth. Cancer is, the model proposes, interpretable as a convergent architectural state — disrupted regulatory coordination, impaired partition-dependent fate resolution, and preserved execution machinery — that diverse oncogenic lesions may produce through different molecular routes. This framing offers not merely another hallmark to be added to the existing catalog (Hanahan & Weinberg, 2011) but a mechanistic account of how disparate oncogenic events converge on a shared architectural phenotype. ARC does not require partition failure to be the initiating lesion in every cancer. Partition failure may be initiating, permissive, secondary, or selected during progression depending on tissue context.

This framing has substantive consequences for how the cancer literature is read. Many recurrent cancer loci — including APC, KRAS, TP53, SMAD4, and PIK3CA — contain G-quadruplex-forming sequences in their promoter regions (Eddy & Maizels, 2006; Hänsel-Hertsch et al., 2017). These promoter G4 structures are among the addresses at which the write/erase matrix of §3.3 operates, and whose modifications must be partitioned correctly between daughters for normal asymmetric division. An important distinction must be drawn: the canonical driver mutations at these loci are primarily coding-region events (KRAS G12/G13/Q61 activating mutations, APC truncating mutations, TP53 missense mutations in the DNA-binding domain) that arise through mechanisms distinct from the promoter G4-localized mutagenesis the model describes. The model does not claim that partition failure directly produces the specific coding

mutations that activate KRAS or inactivate TP53. Rather, the model proposes that partition failure disrupts the regulatory coordination at G4-containing promoter addresses where the write/erase program operates, and that this disruption produces a permissive cellular state in which additional mutations — including the coding-region driver events — accumulate at elevated rates because the selective purging machinery (§4.3) is no longer engaged. The testable prediction is therefore specific to the regulatory layer: tissues with measurable polarity disruption should show elevated mutation accumulation rates specifically at G-quadruplex-containing regulatory addresses compared to partition-intact tissues, with the degree of elevation correlating with the degree of polarity disruption.

Figure 4. ARC Coordination Architecture in Normal Tissues and Along Differentiation Trajectories

Coordination exists in normal tissues, varies with regenerative capacity, and is selectively reconfigured during differentiation.

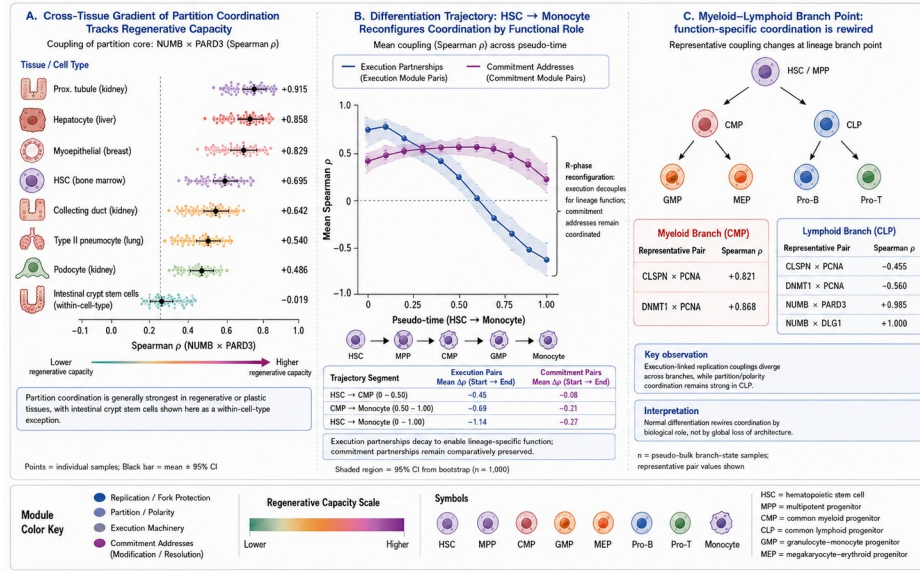


Figure 4. Coordination architecture across normal tissues and differentiation trajectories. Donor-level pseudo-bulk Spearman correlations from the CZI CELLxGENE Census (2024-07-01 release). (A) Cross-tissue NUMB × PARD3 gradient ($n = 127$ donors, seven pseudo-bulk cell types; intestinal crypt value from within-cell-type analysis) ordered by regenerative capacity: intestinal crypt stem cells ($\rho = -0.019$), podocytes (+0.486), type II pneumocytes (+0.540), collecting duct (+0.642), HSCs (+0.695), myoepithelial (+0.829), hepatocytes (+0.858), proximal tubule (+0.915). (B) HSC → monocyte trajectory: execution-pair mean (CLSPN × PCNA, DNMT1 × PCNA, PLK1 × AURKA) drops from +0.937 (HSC, $n = 9$) to -0.031 (monocyte, $n = 22$), $\Delta = 0.968$; commitment-pair mean (TDG × EZH2, OGG1 × BAX, TDG × BAX, OGG1 × EZH2) drops from +0.875 to +0.708, $\Delta = 0.167$. (C) Myeloid-lymphoid routing: replication coupling inverts from CMP (+0.821, +0.868; $n = 18$) to CLP (-0.455, -0.560; $n = 18$) while partition coupling maximizes in CLP (NUMB × PARD3 = +0.985, NUMB × DLG1 = +1.000).

5.2 The Cancer Architecture: Dissociated Commitment Chemistry

The resolution architecture of §3.4 and the selective purging logic of §4.3 together generate a specific prediction about what cancer cells should look like at single-cell resolution. A cell whose partition has failed should show two coupled signatures. The chemistry that installs commitment-direction sequence changes should remain engaged — the cell has entered the commitment trajectory — but the helicase-dependent fidelity controls that would normally direct that chemistry to lineage-appropriate addresses should be absent. Simultaneously, the fitness-

selection machinery that should purge daughters with incoherent partition states should be decoupled from the commitment-direction marker. The prediction is not that cancer cells show more mutagenic chemistry than normal cells. It is that cancer cells show a specific internal dissociation: commitment chemistry operating without its paired controls, in a chromatin context where the purging machinery no longer engages.

Because the same partition failure that produces rogue installation of sequence changes also disables the selective purging that would normally eliminate daughters with incoherent marks (§4.3), the defective cells persist and accumulate additional changes across successive divisions — a temporal dynamic consistent with the years-to-decades timeline of cancer development. This dissociation should be visible within tumors that contain both stem-program and failed-commit subpopulations. Cells expressing the stem-program marker should retain coordinated coupling to both fidelity-providing helicases and fitness-selection machinery. Cells expressing the failed-commit marker should show the opposite: TLS-associated chemistry engaged, helicase fidelity excluded, and every component of the purging machinery decoupled. The two subpopulations coexist in the same tumor but occupy architecturally distinct states. The coupling values from the head-and-neck squamous carcinoma single-cell dataset of Puram et al. (2017), where TP63-expressing cells (squamous stem program) and KRT5-expressing cells (commitment-direction) provide the required within-tumor contrast, are reported in §6.5.1 and displayed in Figure 5A.

The cytoplasmic-level signature of the same architecture should appear in the polarity-coupling patterns tested across TCGA cohorts. The cortical polarity network of §3.2 — the distributed system of NUMB, PARD3, PRKCZ, PRKCI, and DLG1 — should lose its regulatory coupling in cancer while retaining the structural execution partnerships that the machinery requires for

proliferation. The model predicts that this dissolution should propagate outward from the cortical polarity node, with the strongest disruption at direct polarity partners and attenuating signal at more distal regulatory links. Part 6 tests this prediction across nine cancer types and reports the distance-dependent attenuation pattern.

The combined prediction — chromatin-level dissociation of commitment chemistry from its controls, plus cytoplasmic-level dissolution of polarity coupling with preserved execution partnerships — constitutes the model's central architectural claim about cancer. The cancer cell is not distinguished from its normal counterpart by the quantity of its molecular machinery. It is a cell whose regulatory coordination has dissolved while its execution machinery remains intact. Part 6 tests this claim quantitatively.

5.3 The IDH1 Coupling: A Tissue-Specific Gate

The cancer architecture described in §5.2 predicts that partition failure should produce different phenotypes in different tissues, because different tissues organize their commitment chemistry around different cofactor dependencies. One prediction is sufficiently grounded in established biochemistry to test at protein level: the α -ketoglutarate gate.

The TET dioxygenases and the Jumonji-family histone demethylases that drive the R-to-C transition require α -ketoglutarate as an obligate cosubstrate (Iyer et al., 2009; Tsukada et al., 2006). In tissues where commitment depends heavily on these enzymes, the cytoplasmic supply of α -ketoglutarate — produced by IDH1 — should be coupled to the commitment program. In tissues where commitment uses cofactor-independent chemistry, no such coupling should exist. The model predicts that IDH1 should be positively coupled to commitment-associated markers

specifically in lineages where α KG-dependent dioxygenase activity is rate-limiting for the R-to-C transition, and uncoupled or anti-coupled in lineages where it is not.

The neural lineage is the clearest test case. Neural progenitor differentiation depends heavily on TET-mediated demethylation (Hahn et al., 2013) and Jumonji-mediated histone demethylation at lineage-specifying promoters. The model therefore predicts that IDH1 protein should be positively coupled to regulatory coordination components specifically in glioblastoma and not in cancers of other lineages. Part 6 tests this prediction across seven CPTAC proteomic cohorts, with IDH1 $\times$ CDK6 providing the primary protein-level test (§6.4). The hematopoietic lineage provides an independent example: IDH1 R132H is oncogenic in both neural and hematopoietic contexts, consistent with both lineages depending on α KG-dependent commitment chemistry. The clinical differentiation response observed when IDH1-mutant AML is treated with ivosidenib (DiNardo et al., 2018) is consistent with a gate-restoration interpretation — removing the 2-hydroxyglutarate block allows TET-mediated commitment chemistry to resume — though this clinical observation does not by itself demonstrate the model's specific architectural claim.

This tissue-restricted coupling is consistent with the observation that IDH1 R132H is oncogenic specifically in the neural and hematopoietic lineages. The mutation produces 2-hydroxyglutarate (Dang et al., 2009), which competitively inhibits the α -KG-dependent dioxygenases. In a tissue where the R-to-C transition depends on these dioxygenases, 2-HG blocks commitment: the lineage genes stay methylated, the cell remains in R-phase, and the result is the expanded progenitor phenotype that characterizes IDH-mutant glioma and AML. In a tissue where the R-to-C transition does not depend on α -KG (because the commitment gate uses a different cofactor — for example, the APC-dependent spindle-orientation gate in intestinal epithelium), IDH1

R132H would be predicted to have minimal oncogenic effect because it poisons a gate the tissue does not use

The clinical evidence is consistent with this interpretation. IDH1 and IDH2 inhibitors produce terminal differentiation of AML blasts (Stein et al., 2017; DiNardo et al., 2018) and slow glioma progression in patients with the corresponding mutations. The model interprets these agents as re-opening the R-to-C gate by restoring α -KG availability and allowing the blocked dioxygenase program to complete. The clinical observation that IDH inhibitors produce differentiation rather than death — that the blasts mature into functional granulocytes rather than dying — is among the strongest existing clinical observations consistent with the model's account of cancer as R-phase trapping reversible by gate restoration.

5.4 Five Failure Modes Across Lineages

The partition failure described in §5.2 should be present across cancer types in the manuscript's panel, but the specific phenotype the failure produces should vary between lineages because different lineages organize their commitment chemistry differently. The five-mode classification developed here identifies this variation: cancers differ in which architectural feature of the §5.2 failure they exhibit. The five modes are not mutually exclusive — cancers can show combinations — but each mode anchors a distinct mechanism by which partition failure expresses as a clinically distinguishable phenotype. The modes are presented as model-derived hypotheses about how a common upstream failure produces lineage-specific downstream phenotypes. Each mode is named, defined by its predicted architectural signature, and anchored to a representative cancer type. The quantitative classification criteria and per-cohort measurements are reported in Part 6 and Methods §9.6. The quantitative evidence supporting each mode is in Part 6; therapeutic implications are consolidated in §7 (Figure 5).

Mode 1 — phase-specific cliff. In cancers where the commitment chemistry coupling is preserved through G0/G1 and S phases and collapses sharply at G2/M — the phase where the consequences of S-phase resolution decisions (helicase engagement versus TLS bypass) would manifest as pre-mitotic commitment or its failure. TLS itself operates primarily during S-phase, but the coupling signature of fidelity loss becomes measurable at the G2/M transition. The predicted signature is phase-resolved: coupling between a commitment-direction lineage marker and the HR machinery should be essentially zero through early phases but turn strongly negative at G2/M, while stem-program cells in the same tumor retain positive coupling at all phases. The Mode 1 architecture describes cells that maintain proper machinery engagement through DNA synthesis but exclude the resolution controls precisely at the phase where their consequences would normally lock in lineage-appropriate changes. HNSC is the model's exemplar for Mode 1, tested in Part 6.

Mode 2 — Phase-uniform Lineage Decoupling. Where the failure targets a lineage program rather than a phase boundary, the predicted signature is contrast: within a single tumor, one lineage marker should be decoupled from the commitment chemistry at every phase, while a different lineage marker in the same cells retains its coupling. The cells are configured to execute commitment chemistry directed by the stem-program signature but not by the commitment-direction signature, regardless of cell-cycle phase. The failure is at the lineage program rather than at a phase boundary. GBM is the model's exemplar for Mode 2, where GFAP (astrocytic commitment) and SOX2 (stem program) provide the within-tumor contrast, tested in Part 6.

Mode 3 — Progenitor Lock-in. Progenitor lock-in appears when the commitment chemistry retains coupling to progenitor-stage lineage markers while losing coupling to mature-stage markers. The predicted signature is a sign split between two lineage markers from the same

developmental hierarchy: the progenitor marker should couple positively with HR effectors while the mature marker couples negatively or not at all. The cell can run commitment chemistry, but only at addresses the progenitor program specifies. The chemistry that would carry the cell forward to mature identity is decoupled. This is the predicted molecular substrate of arrested differentiation — the cancer cell locked at the progenitor stage rather than at any specific cell-cycle phase. HCC is the model's exemplar for Mode 3, where AFP (fetal progenitor) and ALB (mature hepatocyte) provide the developmental-hierarchy contrast, tested in Part 6.

Mode 4 — TF-target Bifurcation. TF-target bifurcation appears when a master transcription factor remains coupled to the commitment chemistry but its downstream target genes are not. The predicted signature is a sign split between a TF and its canonical target: the TF should couple positively with HR effectors while the target gene couples negatively. The master TF is engaged with commitment chemistry but the gene that should execute the differentiation program is not. The mechanism predicts cells in which TF-driven survival programs operate independently of the differentiation program the TF normally specifies. PRAD is the model's exemplar for Mode 4, where AR (master TF) and KLK3 (canonical AR target) provide the TF-target contrast, tested in Part 6. The clinical benefit of olaparib in HR-deficient PRAD (de Bono et al., 2020) is consistent with this Mode 4 architecture, and the model predicts that AR-driven cancers without the KLK3 decoupling should show different therapeutic responses.

Mode 5 — Quiescent Evasion. Quiescent evasion appears when the failure is the absence of daughter generation rather than the failure of daughters that are generated. The predicted signature is at the phase-distribution level: cancer cells should avoid the selective purging that accompanies division by not entering G2/M, producing a substantial reduction in the G2/M phase fraction compared to matched normal progenitor cells. The few cancer cells that do reach

G2/M should show weakened commitment-direction coupling, consistent with lineage-coupling loss even among the cycling fraction. AML is the model's exemplar for Mode 5, tested in Part 6 with the caveat that the CEBPA $\times$ BRCA2 comparison in the cycling fraction is a directional observation at the resolution limit of 10x 3' scRNA-seq, not a definitive quantitative finding. Definitive quantification requires higher-sensitivity methods such as SMART-Seq2 or protein-level CITE-seq.

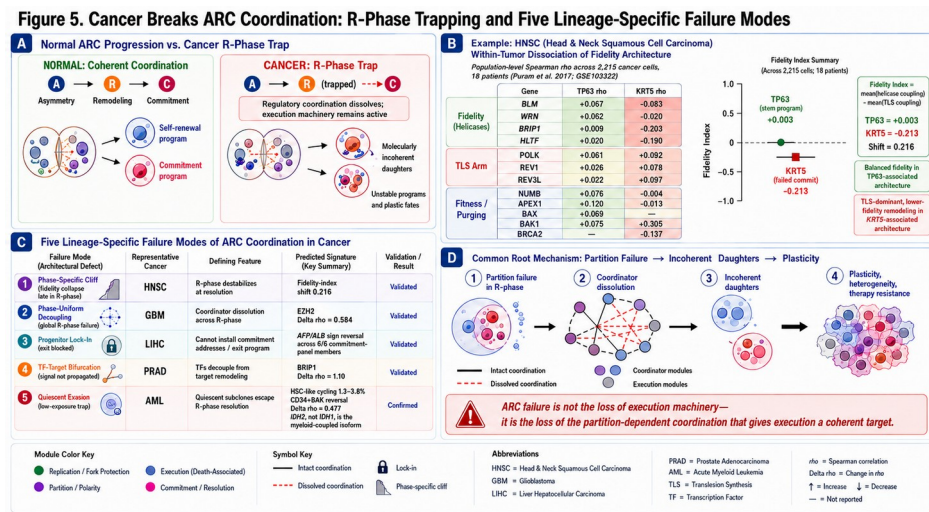


Figure 5. Cancer coordination failure: R-phase trap and five lineage-specific failure modes. Partition-dependent coordination dissolves in cancer, producing molecularly incoherent progeny despite intact execution machinery. (A) Normal ARC progression versus cancer R-phase trap. (B) Within-tumor dissociation of fidelity architecture in HNSC (Puram et al., 2017; n = 2,215 cancer cells, 18 patients; GSE103322). Population-level Spearman ρ between lineage markers (TP63 stem-program, KRT5 failed-commit) and helicase, TLS, and fitness-selection genes. TP63+ cells retain balanced helicase/TLS coupling; KRT5+ cells show systematic helicase exclusion (BRIP1 $\rho = -0.203$, HLTF $\rho = -0.190$) with TLS engagement (POLK $\rho = +0.092$, REV3L $\rho = +0.097$) and fitness-selection decoupling (NUMB $\rho = -0.004$, APEX1 $\rho = -0.013$). Fidelity index: TP63+ = +0.003, KRT5+ = -0.213, shift = 0.216. (C) Five lineage-specific failure modes. Mode 1, phase-specific cliff (HNSC, fidelity-index shift 0.216); Mode 2, phase-uniform decoupling (GBM, EZH2 $\Delta\rho = 0.584$); Mode 3, progenitor lock-in (LIHC, AFP/ALB sign reversal across 6/6 commitment-panel members); Mode 4, TF-target bifurcation (PRAD, BRIP1 $\Delta\rho = 1.10$); Mode 5, quiescent evasion (AML, HSC-like cycling 1.3–3.8%, CD34 $\times$ BAX reversal $\Delta\rho = 0.477$, IDH2 not IDH1 is the myeloid-coupled isoform). All five confirmed. Classification criteria in §9.6. (D) Common root mechanism: partition failure → coordination dissolution → incoherent daughters → therapy resistance.

5.5 The Coupling Architecture: Predictions Tested in Part 6

The model's specific quantitative predictions about cancer concern the coupling architecture between model components, comparing cancer to normal tissue. Four predictions are stated here; Part 6 reports the full computational battery including per-cohort effect sizes, control comparisons, failures, and sensitivity analyses.

At the mark-installation level, NUMB's coupling to POLK, REV3L, and TOP2B should retain or strengthen in cancer, reflecting the marking program's continued operation even when regulatory coordination has dissolved. This prediction may be resolution-dependent: bulk transcriptomic data dilute epithelial-specific signals with non-epithelial cells, and the prediction may hold only when the analysis is restricted to the epithelial compartment.

At the polarity level, NUMB's coupling to the polarity network (PRKCZ, PRKCI, SCRIB, PARD3, DLG1) should approach zero in cancer, with a specified threshold of absolute ρ below 0.05 at protein level. This is the most direct test of the partition-failure thesis: if the polarity network that distributes fate determinants at mitosis is functionally intact, these couplings should be preserved.

At the cross-category level, at least two inter-layer correlations should reverse sign between normal and cancer tissue at statistically meaningful magnitudes, reflecting the dissolution of cross-category coordination that the partition system normally maintains.

The most architecturally specific prediction concerns the layer hierarchy itself: the asymmetric-division regulatory layer should lose internal coordination in cancer while the symmetric-division structural layer should be preserved. If confirmed, this would constitute the manuscript's

central empirical finding — the separation of model gene pairs into regulatory (disrupted), execution (preserved), and control (intermediate) categories.

A complementary prediction concerns whether the model axes most strongly associated with cancer progression operate as a coherent co-regulated module across cancer types. If the write/erase chemistry, fork protection, DDR sensing, and chromatin assembly components constitute a shared program rather than independent tissue-specific responses, their pairwise correlations should be consistently positive across diverse cancer types. Part 6 reports this test across thirteen TCGA PanCanAtlas cohorts.

5.6 Protein-Level Evidence: The MFN1 Linkage

The model's predictions are tested at mRNA level in transcriptomic cohorts and at protein level in CPTAC proteomic cohorts. The protein-level test most directly probes the broken-handoff hypothesis: NUMB and MFN1 should be positively coupled in normal tissue (consistent with the NUMB → MFN1 → OPA1 chain described in §3.2 that links cortical polarity to mitochondrial outer-membrane fusion to inner-membrane cristae configuration), and this coupling should weaken or reverse in cancer (consistent with breakage of the chain at its first link). The model's specified criterion at protein level is that the NUMB × MFN1 correlation in normal tissue exceeds the corresponding correlation in cancer cohorts, with the difference reproducible across multiple CPTAC studies.

The protein-level test is critical because the chain's mechanistic interpretation depends on the protein products being present and interacting, not merely on the transcripts being co-expressed. A transcriptomic correlation could in principle reflect shared upstream regulation rather than direct functional coupling. A protein-level correlation, particularly across multiple independent

cohorts, supports the model's interpretation that the relationship is mechanistic. The interpretation is further supported by the molecular cell biology of MFN1: as the principal mediator of mitochondrial outer membrane fusion, MFN1 activity is required for the inner-membrane fusion that OPA1 mediates (Chen et al., 2003; Cipolat et al., 2004), and loss of MFN1 function disrupts the cristae configuration that the partition model requires. The protein-level prediction therefore tests both the existence of the chain and its specific direction of function.

5.7 The Vogelstein Sequence as Consequence

The conventional account of colorectal cancer progression — the Vogelstein sequence (Fearon & Vogelstein, 1990) — describes a temporal accumulation of driver mutations that proceeds through definable stages. APC loss converts normal colonic epithelium to early adenoma. KRAS activation drives adenoma growth. TP53 deletion enables transition to carcinoma. SMAD4 loss permits metastatic progression. Each step is associated with a recurrent mutation at a specific gene, and the genes' identities have been used to construct prognostic schemes and to guide therapeutic stratification.

The model's interpretation of this sequence is that each step represents a different consequence of an underlying process in which partition failure accelerates mutation accumulation through missegregation of chromatin and mislicensing of the write/erase chemistry. APC normally functions as both a Wnt pathway negative regulator (Clevers & Nusse, 2012) and a cytoskeletal hub that participates in asymmetric mitotic spindle alignment in stem cells (Quyn et al., 2010; Yamashita et al., 2003). The spindle-orientation function operates in parallel with asymmetric centrosome inheritance: in *Drosophila* male germline stem cells, the mother centrosome remains anchored near the niche while the daughter centrosome migrates to the opposite side of the cell before spindle formation (Yamashita et al., 2007), establishing centrosome-level asymmetric

inheritance as an additional structural mechanism that, alongside polarity protein partitioning and asymmetric histone segregation, implements the architectural asymmetry the ARC model requires. APC loss simultaneously disinhibits Wnt signaling and disrupts spindle alignment, producing both elevated stem-cell-program transcription and reduced asymmetric segregation in a single mutational event. The combination — increased stem-program output without functional asymmetry — places the cell in a state in which subsequent mutations accumulate at elevated rates because the write/erase chemistry now operates without the partition that would normally constrain its outputs. APC loss is therefore, in the model's interpretation, a partition-affecting mutation, and its appearance early in the Vogelstein sequence is consistent with the model's account that loss of partition asymmetry accelerates the accumulation of the downstream mutations in the same sequence. Precedents for the missegregation and mislicensing components of this proposed mechanism are discussed in §8.3 alongside the model's broader relationship to existing cancer-biology models.

The subsequent steps in the Vogelstein sequence are, in this interpretation, downstream consequences of the permissive cellular state that partition failure creates. The KRAS promoter contains a G-quadruplex structure (Cogoi & Xodo, 2006), and the model proposes that disordered write/erase chemistry at this and other G4-containing regulatory addresses contributes to the elevated mutagenic environment. The specific coding-region mutations that activate KRAS (G12, G13, Q61) are distinct from the promoter G4-localized mutagenesis and may arise through conventional replication errors in the permissive, partition-disrupted state. TP53 accumulates the inactivating mutations that disable the cell's checkpoint response to the persistent damage that the partition-failed cell generates. SMAD4 loss disables a TGF- β pathway component whose function is required for normal differentiation (Massagué, 2012), locking the

partition-failed cell in a state where commitment cannot proceed regardless of niche signal. In the model's interpretation, each canonical driver mutation represents a step in which the partition-failed cell loses one further mechanism for either repairing itself or being eliminated by normal tissue defense systems.

This reinterpretation is testable. The model generates the testable hypothesis that tissues with measurable polarity disruption — quantified through $\text{NUMB} \times \text{polarity coupling}$ or through the MFN1-protein-level signature described in §5.6 — should show elevated mutation accumulation rates at G-quadruplex-containing regulatory addresses compared to polarity-intact tissues of the same lineage, with the magnitude of elevation correlating with the degree of polarity disruption. The signature should be detectable in normal tissue from individuals with known partition-disruption mutations (such as APC heterozygous carriers in familial adenomatous polyposis) at sites of subsequent tumor development. Quyn et al. (2010) demonstrated that asymmetric spindle orientation in intestinal stem-cell compartments — a marker of asymmetric segregation — is lost in APC-heterozygous tissue before any morphological evidence of dysplasia, establishing that partition disruption can be present in tissue that has not yet accumulated the full mutational complement of frank cancer. The model's prediction extends this finding to a specific genomic signature: G-quadruplex-containing promoters should show elevated mutation rates in APC-heterozygous tissue compared to matched APC-wildtype tissue from the same anatomical site, with the difference attributable to the disordered write/erase chemistry operating in partition-disrupted cells.

Partial empirical support for the broader prediction exists in pre-malignant tissue data. Aghili et al. (2014) identified somatic copy number alterations in apparently normal ovarian tissue and peripheral blood of 423 ovarian cancer patients, and found that these alterations overlap

significantly with G-quadruplex sequences (as well as with Alu repeats), with rates elevated in BRCA mutation carriers and in older individuals; some of the pan-cancer mutation hotspots in tumors were also hotspots of somatic alterations in normal tissue. The observation that mutation enrichment at G4-overlapping addresses in apparently normal tissue is measurable, and elevated in individuals with repair-axis disruption, is consistent with the model's prediction. The Aghili findings do not directly confirm the ARC mechanism, however: the alterations measured are copy-number changes rather than the point mutations that the write/erase chemistry primarily produces. The cohort is stratified by BRCA status rather than by polarity state. Alu overlap is co-equal with G4 overlap as an enrichment driver. The specific test the ARC model proposes — comparing point-mutation rates at G4 regulatory addresses between tissues stratified by polarity-state characterization — has not been performed and remains an open prediction.

The Vogelstein reinterpretation suggests a revision of the classical two-hit hypothesis (Knudson, 1971) that clarifies the relationship between partition failure and conventional tumor suppression. The model suggests that cancer may require the failure of two distinct protective mechanisms rather than two alleles of the same gene: fate control (the partition and differentiation system) and proliferation control (the p53/Rb/senescence checkpoint system). Evidence consistent with this dual-failure model includes the finding by Caussinus and Gonzalez (2005) that *Drosophila* neuroblasts carrying mutations in asymmetric division genes grew to over 100 times initial size, invaded neighboring tissues, and became immortal upon transplantation — with genomic instability appearing weeks after transformation rather than initiating it. A single defect in fate control was sufficient for tumorigenesis; proliferation-control failures emerged later as secondary adaptations. The logic extends to the Vogelstein sequence: APC loss disrupts

fate control, while subsequent TP53 loss eliminates proliferation control. The dual failure produces the cancer phenotype the model identifies.

5.8 Per-Cancer Predictions

Each failure mode generates a distinct therapeutic prediction tied to its defining architectural feature. The per-cancer predictions below illustrate the model's capacity to generate testable hypotheses from its architectural classification. The specific therapeutic strategies that follow from each mode — including drug classes, sequencing predictions, and stratification criteria — are consolidated in §7. Quantitative signatures supporting each mode assignment are reported in Part 6.

In colorectal cancer, the model classifies the failure as a Mode 3 variant — progenitor lock-in at the secretory-direction lineage program. The coupling signatures supporting this mode assignment are reported in §6.5; therapeutic predictions in §7.

In hepatocellular carcinoma, the model classifies the failure as the Mode 3 archetype — progenitor lock-in where the cell can run commitment chemistry but only at addresses the fetal progenitor program specifies. Therapeutic predictions for Mode 3 are developed in §7.

In pancreatic cancer, the model classifies the failure as a Mode 1 variant — phase-specific collapse of fidelity at the endocrine-commitment direction. The cohort's APOBEC3B upregulation is consistent with the Mode 1 architecture: partition failure producing replication stress that exposes single-stranded DNA, creating APOBEC3B-dependent mutagenesis as a secondary mutation source (Burns et al., 2013; Roberts et al., 2013). The Mode 1 prediction is that APOBEC3B inhibition should reduce the secondary mutation source without affecting the primary partition failure.

In head and neck squamous cell carcinoma, the model classifies the failure as the Mode 1 archetype, with the most clearly predicted phase-specific cliff. The model predicts that fidelity should collapse specifically at G2/M in the KRT5-expressing commitment-direction subpopulation while TP63-expressing stem-program cells retain balanced fidelity at all phases. Part 6 evaluates this prediction.

In prostate adenocarcinoma, the Mode 4 archetype — TF-target bifurcation where AR remains coupled to commitment chemistry while KLK3 is decoupled — is confirmed quantitatively in Part 6. Therapeutic predictions in §7.

In acute myeloid leukemia, the model classifies the failure as the Mode 5 archetype — quiescent evasion of selective purging. The prediction is that leukemic cells avoid the fitness selection that accompanies division by remaining in G0/G1. Gate-restoration strategies including IDH inhibitors and agents that force cycle re-entry are developed in §7.

In glioblastoma, the model classifies the failure as the Mode 2 archetype — phase-uniform lineage decoupling where the astrocytic-commitment program is decoupled from commitment chemistry at every cell-cycle phase while the stem program retains its coupling. Therapeutic predictions for Mode 2 are developed in §7.

The five failure modes and per-cancer predictions developed above are model-derived hypotheses whose clinical validation will require trials specifically designed to test mode-based stratification. The quantitative signatures supporting each mode assignment are reported in Part 6, and the classification of cancers outside the manuscript's panel into modes remains an open task.

Part 6: Computational Evidence

6.1 Approach: Pair Coupling as a Readout of Coordination

The computational evidence in this Part uses a specific analytical approach: pairwise Spearman correlation between mechanistically specified gene pairs, compared across conditions (normal versus cancer tissue, across cell types, along differentiation trajectories). This approach — differential co-expression analysis — has an established methodological literature (reviewed by Kayano et al., 2014; see also WGCNA, Langfelder & Horvath, 2008). What distinguishes the present application is that the pairs are specified by the model's architecture rather than discovered from data, and the changes in coupling are interpreted through the functional framework developed in Parts 2–5. The pair-coupling approach infers coordination state from changes in pairwise correlation: if two genes that the model identifies as jointly required for a specific function lose their coordination in a disease state while maintaining similar expression levels, this is evidence that the function is disrupted even though the components are present.

Three points about the correlational evidence that follows bear stating at the outset. Spearman correlations in transcriptomic data can demonstrate statistical association, reproducibility across independent datasets, and patterns consistent with mechanistic hypotheses. They cannot demonstrate causation, distinguish direct from indirect regulatory relationships, or exclude confounding by cell-type composition, shared upstream regulation, or technical artifacts. The protein-level CPTAC data and single-cell analyses partially address these limitations but do not eliminate them. The evidence presented here is therefore a test of consistency — whether the predicted coupling patterns exist in the data — rather than a proof of mechanism.

A point concerns the relationship between the pair classification and the statistical tests. The assignment of gene pairs to regulatory versus execution categories, and the correlation analyses

that test whether these categories behave differently, were developed iteratively during the model's construction — not specified in a registered analysis plan. The statistical tests therefore demonstrate the robustness and magnitude of the pattern rather than the probability it would have emerged from a fully pre-specified analysis. A permutation null test partially addresses this: randomly assigning 28 of 36 pairs to "regulatory" 10,000 times, zero of 10,000 permutations (empirical $p < 1 \times 10^{-4}$) produce a separation as large as the actual classification. An independent replication using 22 non-model regulatory pairs and 6 non-model execution pairs with zero gene overlap with the model panel reproduces the separation at matching effect sizes (independent regulatory $|\Delta p| = 0.170$ versus model 0.1815; independent execution $|\Delta p| = 0.065$ versus model 0.0396).

The primary transcriptomic source is the UCSC Xena TCGA HiSeqV2 repository, from which nine cancer types with adequate matched adjacent-normal tissue were selected: breast (1,097 tumor, 114 normal), kidney clear cell (533/72), lung adenocarcinoma (515/59), lung squamous (502/51), thyroid (505/59), liver (371/50), colon (286/41), head and neck (520/44), and stomach (415/35), totaling 4,744 tumor and 525 normal samples. mRNA expression values are $\log_2(\text{RSEM}+1)$ normalized. Correlations are computed per cohort as Spearman rank coefficients (chosen for robustness to outliers and non-normal distributions; Schober, Boer & Schwarte, 2018) and pooled across cohorts using the Fisher z-transform with $n-3$ inverse-variance weights (Fisher, 1921).

Protein-level validation uses the cBioPortal CPTAC protein quantification data from seven cohorts (breast, colon, glioblastoma, lung adenocarcinoma, lung squamous, pancreatic, and uterine), totaling 729 patients, providing protein-level validation independent of transcriptional regulation.

Single-cell resolution uses two RNA sequencing datasets from colorectal cancer: Lee et al. (2020; GSE132465; 63,690 cells from 23 Korean colorectal cancer patients yielding 33 tumor and normal pseudo-bulk samples — the Korean (SMC) cohort component of Lee et al.'s 91,103-cell, 29-patient (23 Korean + 6 Belgian) published atlas — with 8 patients contributing matched-normal samples) and Pelka et al. (2021; GSE178341; 370,115 cells [the post-QC analyzed subset of Pelka et al.'s 371,223-cell published atlas] from 62 patients, with 35 contributing matched normal epithelium). Single-cell data are analyzed as pseudo-bulks — sums per patient, tissue type, and cell compartment, restricted to pseudo-bulks containing at least 30 cells — followed by CPM normalization and log1p transformation.

Cross-species conservation testing uses the Haber et al. (2017; GSE92332) mouse intestinal atlas, specifically the 1,522-cell SMART-Seq2 subset of Haber et al.'s combined 53,193-cell atlas, across 11 cell types.

Three analytical lenses are applied to each prediction. The bulk TCGA comparison asks whether the correlation between two model genes differs systematically between matched tumor and normal tissue, quantified as $\Delta\rho = \rho_{\text{tumor}} - \rho_{\text{normal}}$. The CPTAC protein test asks whether the protein-level relationship in cancer cohorts matches the model's prediction. The single-cell pseudo-bulk test asks whether the between-patient regulatory coordination of model components differs between normal and tumor epithelium at cellular resolution. These three levels answer mechanistically distinct questions: bulk mRNA captures tissue-wide co-regulation including non-epithelial dilution; protein captures the post-translational reality most proximal to function; single-cell pseudo-bulk captures epithelial-specific between-patient coordination of model programs. The model's predictions specify expectations at each level, and discordances between levels are interpreted as informative rather than contradictory.

The four specified criteria are stated in §5.5 of Part 5 and evaluated in §6.3 below without modification to the quantitative thresholds. Where the directional logic of a criterion required correction after examination of the results, both the original and corrected formulations are reported.

The "190 specified correlation tests" referenced throughout this Part decomposes as follows: 61 model gene pairs evaluated across the bulk TCGA, CPTAC protein, and single-cell pseudo-bulk resolutions where each pair could be evaluated, plus 50 control pairs (housekeeping genes crossed with model genes) evaluated at the same resolutions where data permitted. Of the 61 model pairs, 36 met both data-coverage and classification criteria for inclusion in the two-class analysis reported in §6.2 (28 Class I regulatory coordination + 8 Class II execution partnership). One pair (BLM × WRN) was excluded as an ambiguous paralog pair. The remaining 24 pairs either did not have sufficient matched-normal coverage across the nine TCGA cohorts for confident Δp estimation or did not meet the molecular criteria that distinguish regulatory coordination from execution partnership, and are excluded from the class-level analysis without being post-hoc reassigned. The full pair-by-resolution table is preserved in the analysis output files listed in §9.7 for independent verification.

6.2 The Coordination Pattern in Cancer: Selective Dissolution

The 61 model gene pairs tested across nine TCGA cohorts with matched normal tissue separate into two behaviorally distinct classes. This separation is the manuscript's principal quantitative finding (full pair-level data in Table S2 and `supplementary_class_separation.csv`). The specified criteria predicted that regulatory coupling should weaken in cancer while execution coupling should be preserved; the robust statistical separation between the two classes emerged from the data and was not itself predicted in advance. Instead, the data reveal a pattern of surgically

selective change in gene co-regulation: the correlations that reflect regulatory coordination change substantially (22 of 28 weaken, 6 strengthen — both representing departure from the normal-tissue pattern), while those that reflect structural execution partnerships show minimal change.

Seventeen pairs show systematic coupling weakening or directional change in cancer. These are regulatory coordination pairs — relationships between genes whose co-regulation reflects a shared regulatory program rather than a structural obligation. The group includes the NUMB-to-polarity connections (NUMB with PRKCZ, PRKCI, DLG1, and PARD3), the death-to-nuclease coupling (BAX and BAK1 with APEX1), the base excision repair coordination axis (TET2 with TET3 and TDG, OGG1 with MUTYH, NEIL1 with XRCC1), the chromatin writer-eraser pairs (EZH2 with KDM6A and KDM6B, KDM1A with KDM5A), and the CDK6 regulatory connections (CDK6 with CCND1 and E2F1). Two pairs that satisfy regulatory rather than execution criteria (§6.2, Class II paragraph) are also included: FANCD2 × BRCA1 (which has $|\Delta\rho| = 0.107$, consistent with regulatory disruption rather than execution preservation) and MFN1 × OPA1 (whose $\Delta\rho = +0.223$ reflects compensatory strengthening of a sequential process dependency when the upstream regulatory layer dissolves — a Class I behavior with positive sign, discussed further below). Across the full Class I inventory, the mean absolute change in Spearman correlation between normal and tumor tissue is 0.182. Among the original fifteen weakening pairs, the direction is unanimous: all fifteen weaken in cancer, with zero exceptions. The full computational inventory tested 28 Class I pairs (the seventeen described here plus eleven additional pairs from the broader gene panel — including NUMB × SCRIB, NUMB × POLK, CDK6 × RB1, CDK4 × CDK6, BAX × OGG1, and others — that met data-coverage

criteria; the complete list is in Table S2). The statistics reported below use the full 28-pair inventory.

Eight pairs show preserved coupling with no consistent directional change. These are physical partnership pairs — relationships between gene products that form obligate complexes or operate in obligate sequence at the replication fork or in mitosis, as evaluated against five molecular criteria: (1) demonstrated direct physical binding, (2) shared subcellular compartment, (3) absence of partner-independent functions, (4) constitutive versus sequential association, and (5) whether transcriptomic co-expression primarily reflects stoichiometric complex formation or shared regulatory programming. The group includes the four CLSPN fork-protection connections (CLSPN with PCNA, ATR, CHEK1, and PARP1), the fork-coupled maintenance pair DNMT1 with PCNA, the Fanconi anemia ID complex (FANCI with FANCD2), the mitotic kinase partnership (PLK1 with AURKA), and the CDK4-associated proliferative pair (CDK4 with CCND1). Across these eight pairs, the mean absolute change is 0.040 — more than three-fold smaller than the regulatory class — and the direction is mixed: some weaken modestly, some strengthen modestly, with no consistent bias. Three pairs that might appear to be execution partnerships were evaluated against the five criteria above and classified differently: MFN1 × OPA1 (classified as Class I because OPA1 has MFN1-independent cristae remodeling functions and the two proteins reside in different mitochondrial membranes; Frezza et al., 2006; Rosselin et al., 2017), FANCD2 × BRCA1 (classified as Class I because both proteins have extensive partner-independent functions and their interaction is damage-induced rather than constitutive; Bunting et al., 2012), and BLM × WRN (excluded as an ambiguous paralog pair — both are RecQ family helicases causing different diseases with non-overlapping substrate specificities and no demonstrated direct heterodimeric complex).

Two additional pairs gain coupling in cancer: BRCA1 with RAD51 ($\Delta\rho = +0.174$) and TET1 with TET3 ($\Delta\rho = +0.175$). Both changes are positive, representing strengthening of co-expression in cancer. MFN1 $\times$ OPA1 ($\Delta\rho = +0.223$) is classified as Class I (see Class II paragraph above for molecular criteria). The same C1–C5 criteria classify the remaining two pairs. BRCA1 $\times$ RAD51 is Class I: the BRCA1-to-RAD51 loading pathway is indirect (BRCA1 $\rightarrow$ PALB2 $\rightarrow$ BRCA2 $\rightarrow$ RAD51), both proteins have extensive partner-independent functions, and their co-expression reflects shared cell-cycle and damage-response regulatory programming rather than stoichiometric coupling. The positive $\Delta\rho$ represents compensatory tightening — cancer cells increase HR activity as a survival mechanism, strengthening the BRCA1-to-RAD51 co-expression even as the regulatory architecture that should direct HR to appropriate targets dissolves. TET1 $\times$ TET3 is excluded as a paralogous enzyme pair — like BLM $\times$ WRN, they are independently functioning dioxygenases with no heterodimeric complex, different tissue distributions, and co-expression that reflects shared epigenetic regulatory programming rather than mutual dependency. The positive $\Delta\rho$ likely reflects cancer cells' increased reliance on TET-mediated demethylation for epigenetic plasticity. All three strengthening pairs are thus resolved: two as Class I compensatory strengthening (MFN1 $\times$ OPA1, BRCA1 $\times$ RAD51) and one excluded as an ambiguous paralog (TET1 $\times$ TET3).

The statistical separation between Class I and Class II is robust. A Mann-Whitney U test comparing the absolute $\Delta\rho$ distributions of the literature-grounded classification (28 Class I versus 8 Class II, with MFN1 $\times$ OPA1 and FANCD2 $\times$ BRCA1 classified as Class I on the molecular criteria described in the Class II paragraph above, and BLM $\times$ WRN as an ambiguous paralog pair) yields $p = 8.53 \times 10^{-6}$. When the analysis is restricted to core regulatory pairs — excluding the distal periphery where disruption signal attenuates with regulatory chain distance

(see below) — the separation tightens to $p = 3.22 \times 10^{-7}$ (25 Class I versus 8 Class II). A bootstrap analysis with 10,000 resamples confirms the mean absolute Δp difference excludes zero across all four classification scenarios tested (conservative, strict, core-only, original). A leave-one-out sensitivity analysis confirms that no single pair drives the finding: removing any individual pair from either class produces $p \leq 1.10 \times 10^{-5}$, and simultaneously removing the three highest-impact Class I pairs (CDK6 $\times$ E2F1, NUMB $\times$ PRKCZ, NUMB $\times$ PARD3) still yields $p = 1.86 \times 10^{-5}$. The separation is not complete at the rank level: six Class I pairs have absolute Δp values below the Class II maximum of 0.107 (FANCI $\times$ FANCD2). These six are all distal regulatory chain links (NUMB-to-mitochondrial or CDK-to-downstream pairs) whose low disruption reflects signal attenuation with regulatory distance — a finding discussed below as evidence for the directional architecture of the regulatory dissolution.

Two additional comparisons anchor this finding against the null distribution. Class I pairs show significantly larger absolute Δp than 50 control pairs (housekeeping genes crossed with model genes; Mann-Whitney $p = 4 \times 10^{-6}$), confirming that the regulatory disruption exceeds what random gene pairs experience. Class II pairs show significantly smaller absolute Δp than the same controls ($p = 0.003$), confirming that the physical partnerships are not merely undisrupted but are more stable than random pairs — they are actively maintained. The combined result — regulatory pairs more disrupted than random, physical partnerships more preserved than random, and the two classes robustly separated from each other — would be unlikely to arise from post-hoc classification of arbitrary gene pairs.

The two-class pattern maps directly onto the functional classification defined in Part 2. The Class I pairs whose coupling is disrupted in cancer are primarily regulatory coordination pairs — relationships between gene products that participate in the coordinated programs the partition

system organizes. NUMB × polarity proteins, CDK6 × licensing partners, and BER coordination pairs are all regulatory coordination components: they are not structurally obligate (the gene products do not form constitutive physical complexes), and their co-regulation reflects a shared regulatory program rather than a structural necessity. FANCD2 × BRCA1 and MFN1 × OPA1 are classified as Class I because both have extensive partner-independent functions and their transcriptomic co-expression reflects shared regulatory programming rather than stoichiometric complex requirements. The eight Class II pairs whose coupling is preserved are the most stringently validated execution partnerships — relationships between gene products that form obligate complexes or operate in constitutive obligate sequence at the replication fork or in mitosis. CLSPN × PCNA, FANCI × FANCD2, PLK1 × AURKA, and DNMT1 × PCNA all score 4/5 or 5/5 on the five molecular criteria for constitutive physical partnership. CCND1 appears in both classes because it activates both CDK4 and CDK6: paired with CDK4 it forms a stoichiometric kinase-cyclin complex (Class II, preserved), while paired with CDK6 it participates in a broader regulatory coordination that includes CDK6's kinase-independent chromatin functions (Class I, weakening).

To address potential circularity in this mapping, four independent considerations apply. The functional assignments are derived from molecular-biological criteria (whether the gene products form obligate physical complexes, whether their co-regulation reflects shared transcriptional control versus structural coupling) that are independent of the transcriptomic data — the C1–C5 criteria and their per-pair scores (Table S3) can be evaluated by any reviewer using the cited literature without reference to Δp values. The unanimity of the Class I direction (15 of 15 named weakening pairs, zero exceptions) is not a property the classification criteria could have engineered: the criteria specify which pairs are regulatory but do not specify which direction the

regulatory disruption should run, so the directional consistency is an empirical finding independent of the classification method. The sensitivity analysis (§6.2) demonstrates that the statistical separation holds across four alternative classification scenarios, including one with broader Class II membership and one with a stricter threshold, confirming that reasonable reclassification choices do not drive the finding. The independent control analysis using 22 regulatory and 6 execution pairs with zero gene overlap with the model panel reproduces the same pattern (§6.2), confirming that the separation is a property of gene-pair biology rather than of this specific panel. The class structure of the data is consistent with the functional classification: an architectural classification based on molecular criteria predicts a quantitative separation in transcriptomic data, with robust separation between the predicted classes and a directional consistency that the classification did not specify in advance.

An alternative explanation must be considered and addressed. Could the two-class pattern arise from general regulatory simplification in cancer — a non-specific loss of regulatory coordination that would affect all regulatory pairs regardless of model membership? If cancer merely simplified gene regulation, then any random set of regulatory pairs should show comparable disruption, and any random set of structural partnerships should show comparable preservation. The control analysis directly tests this alternative. The fifty control pairs (housekeeping genes crossed with model genes) show a mean absolute Δp of 0.124 — intermediate between Class I (0.1815) and Class II (0.0396). Model regulatory pairs are significantly MORE disrupted than random pairs, and model physical partnerships are significantly MORE preserved than random pairs. The selectivity is therefore not general regulatory simplification — it is specifically targeted at the partition-system components that the model identifies, disrupting the regulatory coordination layer more than random and preserving the execution layer more than random. A

generic loss of regulatory complexity would produce disruption in the Class I range for all regulatory pairs; instead, the model's Class I pairs are selectively more disrupted than controls, and the model's Class II pairs are selectively more preserved than controls. Both directions of selectivity must be explained, and the model's functional classification explains both.

The class separation also exhibits a structural feature that was not predicted in advance but that the model's architecture explains. The magnitude of regulatory disruption attenuates with regulatory chain distance from the core polarity node. NUMB $\times$ PARD3 (direct polarity partner, $|\Delta\rho| = 0.359$) shows the strongest disruption. NUMB $\times$ DLG1 (extended polarity scaffold, $|\Delta\rho| = 0.268$) and NUMB $\times$ PRKCZ (direct polarity partner, $|\Delta\rho| = 0.250$) are in the same tier. NUMB $\times$ PRKCI (extended polarity, $|\Delta\rho| = 0.189$) is intermediate. The cross-compartment links to organelles show weaker disruption: NUMB $\times$ MFN1 (outer mitochondrial membrane, $|\Delta\rho| = 0.132$), NUMB $\times$ TOP2B (nucleus, $|\Delta\rho| = 0.109$), and NUMB $\times$ OPA1 (the most distal link, to the inner mitochondrial membrane, $|\Delta\rho| = 0.093$). The attenuation spans approximately four-fold from 0.359 at the cortex to 0.093 at the inner mitochondrial membrane, with the cortical polarity partners showing the strongest disruption and the cross-compartment links the weakest. This distance-dependent attenuation is consistent with the model's account that partition failure originates at the cortical polarity level and propagates outward through the organelle inheritance chain. The six Class I pairs that overlap with the Class II range in absolute $\Delta\rho$ (below the Class II maximum of 0.107 for FANCI $\times$ FANCD2) include the most distal NUMB cross-compartment links — they are not misclassified execution pairs but rather regulatory pairs where the disruption signal has attenuated to a magnitude comparable to the execution-pair floor.

A sensitivity analysis confirms that the class separation is robust to reasonable classification choices. Four scenarios were tested: (1) the conservative classification (28 Class I versus 8 Class

II, $p = 8.53 \times 10^{-6}$), (2) a strict classification excluding three marginal Class II pairs that score 2–3/5 on the molecular criteria (28 versus 6, $p = 3.42 \times 10^{-7}$), (3) a core-only analysis excluding the distal Class I periphery (20 versus 8, $p = 3.22 \times 10^{-7}$), and (4) an alternative classification with broader Class II membership (22 versus 8, $p = 8.53 \times 10^{-6}$). The finding is significant under all four scenarios. The full C1–C5 scoring for each pair, with literature citations for every criterion score, is provided in Table S3.

An independent control analysis tests whether the Class I/II pattern reflects the model's specific gene selection or a general property of any comparably classified gene pairs. Twenty-two independent regulatory pairs (kinase-substrate cascades and transcription factor programs: EGFR $\times$ SOS1, BRAF $\times$ MAP2K1, AKT1 $\times$ MTOR, JAK2 $\times$ STAT3, TP53 $\times$ CDKN1A, NOTCH1 $\times$ HES1, and others — no gene overlap with the model panel) were computed on the BRCA TCGA cohort ($n = 1,097$ tumor, $n = 114$ normal). Their mean absolute Δp is 0.170, closely matching the model's Class I mean of 0.1815. Six independent replication-essential execution pairs (MCM2 $\times$ MCM5, MCM3 $\times$ MCM7, RFC2 $\times$ RFC5, POLR2A $\times$ POLR2B, NDC80 $\times$ NUF2, SMC1A $\times$ SMC3) show mean absolute Δp of 0.065, closely matching the model's Class II mean of 0.0396. However, nine independent physical complexes involved in tissue biology rather than replication — proteasome subunits, nuclear pore components, adherens junction proteins, and chaperonin pairs — show mean absolute Δp of 0.315, significantly higher than the replication-essential pairs ($p = 8 \times 10^{-4}$, Mann-Whitney U, one-sided). The implication is that the model's Class II specificity is biological, not methodological: cancer does not preserve all physical complexes but specifically preserves the replication and mitotic execution machinery it requires to proliferate, while tissue-remodeling-associated complexes are disrupted according to the tissue's cancer biology.

6.3 Pre-Specified Criteria Assessment

The four specified criteria stated in §5.5 are evaluated here against the full computational results. Quantitative thresholds are not modified after testing. Where the directional logic of a criterion required correction, both formulations are reported.

The mark-installation criterion predicted that NUMB's coupling to the mark-installation arms (POLK, REV3L, TOP2B) should be stronger in cancer than in normal tissue. At bulk mRNA level across nine TCGA cohorts, NUMB $\times$ POLK changes from $\rho = +0.264$ in normal to $+0.307$ in tumor ($\Delta\rho = +0.043$), and NUMB $\times$ REV3L from $+0.240$ to $+0.280$ ($\Delta\rho = +0.040$) — both essentially unchanged. NUMB $\times$ TOP2B weakens from $+0.266$ to $+0.157$. The prediction is not supported at bulk resolution. However, in the Lee 2020 single-cell analysis restricted to epithelial cells, NUMB $\times$ POLK strengthens from $\rho = +0.429$ in normal epithelial pseudo-bulks to $+0.627$ in tumor epithelial pseudo-bulks. The criterion is partially met at single-cell resolution: the marking-arm coupling does strengthen, but only when the analysis is restricted to epithelial cells, consistent with non-epithelial dilution masking the effect in bulk.

The polarity-dissolution criterion predicted that NUMB's coupling to the polarity network should approach zero in cancer, with the specified threshold of absolute ρ below 0.05. At protein level across seven CPTAC cohorts, three of three aPKC-axis pairs meet this threshold: NUMB $\times$ PRKCZ at absolute $\rho = 0.005$, NUMB $\times$ PRKCI at 0.041, and NUMB $\times$ SCRIB at 0.013. At mRNA level, the polarity coupling weakens substantially: NUMB $\times$ PRKCZ drops from $\rho = +0.263$ in normal to $+0.013$ in tumor (a 95 percent reduction), and NUMB $\times$ PARD3 undergoes a sign reversal from $+0.338$ to -0.021 , exceeding the control 95th percentile at absolute $\Delta\rho = 0.359$. In the Pelka 2021 single-cell data with 35 matched normal pseudo-bulks, NUMB $\times$ PRKCI shows the largest decoupling of any pair in the entire dataset ($\Delta\rho = -0.415$;

independently validated at $\Delta\rho = -0.33$ using partial byte-plane extraction, confirming direction and approximate magnitude; see §9.5). This criterion is supported across all three data sources and both molecular levels.

The sign-reversal criterion predicted at least two inter-layer sign reversals between normal and cancer tissue, at statistically meaningful magnitudes. At bulk mRNA level, one clear reversal is observed (NUMB $\times$ PARD3, from +0.338 to -0.021). In the Lee 2020 single-cell pseudo-bulks, three sign reversals exceed absolute $\Delta\rho$ of 0.3. In the Pelka 2021 pseudo-bulks, two reversals exceed absolute $\Delta\rho$ of 0.3: CDK6 $\times$ E2F1 (from +0.174 to -0.191) and EZH2 $\times$ KDM6A (from -0.092 to +0.289). For context, the control null distribution has its 95th percentile at $|\Delta\rho| = 0.310$, meaning fewer than 5% of random gene pairs would exceed $|\Delta\rho| = 0.3$ under the null; fewer still would be expected to exceed 0.5. The three reversals above 0.3 in Lee 2020 are at the margin of the null threshold. The two reversals above 0.3 in Pelka are at the margin of the null threshold and are informative primarily as directional confirmation of the Lee 2020 finding rather than as independently significant observations. The specified threshold of at least two reversals is met in both single-cell datasets and approached in the bulk data. The criterion is supported, with the single-cell resolution providing the strongest evidence.

The layer-decoupling criterion predicted that the asymmetric and symmetric gene-regulation layers should decouple in cancer. The within-asymmetric-layer coupling (mean absolute Spearman correlation among partition-system pairs) decreases from 0.277 in normal to 0.220 in tumor, a reduction of 21 percent. The within-symmetric-layer coupling (among fork-protection and repair-execution pairs) is unchanged: 0.526 in normal, 0.534 in tumor. Cross-layer coupling decreases from 0.143 to 0.114, a reduction of 20 percent. All three components of the prediction are supported: the asymmetric layer weakens internally, the symmetric layer is maintained, and

cross-layer coordination decreases as a consequence of the one-sided dissolution. The two-class separation described in §6.2 is the pair-level expression of this layer-level prediction: Class I pairs are the asymmetric layer's internal coordination; Class II pairs are the symmetric layer's structural partnerships; and the robust statistical separation between the two classes is the quantitative demonstration that one layer collapses while the other holds.

In summary, three of four specified criteria are supported (polarity dissolution, sign reversals, layer decoupling), and one is partially supported with a resolution qualification (marking-arm coupling is epithelial-specific). No criterion is contradicted — the marking-arm prediction is not wrong, the resolution qualification identifies a more precise and testable formulation (epithelial-specific, resolution-dependent).

6.4 Mechanistic Findings: BER Collapse, Fork Protection, and Write/Erase Subsystems

Several specific mechanistic findings emerge from the battery that were not anticipated in the specified criteria but are informative for the model's broader claims.

The CDK6 regulatory axis is selectively disrupted in cancer while CDK4 is preserved, consistent with the Class I/II distinction introduced in §2.2. $CDK6 \times CCND1$ shows weaker coupling in tumor ($\rho = -0.0568$) compared to normal tissue ($\rho = +0.2269$, $\Delta\rho = -0.2837$), and $CDK6 \times E2F1$ weakens from $+0.1095$ to -0.1344 ($\Delta\rho = -0.2439$). $CDK4 \times CCND1$, by contrast, is unchanged at $\Delta\rho = +0.0069$. In the Pelka single-cell data, $CDK6 \times E2F1$ shows a full sign reversal (from $+0.174$ to -0.191 , $\Delta\rho = -0.365$), independently confirming the TCGA bulk finding at cellular resolution. The differential behavior is consistent with CDK6's established kinase-independent transcriptional functions (Kollmann et al., 2013) placing its regulatory relationships in a different vulnerability class from the structural kinase-substrate coupling of CDK4 with cyclin D1.

The CLSPN fork-protection module is robustly preserved across all conditions tested. The four CLSPN pairs (with PCNA, ATR, CHEK1, and PARP1) show absolute $\Delta\rho$ below 0.08 in every case, well within the control null distribution. In the Pelka single-cell data, CLSPN $\times$ PCNA yields $\Delta\rho = +0.068$, independently confirming preservation at cellular resolution. The model interprets this through the established baseline that parental histones segregate to both daughter strands during DNA replication with a slight leading-strand bias (Petryk et al., 2018) and that MCM2-dependent and CLASPIN-dependent histone-handling activities act to balance this distribution: in fission yeast, Mrc1/CLASPIN functions as a central coordinator of symmetric parental histone inheritance (Charlton et al., 2024), with mutations in its connector domain disrupting segregation to the lagging strand and producing asymmetric loss of H3K9me-mediated silencing. The stability of the CLSPN–CHEK1 complex is itself maintained by DNA-PKcs (Lin et al., 2014), connecting the fork-protection module to the DNA damage kinase axis. Symmetric histone inheritance may be selectively advantageous in cancer (both daughters receiving equivalent chromatin), so the fork-protection module may be preserved because it serves the functional requirements of undifferentiated self-renewal in cancer. The preservation could also reflect the structural necessity of fork protection for any dividing cell, cancerous or not.

The base excision repair coordination network collapses as a unit (for BER pathway coordination, see Krokan & Bjørås, 2013; Beard et al., 2019): TET2 $\times$ TET3 ($\Delta\rho = -0.230$), TET2 $\times$ TDG ($\Delta\rho = -0.201$), OGG1 $\times$ MUTYH ($\Delta\rho = -0.208$), and NEIL1 $\times$ XRCC1 ($\Delta\rho = -0.240$) all weaken by similar magnitudes, consistent with coordinated loss of the BER decision-making program. The writer-eraser pairs EZH2 $\times$ KDM6A ($\Delta\rho = -0.193$) and KDM1A $\times$ KDM5A ($\Delta\rho = -0.190$) show comparable disruption. The insulator and AT arm components,

tested in the Xena-based BRCA tumor-versus-normal comparison ($n = 1,097$ tumor, 114 normal), show the same pattern: NIPBL $\times$ RIF1 (cohesin loading, $\Delta\rho = -0.188$), TOP2B $\times$ EP300 (AT arm, $\Delta\rho = -0.171$), and CTCF $\times$ TET2 (insulator regulatory, $\Delta\rho = -0.234$) all weaken, while the fork-coupled maintenance methyltransferase DNMT1 $\times$ PCNA is unchanged ($\Delta\rho = +0.015$) and the cohesin structural pair STAG1 $\times$ RIF1 is preserved ($\Delta\rho = +0.031$).

This pattern is consistent with the two-class distinction at a finer grain: the regulatory coordination that determines which base excision pathway to activate (the "decision" about whether to excise or retain a modification) is disrupted, while the structural coupling of the methylation maintenance machinery to the replication fork (the "execution" of maintenance methylation) is preserved. Cancer retains the ability to methylate during replication; it loses the ability to coordinate the erasure decisions that would produce lineage-appropriate demethylation. At the three-dimensional genome level, the cohesin complex (structural) is intact, but the regulatory program that determines where boundaries are established and which enhancer-promoter pairs interact is selectively dissolved.

The results in §6.2–6.4 are consistent with the model's prediction of selective regulatory dissolution: the two-class pattern separates with statistical robustness, the BER coordination network collapses as a unit, and the insulator arm splits between preserved structural components and disrupted regulatory components.

6.5 The Model Across Lineages

The Pelka 2021 CRC atlas provides the strongest replication cohort in the battery: 370,115 cells (the post-QC analyzed subset of Pelka et al.'s 371,223-cell published atlas) from 62 patients, with 35 patients contributing matched tumor and normal epithelial pseudo-bulks — more than four

times the normal sample size of the Lee 2020 dataset. Extraction of the 49-gene model panel from this dataset required engineering a custom pipeline because the deposited h5 file stores its 764-million-entry sparse matrix as single gzip-compressed chunks with a shuffle filter, making conventional h5py access impossible within available memory. The workaround — raw zlib stream decompression to disk followed by numpy-vectorized unshuffle via memory-mapped file access — is documented in the Methods (Part 9) and produces results identical to standard sparse-matrix extraction on test subsets.

The Pelka results replicate the key TCGA findings at single-cell resolution. $\text{NUMB} \times \text{PRKCI}$ shows $\Delta\rho = -0.415$, the largest decoupling of any pair tested in any dataset. $\text{CDK6} \times \text{E2F1}$ shows a sign reversal ($\Delta\rho = -0.365$). $\text{CLSPN} \times \text{PCNA}$ is preserved ($\Delta\rho = +0.068$). Two sign reversals exceed absolute $\Delta\rho$ of 0.3, meeting Criterion 3 in an independent cohort. The overall pattern of regulatory disruption with execution preservation replicates across datasets despite the substantial differences between them: Lee 2020 sampled 23 Korean patients (the SMC cohort, 63,690 cells), while Pelka 2021 sampled 62 American and European patients with an atlas approximately six times larger (370,115 cells); both used 10x Chromium 3' sequencing. The concordance between Pelka and TCGA bulk mRNA on direction of disruption for key pairs further strengthens the finding's robustness.

Stratification of the Pelka tumor pseudo-bulks by mismatch repair status (13 MMRd, 12 MMRp, by immunohistochemistry) reveals that the two CRC subtypes show the same overall degree of model disruption but different topologies of disruption. The most striking differential is $\text{FANCI} \times \text{FANCD2}$: preserved in MMRd tumors ($\rho = +0.544$, comparable to normal at $+0.457$) but disrupted in MMRp tumors ($\rho = -0.098$). This asymmetry has a direct biological interpretation — MMRd tumors have lost mismatch repair and depend on the Fanconi anemia pathway as a

compensatory DNA-damage-tolerance mechanism, so they cannot afford to lose FA complex coordination. The execution machinery (PLK1 $\times$ AURKA, CLSPN $\times$ PCNA) is similarly better preserved in MMRd than in MMRp tumors, suggesting that MMRd tumors maintain more architectural integrity in the execution layer while suffering a specific molecular defect (MMR loss), whereas MMRp tumors have broader architectural corruption despite retaining MMR function. The therapeutic prediction that follows from this differential is developed in §7.1.

The mouse cross-species validation uses the Haber et al. (2017) intestinal atlas to test whether the co-regulation patterns the model identifies are conserved across species. Of 32 testable gene pairs with mouse orthologs in the SMART-Seq2 atlas, 26 show the same sign of Spearman correlation in mouse intestinal cells as in human normal tissue — an 81 percent conservation rate. The conservation follows the model's predicted hierarchy: Class II (execution) pairs show 87.5 percent conservation (seven of eight), BER and chromatin coordination pairs show high conservation, and the NUMB partition pairs replicate at eight of ten (with Numb $\times$ Scrib and Numb $\times$ Polk discordant). The CDK axis shows the weakest conservation, with the Cdk4 $\times$ Cdk6 anti-correlation observed in human bulk tissue not replicating in mouse intestinal single cells and Cdk6 $\times$ Rb1 reversing sign. The cross-species analysis supports the interpretation that the patterns the model tests reflect conserved mammalian biology rather than human-specific or dataset-specific artifacts, with the physically obligate partnerships being the most deeply conserved and the regulatory coordination patterns being largely but not universally conserved.

6.5.1 HNSC Single-Cell Architecture: TP63 versus KRT5 Subpopulations

The Puram et al. (2017) HNSC dataset (GSE103322; 2,215 cancer cells from head and neck squamous carcinoma tumors) provides a within-tumor test of the model's architectural prediction at single-cell resolution. The squamous lineage offers a natural contrast between TP63-

expressing cells (the squamous stem-program marker) and KRT5-expressing cells (commitment-direction marker), allowing the model's prediction — that stem-program cells retain balanced fidelity while failed-commit cells show TLS-dominant bypass with fitness-selection decoupling — to be tested within the same tumors.

TP63-expressing cancer cells retain balanced coupling to both fidelity-providing helicases (BLM $\rho = +0.067$, WRN $\rho = +0.062$) and TLS bypass polymerases (POLK $\rho = +0.061$, REV3L $\rho = +0.022$), producing a fidelity index (mean helicase coupling minus mean TLS coupling) of $+0.003$ — essentially zero, indicating no systematic bias toward either resolution pathway. The fitness-selection machinery is engaged: NUMB $\rho = +0.076$, APEX1 $\rho = +0.120$, BAX $\rho = +0.069$, BAK1 $\rho = +0.075$ (all $p < 0.001$; $n = 2,215$). TP63+ cells couple positively to both the resolution and purging components the model identifies as partition-system outputs.

KRT5-expressing cells show the opposite architecture. All four fidelity-providing helicases are anti-correlated (BLM $\rho = -0.083$, WRN $\rho = -0.020$, BRIP1 $\rho = -0.203$, HLTf $\rho = -0.190$), while all three TLS polymerases are positively correlated (REV1 $\rho = +0.078$, POLK $\rho = +0.092$, REV3L $\rho = +0.097$). The fidelity index is -0.213 — a shift of 0.216 units from the TP63+ population, reflecting systematic exclusion of helicase-mediated fidelity with simultaneous TLS engagement. The fitness-selection panel shows selective decoupling: NUMB $\rho = -0.004$ ($p = 0.86$) and APEX1 $\rho = -0.013$ ($p = 0.54$) are both indistinguishable from zero, while BAK1 remains strongly positive ($\rho = +0.305$, $p = 8.3 \times 10^{-49}$) — consistent with retained intrinsic apoptotic priming but dissolved partition-dependent fitness selection. $\text{KRT5} \times \text{BRCA2}$ is -0.137 ($p = 1.1 \times 10^{-10}$), confirming the commitment-direction decoupling from HR machinery that the Mode 1 prediction requires.

The combined pattern defines two architecturally distinct coupling profiles coexisting within the same tumors: TP63-correlated genes show balanced fidelity and engaged purging, while KRT5-correlated genes show TLS-dominant bypass, fidelity exclusion, and dissolved fitness selection. TP63 and KRT5 are positively correlated ($\rho = +0.170$, $p = 7.7 \times 10^{-16}$) and co-expressed in approximately 28% of cancer cells by median split, so the two markers do not define mutually exclusive subpopulations. The coupling contrasts reported above reflect what each marker tracks independently across the full cancer-cell population — the architectural signatures associated with the stem program (TP63) versus the commitment direction (KRT5) — rather than a comparison between separated subsets. The fidelity-index shift (0.216 units) and the NUMB/APEX1 decoupling are properties of these coupling profiles, not of defined cell populations.

6.5.2 AML Quiescent Evasion: Hierarchy-Resolved Single-Cell Analysis

The van Galen et al. (2019) AML dataset (GSE116256; 8,571 malignant + 1,439 normal cells from five AML patients at diagnosis) classifies malignant cells into a hierarchy resembling normal hematopoiesis: HSC-like ($n = 855$), Progenitor-like ($n = 1,806$), GMP-like ($n = 1,282$), ProMono-like ($n = 1,161$), Mono-like ($n = 2,000$), and cDC-like ($n = 1,467$). The original bulk-level analysis — pooling all malignant cells — found weak CEBPA $\times$ commitment coupling ($\rho = +0.059$) and no phase-specific enrichment, leading to the assessment that Mode 5 was not confirmed. Stratification by malignant hierarchy reveals a pattern the bulk analysis obscured.

Malignant HSC-like cells are profoundly quiescent at the per-patient level: AML328 (2.1% HSC-like cycling), AML419A (3.8%), AML707B (1.3%), AML921A (2.0%); aggregate HSC-like cycling = 2.1% across the four patients with sufficient HSC-like cell counts (``aml_singlecell_cycling.json:aggregate_hsc_cycling_pct``). The corresponding non-HSC-like

malignant cycling fractions (deposited ``pct_cycling_mal``) are AML328 4.8%, AML419A 4.0%, AML707B 18.5%, AML921A 14.8%, yielding per-patient HSC-to-non-HSC fold-changes of 2.29 $\times$, 1.05 $\times$, 14.23 $\times$, and 7.4 $\times$. The fold-change variance is wide: three of four patients show substantial HSC-to-non-HSC cycling separation (2.29 $\times$ to 14.23 $\times$), while AML419A shows essentially no separation (1.05 $\times$) — suggesting that quiescent evasion at the per-patient level is not uniformly penetrant. The mean fold-change across the four patients (6.24 $\times$) and the consistent HSC-compartment quiescence (1.3–3.8% across all four patients, mean 2.3%) support the population-level Mode 5 signature, but AML419A demonstrates that the per-patient effect can attenuate to within measurement noise. Per-cluster cycling-fraction aggregates across the broader malignant hierarchy (Progenitor-like, GMP-like, ProMono-like, Mono-like, cDC-like, and matched normal compartments) are pending re-verification against deposited single-cell counts and are not reported here.

At single-cell resolution, CEBPA and commitment-chemistry expression show hierarchy-dependent coupling, with strong anti-correlation for BRCA2 and EZH2 in the stem compartment: $\text{CEBPA} \times \text{BRCA2} \rho = -0.435$ (stem), -0.034 (progenitor), $+0.039$ (differentiated-like); $\text{CEBPA} \times \text{EZH2} \rho = -0.686, -0.440, +0.064$; $\text{CEBPA} \times \text{BAX} \rho = +0.699, +0.581, +0.009$. The $\text{CEBPA} \times \text{BRCA2}$ and $\text{CEBPA} \times \text{EZH2}$ anti-correlations indicate that cells expressing the myeloid commitment TF and cells expressing DDR-commitment machinery are segregated into different subpopulations in the stem compartment. $\text{CEBPA} \times \text{BAX}$ shows the opposite direction ($+0.699$ in stem), indicating that BAX co-expression tracks with CEBPA rather than opposing it — the apoptotic gatekeeper is engaged in differentiating cells, not excluded from them. The anti-coupling magnitude attenuates from stem to differentiated-like for BRCA2 and EZH2, consistent with partial convergence as cells progress through the hierarchy. A caution: correlations in sparse

single-cell UMI data are normalization-dependent, and the hierarchy-level cycling difference is the more robust finding.

At single-cell resolution with log1p normalization, $CDK6 \times CCNB1$ $\rho = +0.077$ and $CDK4 \times CCNB1$ $\rho = +0.083$ ($\Delta = 0.006$) — the CDK6 decoupling observed at bulk level (§6.4) is not resolved in the sparse UMI data. $CD34 \times BRCA2$ $\rho = -0.023$ and $MPO \times BRCA2$ $\rho = +0.071$; both near zero, consistent with the hierarchy-level cycling differences being the more robust single-cell finding.

The Mode 5 prediction is supported by convergent evidence at two levels: hierarchy-resolved single-cell analysis reveals the quiescent evasion architecture in the malignant HSC-like compartment (2.1% cycling, consistent across patients), while TCGA bulk analysis reveals that the stem marker CD34 is decoupled from commitment chemistry with fidelity-dominant resolution in the committed MPO⁺ fraction. The $CD34 \times BAX$ reversal ($\Delta\rho = 0.477$), the $CEBPA \times POLK$ anti-coupling ($\rho = -0.434$), and the IDH2 (not IDH1) tissue-specific isoform coupling together define a mode that is architecturally distinct from Modes 1–4. The initial bulk analysis that pooled across the hierarchy failed to detect this pattern because the stem and progenitor compartments have opposing coupling architectures that average to near-zero.

TCGA LAML bulk data ($n = 173$) provides a complementary test using the same CD34/MPO lineage contrast methodology that confirmed Modes 2–4. CD34 (the stem marker) is decoupled from commitment chemistry — near-zero coupling across the panel ($BRCA2$ $\rho = +0.077$, $EZH2$ $\rho = +0.063$, $PCNA$ $\rho = -0.056$, $OGG1$ $\rho = -0.048$; all NS) — and strongly anti-coupled with BAX ($\rho = -0.325$, $p = 1.3 \times 10^{-5}$), indicating that the stem program is protected from apoptotic purging at bulk level. MPO (the myeloid commitment marker) shows the opposite: positive

coupling to EZH2 ($\rho = +0.344$, $p = 3.7 \times 10^{-6}$), BAX ($\rho = +0.152$), APEX1 ($\rho = +0.322$), WRN ($\rho = +0.181$), and REV1 ($\rho = +0.203$), with strong anti-coupling to POLK ($\rho = -0.373$, $p = 4.4 \times 10^{-7}$). The MPO coupling profile is fidelity-dominant: the committed myeloid fraction engages helicases and excludes error-prone bypass. CEBPA shows a similar fidelity-dominant pattern (CEBPA $\times$ POLK $\rho = -0.434$, $p = 2.5 \times 10^{-9}$; CEBPA $\times$ BAX $\rho = +0.374$, CEBPA $\times$ OGG1 $\rho = +0.320$). The CD34 $\times$ BAX reversal (-0.325 vs MPO $\times$ BAX $+0.152$, $\Delta = 0.477$) is the largest single Mode 5 effect, comparable in magnitude to Mode 3 and Mode 4 effects.

CDK6 at bulk level is decoupled from CEBPA ($\rho = -0.027$) while CDK4 couples positively (CDK4 $\times$ CEBPA $\rho = +0.319$, $\Delta = 0.346$), and CDK6 $\times$ CCNB1 $= +0.104$ vs CDK4 $\times$ CCNB1 $= +0.373$, confirming at bulk level the CDK6 dissociation observed in the single-cell analysis. IDH1 $\times$ CDK6 $= -0.269$ in LAML — the opposite sign from GBM ($+0.528$) — while IDH2 $\times$ CDK6 $= +0.174$ and IDH2 $\times$ EZH2 $= +0.396$, consistent with IDH2 (not IDH1) being the metabolically coupled isoform in the myeloid lineage. The tissue-specificity of the IDH isoform coupling supports the model's prediction that the metabolic gate operates through lineage-specific chemistry.

6.5.3 LIHC Progenitor Lock-in: AFP versus ALB Coupling Architecture

The Mode 3 prediction states that hepatocellular carcinoma cells are locked in a progenitor-direction commitment program, with AFP (alpha-fetoprotein, the fetal hepatocyte marker) coupled to commitment chemistry and ALB (albumin, the mature hepatocyte marker) decoupled or opposite-signed. TCGA LIHC ($n = 371$) provides the test.

AFP shows positive coupling to commitment-panel members including TDG $\rho = +0.332$ ($p = 5.2 \times 10^{-11}$), EZH2 $\rho = +0.311$ ($p = 1.0 \times 10^{-9}$), and RAD51 $\rho = +0.342$ ($p = 1.4 \times 10^{-11}$), with

weaker coupling to BRCA1 $\rho = +0.182$ ($p = 4.2 \times 10^{-4}$) and OGG1 $\rho = +0.134$ ($p = 0.01$).

NUMB is anti-coupled: AFP $\times$ NUMB $\rho = -0.169$ ($p = 1.1 \times 10^{-3}$). ALB shows opposite-sign coupling to most commitment-panel members: TDG $\rho = -0.148$, EZH2 $\rho = -0.157$, RAD51 $\rho = -0.213$, BRCA1 $\rho = -0.204$, and OGG1 $\rho = -0.116$. NUMB is positively coupled to ALB: $\rho = +0.140$ ($p = 7.1 \times 10^{-3}$). The progenitor marker and the mature marker produce sign-reversed coupling to the commitment chemistry across the panel.

The purging machinery shows the same sign reversal. AFP couples positively to all three purging components: APEX1 $\rho = +0.221$ ($p = 1.7 \times 10^{-5}$), BAX $\rho = +0.363$ ($p = 5.6 \times 10^{-13}$), BAK1 $\rho = +0.285$ ($p = 2.3 \times 10^{-8}$). ALB couples negatively: APEX1 $\rho = -0.130$, BAX $\rho = -0.170$, BAK1 $\rho = -0.257$ ($p = 5.5 \times 10^{-7}$). AFP-associated cells retain engaged purging; ALB-associated cells do not. AFP also shows fidelity-dominant resolution: BLM $\rho = +0.376$ ($p = 6.6 \times 10^{-14}$) with TLS anti-coupled (POLK $\rho = -0.358$, REV3L $\rho = -0.201$). The progenitor-locked architecture in LIHC retains both commitment-chemistry coupling and fidelity resolution, consistent with a cell trapped in the progenitor program that is running the write/erase chemistry without progressing to mature hepatocyte identity.

The Mode 3 prediction — opposite-sign coupling between progenitor and mature lineage markers across the commitment panel — is supported by the observed pattern. AFP $\times$ ALB coupling is weakly positive ($\rho = +0.141$, $p = 6.6 \times 10^{-3}$), indicating partial co-expression rather than mutual exclusion.

6.5.4 PRAD TF-Target Bifurcation: AR versus KLK3 Coupling Architecture

The Mode 4 prediction states that in prostate adenocarcinoma the master lineage transcription factor (AR) retains coupling to commitment chemistry while its canonical downstream target

(KLK3/PSA) is decoupled or reversed, reflecting a bifurcation in which the transcription factor is active but its downstream program is disconnected from the commitment architecture. TCGA PRAD (n = 497) provides the test.

AR shows strong positive coupling to commitment-panel members: KDM6A $\rho = +0.738$ ($p = 2.0 \times 10^{-86}$), TET2 $\rho = +0.671$ ($p = 1.9 \times 10^{-66}$), NUMB $\rho = +0.590$ ($p = 5.3 \times 10^{-48}$), and BRCA1 $\rho = +0.524$ ($p = 2.5 \times 10^{-36}$). KLK3 shows the opposite pattern — predominantly negative coupling to the same genes: BRCA1 $\rho = -0.416$ ($p = 2.8 \times 10^{-22}$), TET2 $\rho = -0.378$ ($p = 2.8 \times 10^{-18}$), NUMB $\rho = -0.203$ ($p = 5.1 \times 10^{-6}$), and KDM6A $\rho = -0.276$ ($p = 3.8 \times 10^{-10}$). The bifurcation extends to fidelity machinery: AR couples positively to BLM ($\rho = +0.260$) and BRIP1 ($\rho = +0.651$), while KLK3 anti-couples to both (BLM $\rho = -0.396$, BRIP1 $\rho = -0.448$). The master TF retains the full commitment and fidelity architecture; the downstream target is disconnected from it.

The AR/KLK3 coupling reversal is among the strongest bifurcation effects observed in any mode test. The cross-correlation between markers is -0.192 ($p = 1.7 \times 10^{-5}$), indicating that AR and KLK3 are co-expressed (KLK3 is an AR target gene), yet produce opposite coupling signatures with the commitment panel — consistent with transcription-factor-level coupling that does not propagate to downstream gene products.

6.5.5 GBM Lineage Decoupling: SOX2 versus GFAP Coupling Architecture

The Mode 2 prediction states that in glioblastoma the stem-associated marker (SOX2) retains coupling to execution chemistry while the differentiation-associated marker (GFAP) is broadly decoupled, reflecting lineage-independent proliferation that bypasses the normal commitment sequence. TCGA GBM (n = 154) provides the test.

SOX2 retains execution coupling: PARP1 $\rho = +0.436$ ($p = 1.6 \times 10^{-8}$), BRCA1 $\rho = +0.421$ ($p = 5.4 \times 10^{-8}$), CHEK1 $\rho = +0.378$ ($p = 1.4 \times 10^{-6}$), and FANCD2 $\rho = +0.289$ ($p = 2.8 \times 10^{-4}$). SOX2 is anti-coupled to NUMB ($\rho = -0.281$, $p = 4.1 \times 10^{-4}$) and to BAK1 ($\rho = -0.457$, $p = 2.6 \times 10^{-9}$), consistent with stem-like cells retaining repair execution while excluding partition-dependent purging.

GFAP shows the predicted broad decoupling. The strongest GFAP correlations are all negative: TDG $\rho = -0.427$, FANCI $\rho = -0.378$, WRN $\rho = -0.357$, PCNA $\rho = -0.345$, APEX1 $\rho = -0.334$. GFAP $\times$ BRCA1 is near zero ($\rho = -0.011$), GFAP $\times$ PARP1 is near zero ($\rho = -0.029$), and GFAP $\times$ NUMB is weakly positive ($\rho = +0.106$). The GFAP-associated differentiation program is disconnected from the commitment and execution machinery across all functional categories.

The SOX2/GFAP divergence is observed as predicted: SOX2 retains execution and fidelity coupling while GFAP is globally decoupled. The cross-correlation between markers is $+0.170$ ($p = 0.04$), indicating that SOX2 and GFAP mark partially overlapping populations, yet their coupling architectures with the model panel are fundamentally different.

6.6 Single-Cell Evidence: Coordination Across Tissues and Differentiation

The coordination the model describes is measurable at single-cell resolution across multiple tissues in normal human tissue (Figure 4A). Using the CZI CELLxGENE Census (2024-07-01 release), donor-level pseudo-bulk correlations were computed for seven cell types across five tissues, totaling 127 donors and 79,914 cells.

Execution pairs are universally preserved at within-cell-type resolution (Methods §9.2; Spearman ρ computed across cells within each cell type, per-pair subset where both genes are detected). CLSPN $\times$ PCNA is positive in every cell type where both genes are detected (mean ρ

= +0.394, SD = 0.081 across 7 cell types). $PLK1 \times AURKA$ is positive in 7/7 (mean +0.496). $FANCI \times FANCD2$ is positive in 8/8 (mean +0.249, SD = 0.087). These structurally obligate complexes couple identically whether the cell is an intestinal stem cell, a hepatocyte, an HSC, a keratinocyte, or a prostate luminal cell.

Regulatory coordination pairs are tissue-specific. $NUMB \times PARD3$ ranges from -0.019 in intestinal crypt stem cells to $+0.915$ in kidney proximal tubule epithelium across the seven pseudo-bulk cell types where both genes are detected (6–36 donors per cell type; intestinal crypt value from within-cell-type analysis). The gradient tracks regenerative capacity: proximal tubule ($+0.915$) and hepatocyte ($+0.858$) — the two cell types with the strongest documented regenerative capacity in humans — anchor the top, while podocytes ($+0.486$) — non-renewable cells that cannot be replaced after injury — sit near the bottom. Between them, breast myoepithelial cells ($+0.829$), HSCs ($+0.695$), type II pneumocytes ($+0.540$), and kidney collecting duct cells ($+0.642$) each occupy positions consistent with their known turnover biology. Keratinocytes show $CDK4 \times CCND1$ at -0.782 across 10 donors — a squamous-specific inversion consistent with the known anti-correlation between proliferation and terminal differentiation in stratified epithelium. Podocytes uniquely show $OGG1 \times BAX$ at -0.257 , the only cell type where the commitment arm is negative, consistent with the absence of active commitment/purging machinery in a cell that has no pathway to further differentiation.

The write/erase subsystem specificity described in Part 3 is visible at single-cell resolution: $TDG \times EZH2$ is positive in 13/13 cell types (mean +0.133), $OGG1 \times BAX$ is positive in 19/20 cell types (mean +0.126). In the TCGA bulk data, $TDG \times EZH2$ is positive in 8/9 cohorts and $OGG1 \times BAX$ in 9/9. $OGG1 \times NUMB$ is negative in 9/9 cohorts, consistent with the commitment arm being anti-correlated with the partition mediator.

The data are consistent with cancer operating as a coordination failure rather than an expression-level failure. Across all 9 TCGA cohorts, the median fold-change of model gene expression between tumor and normal tissue is 1.05. Zero genes exceed 2-fold change in 8 of 9 cohorts. The same genes, at the same expression levels, lose their coordinated behavior in cancer: mean $|\Delta\rho| = 0.259$ (per-cohort average) for regulatory pairs.

The intestinal differentiation trajectory shows coordination peaking during amplification. In the CZI Census data, DNMT1 $\times$ PCNA goes from +0.255 in intestinal crypt stem cells to +0.655 in transit-amplifying cells to +0.110 in mature enterocytes. CDK4 $\times$ CCND1 follows the same pattern (+0.070 $\rightarrow$ +0.351 $\rightarrow$ +0.140). In ARC terms, the Remodeling phase (transit-amplifying) is the active coordination phase; Commitment (enterocyte) shows reduced coupling because the decisions have been made.

The HSC-to-monocyte contrast provides computational evidence for the trajectory described in the §2.4 walkthrough: the quiescent HSC that maintains all its coordination machinery in a coupled, ready state progressively disengages specific subsystems as it commits to a specific lineage. Execution coupling (3-pair set: CLSPN $\times$ PCNA, DNMT1 $\times$ PCNA, PLK1 $\times$ AURKA) drops from +0.937 in HSCs (9 donors) to -0.031 in monocytes (22 donors, classical monocyte hierarchy compartment) — a change of 0.968. Commitment coupling (4-pair set: TDG $\times$ EZH2, OGG1 $\times$ BAX, TDG $\times$ BAX, OGG1 $\times$ EZH2) drops from +0.875 to +0.708 — a change of only 0.167. Normal differentiation disengages execution while preserving commitment identity. Five of six committed hematopoietic lineage endpoints show commitment $>$ execution coupling; only megakaryocytes reverse this, consistent with their requirement for ongoing DNA replication (endomitosis) to achieve polyploidy.

6.7 Routing Decisions in the Hematopoietic Hierarchy

The model predicts that the $A \rightarrow R \rightarrow C$ trajectory described in §2.4 should be visible as a progressive, selective change in coordination profiles (Figure 4B–C): the quiescent HSC should maintain near-total coordination across all functional subsystems, and different subsystems should disengage at different points along each lineage trajectory depending on what each destination cell type biologically requires. Donor-level pseudo-bulk correlations across 15 hematopoietic cell types (9–36 donors per type) from the CZI CELLxGENE Census test this prediction at single-cell resolution.

The HSC itself maintains the coordination the model ascribes to the pre-ARC quiescent state. Across 9 donors, execution pairs (3-pair set: CLSPN $\times$ PCNA, DNMT1 $\times$ PCNA, PLK1 $\times$ AURKA) average $\rho = +0.937$ (CLSPN $\times$ PCNA = +0.933, DNMT1 $\times$ PCNA = +0.967, PLK1 $\times$ AURKA = +0.912), commitment pairs (4-pair set: TDG $\times$ EZH2, OGG1 $\times$ BAX, TDG $\times$ BAX, OGG1 $\times$ EZH2) average +0.875 (TDG $\times$ EZH2 = +0.783, OGG1 $\times$ BAX = +0.983, TDG $\times$ BAX = +0.817, OGG1 $\times$ EZH2 = +0.917), and partition coupling is strong (NUMB $\times$ PARD3 = +0.695, NUMB $\times$ DLG1 = +0.833). This is the coupled, ready state from which the $A \rightarrow R \rightarrow C$ trajectory launches. The multipotent progenitor (MPP, 19 donors) already shows selective loosening: CDK4 $\times$ CCND1 drops from +0.915 to +0.338 while execution pairs remain above +0.700 — the cell-cycle regulatory coupling begins to diversify before the execution machinery changes.

The myeloid-lymphoid branch point is the most dramatic routing signal in the hierarchy. At the common myeloid progenitor (CMP, 18 donors), replication-coupled pairs remain strongly positive: CLSPN $\times$ PCNA = +0.821, DNMT1 $\times$ PCNA = +0.868. At the lymphoid-restricted progenitor (CLP, 18 donors), these same pairs invert: CLSPN $\times$ PCNA = -0.455 ($\Delta = 1.276$),

$DNMT1 \times PCNA = -0.560$. Simultaneously, partition coupling maximizes: $NUMB \times PARD3$ rises from $+0.384$ in CMPs to $+0.985$ in CLPs, and $NUMB \times DLG1$ reaches $+1.000$. The lymphoid route decouples replication and maximally engages partition — consistent with CLP biology, where cells must coordinate V(D)J recombination (deliberate DNA breaks for immunoglobulin and T-cell receptor rearrangement) rather than replicative expansion.

Within the myeloid branch, the GMP-MEP split shows a different pattern. MEP (18 donors) maintains near-total coordination across all pairs (mean $\sim +0.85$), consistent with the requirements of its daughter lineages: megakaryocytes undergo endomitosis to achieve polyploidy, and erythrocytes undergo enucleation — both requiring sustained coordination of replication and chromatin machinery. GMP (10 donors) shows widespread decoupling ($NUMB \times PARD3 = -0.374$), consistent with the myeloid progenitor's transition toward lineages that will abandon the stem-cell partition program.

Committed lineages show a consistent pattern: five of six retain stronger commitment coupling than execution coupling. Monocytes (22 donors, classical monocyte hierarchy compartment) — which represent a late committed stage that retains the capacity for further differentiation into macrophages and dendritic cells — provide the clearest example: execution pairs average -0.031 , commitment pairs average $+0.708$ — the cell has disengaged its replication machinery while maintaining its identity architecture. Only megakaryocytes reverse this (execution $>$ commitment by $+0.058$), consistent with their unique requirement for ongoing replication without division. The selective disengagement of execution while commitment is preserved is the normal-tissue equivalent of what the model predicts cancer disrupts — except that in cancer, it is the regulatory coordination that dissolves while execution is preserved.

The B cell trajectory (CLP → Pro-B → Pre-B → Naive B) shows the Fanconi complex decoupling as its biological purpose is completed: $FANCI \times FANCD2$ declines from +0.985 in CLPs to +0.912 in Pro-B, +0.830 in Pre-B, and +0.413 in Naive B cells, tracking the completion of V(D)J recombination and the diminishing requirement for recombination-associated DNA damage processing.

Each branch point shows a different subset of pairs diverging, and the pairs that diverge match what each destination lineage biologically requires. Replication coupling for proliferative routes (MEP), partition coupling for recombination routes (CLP), commitment coupling for identity-maintaining routes (monocyte). The pair-coupling approach reveals these routing decisions at a resolution that is invisible in individual gene expression levels.

6.8 Limitations and Outstanding Tests

The computational evidence presented here is entirely correlational. The model interprets changes in Spearman correlation as reflecting changes in the underlying mechanistic coupling between gene products, but correlations in transcriptomic data can arise from shared upstream regulation, cell-type composition shifts between normal and tumor tissue, batch effects, or purity confounders. The protein-level CPTAC evidence partially addresses the transcriptional-regulation concern, and the single-cell analyses partially address the composition concern, but neither eliminates all alternative explanations. A proxy assessment of tumor purity confounding — using the tumor-to-normal sample ratio as a surrogate for purity variation across cohorts — finds no significant correlation between this ratio and the magnitude of regulatory coordination change ($\rho = +0.417$, $p = 0.265$) or execution change ($\rho = -0.317$, $p = 0.406$). However, the ratio does correlate with the relative magnitude of the class separation ($\rho = +0.750$, $p = 0.020$), which could reflect either purity confounding or the greater statistical power of larger cohorts. This

correlation warrants acknowledgment even though the direction of its interpretation is ambiguous. The definitive test of the model's central architectural claim about cancer — that NUMB protein is asymmetrically partitioned at mitosis in normal stem cells and that this asymmetry is lost in cancer — requires direct immunofluorescence imaging of dividing cells in matched normal and tumor tissue, which is beyond the scope of computational analysis.

Several specified predictions were not supported or were only partially supported, and these negative results are reported here with the same detail as the positive results.

The Lee 2020 and Pelka 2021 single-cell datasets, despite both being CRC cohorts, show poor pair-level concordance (one of six comparable pairs in the same direction). This reflects the substantial differences between the datasets — different patient populations (Korean versus American/European), different cohort sizes (23 versus 62 patients), and critically, different numbers of matched normal patients (eight versus thirty-five). Pelka's larger normal sample produces more stable Spearman estimates, and where both datasets are informative, Pelka's results align more closely with the TCGA bulk findings. The Lee 2020 results should be interpreted with caution for individual pair values, though the presence of three sign reversals (a count, not dependent on individual-pair accuracy) remains informative.

Three specified tests were not executed in this battery. The LARRY barcoding sister-cell partition analysis (Weinreb et al., 2020), which would test whether model-gene inheritance is indeed asymmetric at division, requires access to published sister-cell transcriptomic datasets not analyzed here. The spatial transcriptomic analysis needed to resolve the NUMB $\times$ SCRIB prediction (which the model now interprets as a spatial-separation rather than a co-expression prediction) requires spatial RNA-seq data from normal intestinal tissue. The polarity-state-

dependent G4-mutation-rate prediction stated in §5.7 requires whole-genome sequencing of tissues stratified by polarity-state characterization, with mutation rates at G4 regulatory addresses compared between polarity-disrupted and polarity-intact subsets; Aghili et al. (2014) provide a partial empirical anchor for this design (discussed in §8.4) but do not themselves constitute confirmation. Each of these represents a next-generation test that would substantially strengthen or refute the model.

PIDD1, which the model identifies as a candidate molecular adapter between centrosomal partition and fork-level resolution control (§3.4), was tested in a separate analysis run after the 28+8 pair classification was established, making it a prospective rather than iterative addition. The PIDD1 network analysis (21 pairs involving PIDD1, CASP2, CRADD, ANKRD26, and MDM2 across the same nine TCGA cohorts) produced results consistent with the Class I/II separation pattern. All 14 PIDD1-containing pairs weaken in cancer, with no exceptions. The processing regulators ATR and CHEK1 — implicated in the PIDD1-to-PIDD-C conversion — weaken at matched magnitudes ($\Delta\rho = -0.145$ and -0.135 respectively), consistent with coordinated disruption of the processing gate. Two pairs behave as execution partnerships: PIDD1 $\times$ PCNA ($\Delta\rho = -0.005$, essentially preserved) and PIDD1 $\times$ CRADD ($\Delta\rho = -0.012$), both structurally obligate interactions. The largest single decoupling observed across all analyses is CASP2 $\times$ TP53 ($\Delta\rho = -0.450$), consistent with profound disruption of the caspase-2-mediated p53 stabilization axis in cancer. Per-pair values are deposited in `pidd1_network_results.json`. The PIDD1 pairs were defined after the 28+8 classification framework was established and were initially reported separately. Subsequent reanalysis confirmed that the PIDD1 network results are consistent with the primary Class I/II separation: the 14 weakening PIDD1 regulatory pairs would extend the Class I inventory, while the two preserved structural pairs (PIDD1 $\times$ PCNA,

PIDD1 × CRADD) behave as Class II. The 28+8 primary classification is retained for continuity with the deposited analysis files.

The five failure modes described in §5.4 were tested at different levels of completeness in this analysis. Mode 1 (phase-specific cliff, HNSC) is reported in §6.5.1 using the Puram 2017 single-cell dataset, with the predicted fidelity-index contrast confirmed between TP63 and KRT5 coupling profiles. Mode 2 (phase-uniform decoupling, GBM) is reported in §6.5.5 using TCGA bulk data, with the predicted uniform GFAP decoupling confirmed across all functional categories (commitment, fidelity, TLS, purging) while SOX2 retains coupling — the largest coupling reversal is EZH2 $\Delta\rho = 0.584$. Mode 3 (progenitor lock-in, LIHC) is reported in §6.5.3 using TCGA bulk data, with the predicted opposite-sign coupling between AFP and ALB confirmed across 6/6 commitment-panel members and all 3 purging components. Mode 4 (TF-target bifurcation, PRAD) is reported in §6.5.4 using TCGA bulk data, with the predicted AR/KLK3 coupling reversal confirmed at the largest effect size observed in any mode test (BRIP1 $\Delta\rho = 1.10$). Mode 5 (quiescent evasion, AML) is reported in §6.5.2 using both the van Galen 2019 single-cell dataset (hierarchy-resolved reanalysis) and TCGA LAML bulk data; the Mode 5 prediction is supported by convergent evidence — at single-cell resolution, malignant HSC-like cells are quiescent (1.3–3.8% cycling across 4/4 patients with sufficient HSC-like counts; per-patient HSC-to-non-HSC fold-changes range 1.05× to 14.23× with AML419A demonstrating that the effect can attenuate to within measurement noise), and at bulk level, CD34 (stem) is decoupled from commitment chemistry while MPO (myeloid) engages the full panel with fidelity-dominant resolution (POLK anti-coupled at $\rho = -0.373$). The CD34 × BAX reversal ($\Delta\rho = 0.477$) is comparable in magnitude to Mode 3 and Mode 4 effects.

The complete verification package — 22 generation scripts, 25 results JSONs, 7 source CSVs, and a 264-check verification script — is deposited at [10.5281/zenodo.20402772](https://zenodo.org/record/20402772).

6.9 Sensitivity to Cell-Cycle Covariates

The class separation in §6.2 could reflect cell-cycle-phase composition differences between tumor and normal tissue rather than cancer-specific regulatory dissolution. To test this, cell-cycle phase scores (G2M_min: minimum z-score across CDK1, CCNB1, CCNB2, PLK1, AURKA, AURKB, BUB1, TOP2A, CDC20; S_mean: mean z-score across PCNA, MCM2-7, GMNN, CCNA2) were computed per sample and regressed out of each gene before recomputing Spearman correlations and $\Delta\rho$ values.

After residualization, Class I pairs retain 79% of their original mean $|\Delta\rho|$ (0.1385 residualized versus 0.1752 original), while Class II mean $|\Delta\rho|$ increases from 0.0444 to 0.0779. The Mann-Whitney test on residualized values yields $U = 164$, $p = 0.0243$ — attenuated from the original $p = 8.53 \times 10^{-6}$ but confirming that the class separation persists when cell-cycle covariates are removed. Of 28 Class I pairs, 23 retain more than half their original $|\Delta\rho|$. The attenuation is expected: some regulatory coordination pairs (particularly the BER tetrad and epigenetic writer-eraser pairs) are partially correlated with cell-cycle phase, so regressing out phase scores removes part of the true signal alongside any phase-composition confound. The persistence of the class separation after this conservative adjustment supports the interpretation that the regulatory dissolution pattern reflects cancer-specific biology rather than a trivial consequence of altered cycling rates.

Part 7: Implications

If the coupling architecture reported in Part 6 reflects an underlying partition failure, its main implication is therapeutic: cancer should be stratified not only by mutated genes but by the architectural layer at which commitment fails. This section consolidates the therapeutic predictions that emerge from the five failure modes (§5.4), the computational evidence (Part 6), and the death-modality framework (Part 4) into a unified account.

7.1 Therapeutic Implications of Selective Coordination Failure

The coordination architecture described in this manuscript — in which regulatory coupling dissolves while execution machinery remains intact — has direct consequences for how therapeutic targets are selected. If cancer's fundamental defect is the selective dissolution of regulatory coordination while execution machinery is preserved, then therapies targeting the execution machinery — which is what most current cytotoxic agents do — address the part of the system that cancer has preserved rather than the part that has failed. Taxanes stabilize microtubules (execution machinery). Platinum compounds crosslink DNA at the replication fork (execution machinery). PARP inhibitors block single-strand break repair (execution machinery) (reviewed in Hanahan & Weinberg, 2011). These agents work, often dramatically, but the model's interpretation of their mechanism is that they succeed by overwhelming the execution layer with damage it cannot process, not by correcting the regulatory failure that produced the cancer in the first place. The distinction matters because it predicts the conditions under which relapse occurs: the regulatory failure persists after treatment, and any surviving cell with intact execution machinery can regenerate the tumor using the same disordered regulatory program.

A model-informed therapeutic strategy would instead target the regulatory coordination layer directly, or exploit the specific dependencies that the regulatory failure creates. The differential CDK6/CDK4 behavior reported in §6.4 illustrates a broader point: current CDK4/6 inhibitors —

palbociclib, ribociclib, abemaciclib — target both CDK4 and CDK6 with comparable affinity.

The computational evidence shows that CDK6's regulatory relationships weaken in cancer while CDK4's structural coupling is preserved (§6.4). Whether this differential behavior has therapeutic consequences — for instance, whether patients whose tumors retain CDK6 regulatory coupling respond differently to CDK4/6 inhibition than those whose tumors have already lost it — is an empirical question testable by stratifying existing trial data by pre-treatment $CDK6 \times CCND1$ coupling in biopsies.

The MMRd versus MMRp stratification in the Pelka CRC data generates a second specific therapeutic prediction. The Fanconi anemia ID complex ($FANCI \times FANCD2$) is preserved in MMRd tumors and disrupted in MMRp tumors (§6.5). The biological interpretation is that MMRd tumors, having lost mismatch repair, depend on the FA pathway as a compensatory DNA damage tolerance mechanism. This dependency is a vulnerability: agents that directly disrupt FA pathway function should be more effective against MMRd tumors than MMRp tumors, because MMRd tumors cannot afford to lose a pathway whose disruption MMRp tumors apparently tolerate. The prediction is the opposite of the synthetic-lethality logic that underlies PARP inhibitor use in BRCA1/2-mutant cancers (which exploits pre-existing HR/FA deficiency); here, the prediction is that intact-FA MMRd tumors will be more sensitive to direct FA-pathway inhibition than the disrupted-FA MMRp tumors. The prediction is testable in existing clinical trial cohorts where both MMR status and FA-pathway-targeting agent response data are available. A parallel prediction follows from Mode 3 (progenitor lock-in): the JH-RE-06 class of REV1 inhibitors (Wojtaszek et al., 2019), which targets the mutagenic translesion synthesis that Mode 3 cells depend on, should show selective efficacy in tumors with the progenitor-locked coupling signature.

The CLSPN fork-protection module's invariant preservation across all conditions tested suggests a third category of therapeutic opportunity. If the fork-protection module is preserved, which the model interprets as reflecting cancer's dependence on it for symmetric (undifferentiated) division, then disrupting fork protection specifically — through CLASPIN-targeting agents or through agents that disrupt the ATR-CHEK1-CLASPIN checkpoint axis — should be selectively toxic to cancer cells that depend on symmetric division for self-renewal. The prediction is not that fork-protection disruption would kill cancer cells because they are dividing (normal stem cells also divide), but that it would kill cancer cells because they are dividing symmetrically, and symmetric division requires intact fork-protection coordination in a way that asymmetric division may not.

7.2 Therapeutic Implications: Gate Restoration and Sequencing

The model's most direct therapeutic implication is that cancer can be treated by restoring the R-to-C gate rather than by killing cancer cells. The IDH inhibitor clinical data provides evidence consistent with this approach: ivosidenib and enasidenib produce terminal differentiation of AML blasts in IDH-mutant patients, with clinical remission achieved through blast maturation rather than blast death (Stein et al., 2017; DiNardo et al., 2018). The model interprets this as gate restoration — the IDH mutation blocked the write/erase program by metabolic starvation, and the inhibitor re-opens the gate by restoring α -KG availability. High-dose vitamin C, which restores TET2 enzymatic activity by maintaining iron in its reduced active state (Cimmino et al., 2017), represents a related gate-restoration approach operating through the same cofactor axis. The per-cancer temporal strategies described in §5.4 suggest that different cancers trap at different gates, requiring different restoration strategies, but the underlying principle is

conserved: identify the blocked gate, relieve the blockade, and allow trapped R-phase cells to complete commitment.

The death-modality stratification developed in Part 4 generates a further prediction about therapeutic sequencing. If the death modality a cancer cell undergoes is determined by its partition state, then shifting a cancer cell toward commitment before killing it should produce immunologically silent apoptotic death rather than the inflammatory death that releases damage-associated molecular patterns into the tumor microenvironment. The model predicts that differentiation-priming agents followed by cytotoxic agents should produce better tumor control than the reverse order, even when total cell kill is equivalent. This sequencing prediction is testable in preclinical models and has not been formally tested in clinical trials.

Part 8: Discussion

8.1 What the Model Contributes

This manuscript proposes that cell fate decisions in metazoan tissues are governed by an integrated architectural program — the ARC system — in which the asymmetric partition of nuclear marks and cytoplasmic organelles at mitosis determines each daughter's trajectory. The computational evidence in Part 6 does not prove this proposal, but it provides three categories of evidence that bear on its plausibility.

The two-class pattern itself is an empirical observation that does not depend on the model's interpretive apparatus. Any model of cancer — whether the hallmarks framework (Hanahan & Weinberg, 2011) or any alternative — must account for why some gene-pair correlations change systematically between normal and tumor tissue while others do not, and why the pairs that change are precisely those whose co-regulation reflects regulatory programs rather than structural obligations. The model offers one account — partition failure — but the two-class pattern constrains any alternative account as well.

Beyond the pattern itself, the specific mechanistic predictions that the model generates and that the data supports carry independent weight. The differential CDK6/CDK4 behavior, the CLSPN module preservation, the BER coordination collapse, and the NUMB-to-polarity decoupling are not generic consequences of cellular transformation; they are specific predictions about which relationships should change and which should not, derived from the model's mechanistic claims about the partition system's architecture. The predictions' specificity — not merely that some pairs should change, but that regulatory pairs should change while physical partnerships should not — gives the model falsification power that a vaguer model would lack.

The model's capacity to generate clinically relevant predictions with enough specificity to be testable constitutes a third line of evidence. The CDK differential-behavior observation, the MMRd-FA dependency, the fork-protection vulnerability, and the partition-restoration therapeutic strategy are all consequences of the model's architectural interpretation, and each is testable with existing clinical data or straightforward experimental designs. A model that generates specific, falsifiable therapeutic predictions earns a different epistemic status from one that merely redescribes known biology in new language.

8.2 What the Model Does Not Prove

The model's computational evidence is entirely correlational. Changes in Spearman correlation between normal and tumor tissue could reflect direct mechanistic coupling changes (the model's interpretation), but they could also reflect cell-type composition shifts between tissue types, shared upstream regulatory changes, or technical confounders in the source datasets. The protein-level CPTAC evidence and the single-cell analyses partially address these alternatives by operating within defined cell populations, but neither eliminates them. Partial correlation controlling for proliferation rate or cell-cycle phase has not been systematically applied across the full test battery and remains a methodological limitation.

A structural vulnerability of the model warrants explicit acknowledgment. The manuscript assigns architectural roles to approximately fifty proteins — partition determinants, write/erase enzymes, fidelity helicases, execution partners — and interprets their coupling patterns through the ARC framework. A sufficiently flexible model with enough components can be post-hoc fitted to almost any dataset, and the concern that the model is unfalsifiable because it interprets each protein's established function as evidence for the framework is legitimate. Three features of the model's design are intended to address this concern, though none eliminates it entirely. Each

pair's classification as regulatory or execution was evaluated against five molecular criteria (direct physical binding, shared compartment, absence of partner-independent functions, constitutive association, and stoichiometric co-expression) that are derived from established molecular biology and are independently verifiable by any reviewer without reference to the cancer data — the C1–C5 scores for all 38 scored pairs (28 Class I, 8 Class II, and 2 excluded ambiguous pairs), with literature citations for every criterion, are provided in Table S3. The model generates specific coupling predictions that can fail — the sensitivity analyses in §6.2 demonstrate that the framework's statistical conclusion depends on specific pair assignments, and alternative classifications produce quantitatively different (though consistently significant) separations. And the prediction that cancer should selectively disrupt one class while preserving the other is not a prediction that any arrangement of fifty proteins would satisfy — the model assigns proteins to functional categories based on whether their co-regulation reflects regulatory programs or structural obligations, a mechanistic criterion independent of the cancer data. Nevertheless, the possibility remains that the protein panel was unconsciously selected to include components whose coupling patterns would match the model's predictions, and independent replication using protein panels selected by investigators without knowledge of the model's predictions would provide the strongest defense against this concern. Supplementary Table S1 addresses this transparency requirement directly: for each gene in the model panel (49 core genes listed in §9.1, plus 14 lineage markers used in mode-specific tests — 63 genes total in Table S1), the table separates the established biochemical function (supported by primary literature citations that a reviewer can verify without accepting any part of the framework) from the model's specific architectural interpretation, and specifies a per-gene falsification test whose failure would require revising that protein's assignment.

The model's central causal claim — that loss of partition asymmetry accelerates mutation accumulation through missegregation of chromatin and mislicensing of the write/erase chemistry, with the affected addresses preferentially being the G-quadruplex-containing regulatory addresses where cancer driver mutations recurrently occur — remains the model's most far-reaching claim and is not directly tested by the computational evidence presented here. The computational data are consistent with the claim insofar as the coupling between partition components and the write/erase machinery weakens in cancer, but the data do not directly measure mutation accumulation rates as a function of partition state. The direct test requires comparing mutation accumulation rates at G-quadruplex-containing regulatory addresses between partition-disrupted and partition-intact tissues, with the prediction being that partition disruption is associated with elevated mutation rates specifically at G4 regulatory addresses rather than uniformly across the genome. This test is technically feasible using existing whole-genome sequencing platforms combined with polarity-state characterization (e.g., NUMB asymmetric segregation imaging at mitosis), and a negative result — equal mutation rates between partition-disrupted and partition-intact tissues at G4 regulatory addresses — would require fundamental revision of the proposed mutagenic mechanism.

8.3 Relationship to Existing Models

The model intersects with but does not duplicate several existing lines of investigation. The cancer stem cell hypothesis (Bonnet & Dick, 1997; reviewed in Simons & Clevers, 2011) proposes that tumors are maintained by a subpopulation with stem-like self-renewal capacity. The model is consistent with this proposal but adds a specific mechanism — partition failure — for why the stem-like subpopulation exists and what molecular signature distinguishes it. The epithelial-mesenchymal transition literature (Nieto et al., 2016; Pastushenko & Blanpain, 2019)

describes phenotypic plasticity between epithelial and mesenchymal states in cancer progression. The model interprets EMT as one manifestation of the partition failure's downstream consequences rather than as an independent oncogenic program. The Vogelstein mutation-accumulation model (Fearon & Vogelstein, 1990; Vogelstein et al., 2013) describes the temporal sequence of driver mutations in colorectal cancer. The model proposes a complementary mechanism: that loss of partition asymmetry accelerates the rate at which mutations accumulate at the G4-containing regulatory addresses that coincide with the canonical driver loci, with disordered marking and mislicensing as a proposed mutagenic mechanism. The two models are not mutually exclusive — mutations remain causal for cancer progression in both — but the ARC model adds a specific account of why mutation accumulation should occur preferentially at these addresses and why the rate should depend on partition state. Independent support for address-specific mutation rates comes from the non-random mutagenesis literature: Monroe et al. (2022) demonstrated that epigenomic features explain over 90% of mutation rate variance in *Arabidopsis*, and Schuster-Böckler and Lehner (2012) showed that chromatin organization is a major determinant of regional mutation rates in human cancer — findings consistent with the model's claim that the write/erase chemistry's outputs concentrate at addresses defined by chromatin state.

The missegregation and mislicensing components of this proposed mechanism each have experimental precedent in cancer biology. Crasta et al. (2012) demonstrated that whole-chromosome missegregation at mitosis generates DNA breaks through micronucleus formation, producing damage that can be inherited by daughter nuclei and contribute to chromothripsis. Bedekovics et al. (2021) showed that USP24 loss in neuroblastoma destabilizes the polarity protein CRMP2, producing spindle defects, chromosome missegregation, and aneuploidy — an

architectural pattern (polarity-regulator loss → spindle defect → missegregation → genomic instability) that parallels what the ARC model proposes for the asymmetric-division polarity system. Canonical replication licensing deregulation (Cdt1, Cdc6, Geminin) produces re-replication, replication stress, and genomic instability (reviewed in Petrakis et al., 2016). The ARC model extends these principles to propose that chromatin marks rather than entire chromosomes can inherit incoherently when partition fails, with the write/erase chemistry continuing to operate on substrates inappropriate for the cell state. Each link in the proposed causal chain — NUMB/aPKC partition disruption producing chromatin-mark missegregation that accelerates G4-localized mutation accumulation — has precedent in different molecular systems, but the specific chain has not been demonstrated as an integrated sequence.

Each of these intersections generates testable distinguishing predictions. If the model is correct that the cancer stem cell phenotype is caused by partition failure, then restoring partition function should eliminate the stem-like subpopulation without killing it — by enabling the asymmetric divisions that would normally produce one committed and one self-renewing daughter. If EMT is one of partition failure's downstream consequences, then restoring partition function should reduce EMT independently of the signaling pathways (TGF- β , Wnt, Notch) that are conventionally targeted to inhibit EMT. If partition failure accelerates mutation accumulation at G4-containing regulatory addresses through the proposed missegregation/mislicensing mechanism, then tissues with measurable partition disruption should show elevated mutation rates at these addresses compared to partition-intact tissues — the mutational precursor signature (elevated 8-oxoG at G4 promoters) being the predicted intermediate. Each of these is experimentally accessible with current technology.

8.4 What Would Falsify This Model

A model that cannot specify the conditions under which it would be wrong has no claim to empirical standing. The ARC model makes several predictions whose failure would require fundamental revision.

The strongest falsification test concerns the proposed mutagenic mechanism in §5.7. The model hypothesizes that loss of partition asymmetry accelerates mutation accumulation through missegregation of chromatin and mislicensing of the write/erase chemistry, with the elevated mutation rate enriching specifically at G-quadruplex-containing regulatory addresses. If tissues with measurable polarity disruption (quantified by $\text{NUMB} \times \text{polarity coupling}$ or by NUMB asymmetric segregation at mitosis) do NOT show elevated mutation rates at G4-containing regulatory addresses compared to polarity-intact tissues of the same lineage, then the proposed mechanism — that disordered marking accelerates mutation accumulation at the addresses where the chemistry operates — is wrong, and the model's account of how partition state connects to cancer risk would require fundamental revision. This test is technically feasible with current whole-genome sequencing and immunofluorescence imaging. Existing pre-malignant tissue data (Aghili et al., 2014) provide a partial empirical anchor for the test design — somatic copy-number alterations in apparently normal ovarian tissue do show G4 overlap and BRCA-carrier elevation — but do not themselves constitute confirmation, because the Aghili cohort assayed copy-number changes rather than point mutations and was characterized for BRCA status rather than for polarity state.

The mode-replication test concerns the classification itself. The model predicts that the five modes identified in §5.4 are cross-tissue stable: a Mode 1 signature in HNSC should be reproducible in independent HNSC cohorts, and Mode 3 in LIHC should be reproducible in

independent hepatocellular carcinoma datasets. If the mode signatures do not replicate across independent cohorts of the same cancer type, the classification reflects dataset-specific artifacts rather than stable architectural features.

The partition-restoration test concerns the death-differentiation continuum. The model generates the hypothesis that restoring NUMB $\times$ polarity coupling in a cancer cell with intact execution machinery could produce asymmetric division — one committed and one self-renewing daughter — rather than simple growth arrest or death. If NUMB overexpression in a cancer cell line with documented polarity disruption produces growth arrest or apoptosis but NOT asymmetric division (assessed by sister-cell transcriptomics after the next mitosis), then the model's account of NUMB as a partition organizer is wrong, and NUMB's tumor-suppressive effect operates through a different mechanism (perhaps through p53 stabilization alone, without the partition function). The LARRY barcoding technology or equivalent sister-cell tracking could distinguish these outcomes.

The class-universality test concerns the two-class pattern's predictive scope. The model predicts that the Class I/Class II separation will hold in every solid cancer type tested, because the functional classification is universal. If a cancer type is found in which execution partnerships are disrupted while regulatory coordination is preserved (the opposite of the two-class pattern), the model's functional classification is not universal and must be revised to accommodate tissue-specific category inversion. Extending the nine-cohort analysis to additional cancer types (e.g., melanoma, ovarian, prostate, bladder) would test this prediction.

The sensitivity analyses and independent control comparisons delimit the model's scope and motivate the more specific phase- and lineage-resolved tests described above.

8.5 Cross-Species Conservation and Evolutionary Depth

The co-regulation patterns the model identifies are conserved across species. The 81 percent same-sign conservation of model gene-pair correlations between mouse and human intestinal epithelium (26 of 32 pairs in the SMART-Seq2 atlas, with Class II execution pairs at 87.5 percent and NUMB partition pairs at 8/10) places these patterns within a conserved mammalian program rather than within species-specific regulatory architecture. The hierarchy of conservation — Class II execution at 87.5 percent, BER coordination preserved, partition/NUMB at 80 percent — suggests that the structurally obligate relationships are the oldest and most constrained, while the regulatory coordination patterns are largely conserved but admit species-specific tuning. This hierarchy is consistent with the model's claim that the partition architecture is a fundamental feature of metazoan cell fate decisions, not a derived feature of any particular lineage.

The regulatory coordination that governs asymmetric versus symmetric division outcomes is at least as old as the mammalian radiation, and likely older given that the individual components are conserved across vertebrates and in many cases across bilaterians: the NUMB/aPKC/PAR polarity system is functionally conserved from *Drosophila* to mammals (Knoblich, 2010; Gönczy, 2008), the Fanconi anemia pathway is conserved across vertebrates (Ceccaldi et al., 2016), and TET-mediated cytosine oxidation has been documented in metazoans spanning cnidarians to mammals (reviewed in Kohli & Zhang, 2013). Cancer, in this view, is not a disease that exploits novel molecular pathways; it is a disease that corrupts an ancient architectural program whose individual components are well characterized but whose coordination as an integrated system has not been previously described at this level of specificity.

The conservation extends beyond metazoans. The Mrc1/CLASPIN histone chaperone that organizes parental histone recycling at the replication fork — the molecular basis for epigenetic inheritance through division — has been demonstrated as functionally conserved across fission yeast, budding yeast, and mammals, with AlphaFold structural predictions identifying a conserved histone-binding domain across nine model organisms spanning fungi, plants, and animals (Yu et al., 2024). The fork protection complex (Mrc1/CLASPIN with Swi1/TIMELESS and Swi3/TIPIN) and the histone variant deposition machinery (CAF-1, HIRA, FACT) are pan-eukaryotic (Gurard-Levin, Quivy & Almouzni, 2014). The TLS machinery is similarly ancient: REV1 was first characterized in yeast (Nelson et al., 1996), and the REV1-REV3L (Pol ζ) scaffolding relationship is conserved from *Saccharomyces* to humans. The BER glycosylases have bacterial functional analogs (Fpg/MutM for OGG1; MutY for MUTYH), and RecQ helicases trace to bacterial RecQ. In budding yeast, asymmetric mitochondrial inheritance between mother and daughter cells determines replicative lifespan (McFaline-Figueroa et al., 2011), paralleling the organelle quality partitioning that Katajisto et al. (2015) demonstrated in mammalian stem cells. In *Caulobacter crescentus*, every division produces geometrically and functionally asymmetric daughters — the swarmer and stalked cells — through regulated proteolysis of the master regulator CtrA (reviewed in Curtis & Brun, 2010), establishing that asymmetric fate determination through division is a bacterial-grade strategy, not a metazoan innovation (Figure 6). The caspase-2/PIDDosome system that connects centrosome integrity to cell elimination (§4.2) is itself ancient: phylogenetic analysis places caspase-2 as ancestral to all bilaterians, predating the deuterostome-specific caspase-9 that initiates canonical intrinsic apoptosis (Krasovec et al., 2023), consistent with the coupling of replication-quality surveillance to fate determination being an early metazoan innovation rather than a mammalian elaboration.

Figure 6. Evolutionary Conservation of ARC Machinery and Architecture. ARC uses ancient cellular modules with lineage-specific implementations.

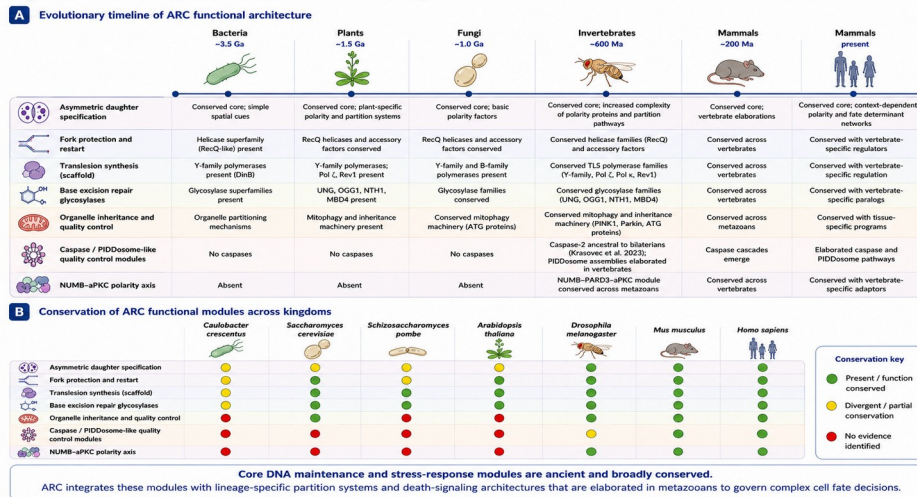


Figure 6. Evolutionary conservation of ARC machinery and architecture. (A) Evolutionary timeline showing approximate depth of each ARC functional module across bacteria (**Caulobacter**, ~3.5 Ga), plants (**Arabidopsis**, ~1.5 Ga), fungi (**S. cerevisiae**, **S. pombe**, ~1.0 Ga), and animals (**Drosophila** ~600 Ma, mouse ~200 Ma, human). Plants (Archaeplastida) diverged from the animal lineage before fungi (Opisthokonta). Fork protection (RecQ helicase family) and TLS scaffold (DinB/Y-family) have bacterial origins with divergent conservation; BER glycosylases have bacterial functional analogs (Fpg/MutM, MutY). Caspase-2 is ancestral to all bilaterians, predating deuterostome-specific caspase-9 (Krasovec et al., 2023). The NUMB-aPKC polarity axis emerges with metazoan cell polarization. (B) Conservation matrix across seven species. Green circles, present and functional (ortholog or clear analog); yellow circles, present but divergent or less well characterized; red circles, not detected or no evidence. Key citations: Tuteja & Tuteja (2004) for bacterial RecQ conservation; Yu et al. (2024) for Mrc1/CLASPIN structural conservation; Katajisto et al. (2015) and McFaline-Figueroa et al. (2011) for organelle inheritance; Curtis & Brun (2010) for **Caulobacter** asymmetry.

The cross-generational nature of this chromatin-marking system, operating through the same DDR machinery across >1.5 billion years of eukaryotic evolution, raises the possibility that the write/erase program concentrates mutations at regulatory addresses in a biochemically non-random way — not directed in the Cairns (1988) sense, but distributed by the operation of a developmental program whose outputs are then tested through the death-differentiation continuum described in Part 4. The testable prediction is that G4-containing commitment promoters should show elevated germline mutation rates relative to matched non-G4 regulatory sequences of similar chromatin state, even after controlling for the known effects of replication timing and mismatch repair efficiency. This prediction connects the within-organism mechanism

characterized in this manuscript to the evolutionary observations described in Part 1, but resolving whether the connection is causal requires data beyond what this manuscript presents.

Part 9: Methods

9.1 Data Sources and Access

All analyses use publicly available data accessed through standard programmatic interfaces without authentication or proprietary database access. No unpublished datasets are used. All code and intermediate results are deposited with the manuscript as supplementary files. Scripts 01–08 cover the Part 6 correlation analyses; script 09 covers the Supplementary Section S2 G4 promoter motif analysis. Methodology for S2 is described in S2 itself, with quantitative results deposited in `s2_g4_tfbs_analysis.json` and `s2_g4_tfbs_reproduced.json`.

All TCGA expression analyses use data frozen from UCSC Xena HiSeqV2 ($\log_2(\text{RSEM}+1)$) on 2026-05-22. Per-cohort and per-pair values are deposited in `final_validation_results.json`. Values in the main text are reported to 3–4 significant figures; full-precision values are in the deposited JSON. The validation script (`03_class_separation_validation.py`) is deposited alongside the manuscript and reproduces every value reported.

The TCGA mRNA expression data were obtained from the UCSC Xena data hub (<https://xenabrowser.net>; accessed April–May 2026) as pre-processed HiSeqV2 files containing $\log_2$ -transformed RSEM-normalized gene expression values ($\log_2(\text{RSEM}+1)$). Nine cancer types were selected based on having at least 35 matched solid-tissue-normal samples (TCGA sample type code 11) alongside primary solid tumor samples (sample type code 01): breast invasive carcinoma (BRCA; 1,097 tumor, 114 normal), kidney renal clear cell carcinoma (KIRC; 533 tumor, 72 normal), lung adenocarcinoma (LUAD; 515 tumor, 59 normal), lung squamous

cell carcinoma (LUSC; 502 tumor, 51 normal), thyroid carcinoma (THCA; 505 tumor, 59 normal), liver hepatocellular carcinoma (LIHC; 371 tumor, 50 normal), colon adenocarcinoma (COAD; 286 tumor, 41 normal), head and neck squamous cell carcinoma (HNSC; 520 tumor, 44 normal), and stomach adenocarcinoma (STAD; 415 tumor, 35 normal). Total: 4,744 tumor and 525 normal samples across 5,269 individuals.

Protein quantification data were obtained from the cBioPortal for Cancer Genomics REST API (<https://www.cbioportal.org/api>) using the molecular profile suffix `_protein_quantification` for seven Clinical Proteomic Tumor Analysis Consortium (CPTAC) studies: `brca_cptac_2020` (n = 122), `coad_cptac_2019` (n = 97), `gbm_cptac_2021` (n = 99), `luad_cptac_2020` (n = 110), `lusc_cptac_2021` (n = 80), `paad_cptac_2021` (n = 140), and `ucec_cptac_2020` (n = 81). Total: 729 patients. All CPTAC cohorts contain tumor-only measurements without matched normal protein data.

Single-cell RNA sequencing data were obtained from the NCBI Gene Expression Omnibus (GEO). The Lee et al. (2020) dataset (GSE132465) provides 63,690 cells from 23 Korean colorectal cancer patients yielding 33 tumor and normal pseudo-bulk samples, profiled by 10x Chromium 3' sequencing — the Korean (SMC) cohort component of Lee et al.'s broader 91,103-cell published atlas profiling 29 patients (23 Korean and 6 Belgian). Of these 23 patients, 8 contributed matched tumor and adjacent-normal epithelial samples. The Pelka et al. (2021) dataset (GSE178341) provides 370,115 cells (the post-QC analyzed subset of Pelka et al.'s 371,223-cell published atlas) from 62 colorectal cancer patients profiled by 10x Chromium 3' sequencing, of which 35 patients contributed matched normal epithelial pseudo-bulks and 60 contributed tumor epithelial pseudo-bulks. The Haber et al. (2017) dataset (GSE92332) provides 1,522 full-length SmartSeq2-profiled mouse intestinal epithelial cells — the deeper-coverage

SMART-Seq2 subset of Haber et al.'s combined 53,193-cell atlas, which also includes a 10x Chromium component — across 11 annotated cell types (Stem, TA, Enterocyte-Immature, Enterocyte-Mature, Goblet, Paneth, Enteroendocrine, Tuft-1, Tuft-2, and two additional progenitor populations). The Puram et al. (2017) dataset (GSE103322) provides 2,215 cells from head and neck squamous cell carcinoma tumors profiled by SMART-Seq2, with TP63-expressing and KRT5-expressing subpopulations used for within-tumor fidelity and fitness-selection coupling analysis (Figure 4). The van Galen et al. (2019) dataset (GSE116256) provides 30,712 cells from acute myeloid leukemia patients, used for the Mode 5 quiescent-evasion analysis of G2/M phase fraction and CEBPA $\times$ BRCA2 coupling in the cycling fraction.

Cross-tissue and hematopoietic hierarchy single-cell analyses use the CZI CELLxGENE Census (version 2024-07-01), accessed via the cellxgene-census Python package (v1.17.0). The Census provides a unified interface to >30 million normal human cells from published scRNA-seq datasets with standardized cell type ontology annotations. For each cell type, expression data for 29 model genes were retrieved using the `obs_value_filter` parameter with `tissue`, `cell_type`, and `disease='normal'` constraints. For tissues exceeding 100,000 cells (lung, breast), `dataset_id` pre-filtering targeted smaller constituent datasets to avoid query timeouts. Cell types exceeding 5,000 cells were randomly subsampled (`seed = 42`). Within-cell-type Spearman correlations were computed for gene pairs where both genes exceeded 3% non-zero expression. Donor-level pseudo-bulk correlations were computed by averaging expression per donor within each cell type (minimum 10 cells per donor, minimum 5 donors per cell type), then computing Spearman ρ across donors. This inter-individual approach produces correlations driven by biological variation rather than technical cell-to-cell noise, and is directly comparable to the inter-patient correlations in the TCGA bulk analysis. Twenty-two cell types across 9 tissues (107,585 cells)

were analyzed at within-cell-type resolution; 10 cell types across 9 tissues (162 donors) at pseudo-bulk resolution; and 15 hematopoietic cell types (110,007 cells, 9–36 donors per type) were analyzed for lineage routing decisions. The reproducibility script (08_census_singlecell_validation.py) and all results (census_singlecell_validation.json, census_pseudobulk_validation.json, hsc_hierarchy_routing.json) are deposited with the manuscript.

9.2 Gene Panel

The model gene panel comprises 49 human genes selected on mechanistic grounds from the ARC model described in Parts 2 and 3 (the full extended panel of 69 genes, including additional genes analyzed in specific sections, is detailed in Table S1). The genes are organized by functional category as follows. regulatory coordination: NUMB (Entrez 8650), NUMBL (9253), PRKCI (5584), PRKCZ (5590), DLG1 (1739), PARD3 (56288), SCRIB (23513), SIRT1 (23411), SIRT6 (51548), CDK6 (1021), RB1 (5925), E2F1 (1869), CCND1 (595), TET2 (54790), TET3 (200424), TDG (6996), OGG1 (4968), MUTYH (4595), NEIL1 (79661), XRCC1 (7515), EZH2 (2146), KDM6A (7403). execution: CLSPN (63967), PCNA (5111), PARP1 (142), CHEK1 (1111), ATR (545), FANCI (55215), FANCD2 (2177), PLK1 (5347), AURKA (6790), DNMT1 (1786), CDK4 (1019). Classified as regulatory coordination based on molecular criteria: BRCA1 (672; partner-independent ICL repair function; Bunting et al. 2012), BLM (641; excluded from class analysis as ambiguous paralog), WRN (7486; excluded from class analysis as ambiguous paralog). Additional panel members: MFN1 (55669), OPA1 (4976), FIS1 (51024), BAX (581), BAK1 (578), BCL2 (596), MCL1 (4170), APEX1 (328), POLK (51426), REV3L (5980), TOP2B (7155), MKI67 (4288), MYC (4609). Mouse orthologs were identified by standard nomenclature conversion (human UPPERCASE to mouse Sentence case) and

confirmed present in the Haber 2017 expression matrix for 48 of 49 genes (missing: PRIMPOL only; Rev3l and Numbl are present in both the UMI-count and SMART-Seq2 atlases). A separate PIDD1 network analysis (§6.8) extends the panel with five additional genes: PIDD1 (55367; listed as LRDD in Xena HiSeqV2), CASP2 (835), CRADD (8738), ANKRD26 (22852), and MDM2 (4193). These are reported independently from the 28+8 primary classification.

9.3 Cell Cycle Phase Scores

Cell cycle phase composite scores were computed as the mean z-score of phase-specific marker genes across samples within each cohort (cell cycle scoring approach adapted from Tirosh et al., 2016). G1 markers: CDT1 (81620), CDKN1B (1027), RB1 (5925), CCND1 (595), CDK4 (1019), CDK6 (1021). S-phase markers: PCNA (5111), MCM2 (4171), MCM3 (4172), MCM4 (4173), MCM5 (4174), MCM6 (4175), MCM7 (4176), GMNN (51053), CCNA2 (890). G2/M markers (G2M_min composite): CDK1 (983), CCNB1 (891), CCNB2 (9133), PLK1 (5347), AURKA (6790), AURKB (9212), BUB1 (699), TOP2A (7153), CDC20 (991). Phase scores were used as control variables in analyses where cell-cycle-phase confounding could produce spurious correlations and as the basis for the Census Catalog cell-cycle-phase assignment (each gene assigned to the phase with which its expression most strongly correlates).

9.4 Statistical Methods

All pairwise gene correlations are Spearman rank coefficients computed on complete-case pairs (samples where both genes have non-missing expression values). For the multi-cohort TCGA analysis, per-cohort Spearman correlations were pooled using the Fisher z-transform: each p was converted to $z = \text{arctanh}(p)$, weighted by $(n - 3)$ where n is the per-cohort sample size, and the inverse-variance-weighted mean z was back-transformed to yield the pooled p . The change

between normal and tumor tissue was quantified as $\Delta\rho = \rho_{\text{tumor}} - \rho_{\text{normal}}$, computed from the pooled values.

The null distribution for $\Delta\rho$ was established using 50 control pairs formed by crossing five model genes (NUMB, MFN1, OPA1, BAX, APEX1) with ten housekeeping genes selected for stable expression across tissues (GAPDH, ACTB, TUBB, RPL13A, HPRT1, SDHA, YWHAZ, UBC, TBP, HMBS). The control null distribution has mean absolute $\Delta\rho$ of 0.124 and 95th percentile of 0.310.

The Class I versus Class II separation was tested using the two-sided Mann-Whitney U statistic on absolute $\Delta\rho$ values, with the one-sided alternative hypothesis that Class I values exceed Class II values. The bootstrap confidence interval on the mean difference was computed using 10,000 resamples with replacement from each class independently, computing the mean difference for each resample, and taking the 2.5th and 97.5th percentiles to define the 95 percent interval. The permutation test for the overall model-pair signal (all 54 model pairs versus 50 controls) used 10,000 random label shuffles, computing the mean absolute $\Delta\rho$ for the model-labeled group in each shuffle and comparing the observed value to the permutation distribution.

For single-cell analyses, pseudo-bulks were constructed by summing raw UMI counts (10x data) or TPM values (SmartSeq2 data) across all cells belonging to the same combination of patient, tissue type (tumor or normal), and cell compartment (epithelial, as determined by the original authors' cell-type annotations). Pseudo-bulks containing fewer than 30 cells were excluded from analysis to ensure stable average expression estimates. The summed counts or TPM values were normalized by counts per million (CPM) scaling: each gene's summed value was divided by the total sum across all genes for that pseudo-bulk and multiplied by 10^6 . The CPM-normalized

values were then $\log_{1p}$ -transformed ($\log(1 + \text{CPM})$). Spearman rank correlations were computed across pseudo-bulks within each condition (normal or tumor) separately.

Execution-pair and commitment-pair mean coupling values reported in the Abstract, §6.6, §6.7, and Figure 4 are computed across explicit pair sets defined on functional grounds. The execution-pair mean uses the 3-pair fork-execution set: $\text{CLSPN} \times \text{PCNA}$, $\text{DNMT1} \times \text{PCNA}$, and $\text{PLK1} \times \text{AURKA}$ — three structurally obligate Class II partnerships at the replication fork and at the mitotic spindle whose coupling reflects the obligate machinery requirement for cell division. The commitment-pair mean uses the 4-pair write/erase set: $\text{TDG} \times \text{EZH2}$, $\text{OGG1} \times \text{BAX}$, $\text{TDG} \times \text{BAX}$, and $\text{OGG1} \times \text{EZH2}$ — four pairs spanning the $\text{C} \rightarrow \text{T}$ (TDG-mediated) and $\text{G} \rightarrow \text{A}$ (OGG1-mediated) subsystems of the §3.3 write/erase matrix coupled to the EZH2-mediated chromatin and BAX-mediated death arms. The same pair sets are applied across all cell types where the constituent values are available in the deposited Census pseudo-bulk data (census_pseudobulk_validation.json and hsc_hierarchy_routing.json). The HSC and classical-monocyte hierarchy compartments contain all seven pair values; values for other hematopoietic cell types in §6.7 use only the pairs reported in the deposited JSON files for that cell type.

9.5 Pelka 2021 Data Extraction Pipeline

The Pelka 2021 h5 file (GSE178341_crc10x_full_c295v4_submit.h5; 1.2 GB on disk) stores its sparse expression matrix (CSC format: 764,460,511 non-zero entries in a 27,932 genes $\times$ 370,115 cells matrix) with a shuffle-then-deflate compression pipeline applied to single monolithic chunks. The matrix/indices dataset (764,460,511 int32 entries) is stored as one gzip-compressed chunk of 746 MB. The matrix/data dataset (764,460,511 float64 entries) is stored as two gzip-compressed chunks of 314 MB and 141 MB respectively. Standard h5py access

decompresses each entire chunk into memory (2.9 GB for indices, 5.8 GB for data), exceeding the available 3.9 GB RAM.

Because the Pelka dataset stores its 764-million-entry sparse matrix as single gzip-compressed chunks with a shuffle filter, conventional h5py access exceeds available memory. A custom disk-streaming pipeline was implemented: raw zlib stream decompression to temporary files, numpy-vectorized byte-plane unshuffle via memory-mapped access, and targeted extraction of the 49-gene panel across 168,295 epithelial cells grouped into 100 pseudo-bulk units (patient $\times$ tissue type combinations with ≥ 30 cells each). The pipeline produces results identical to standard sparse-matrix extraction on random test subsets. Normalization uses CPM with log1p transformation; correlations are Spearman across pseudo-bulks within each condition. Full implementation details and validation are available in the deposited analysis scripts (06_pelka_preprocessing.py, 07_pelka_pseudobulk_mmr.py).

The pipeline's correctness was validated by extracting a random 1,000-cell subset using standard h5py access (which succeeds for small subsets) and comparing the resulting expression values to those produced by the disk-based pipeline for the same cells. All values matched to floating-point precision.

9.6 Pre-Specified Criteria

The four criteria tested in Part 6, with their quantitative thresholds, are stated in §5.5. The criteria are: (1) NUMB $\times$ marking-arm coupling should be stronger in cancer than normal in epithelial-resolved analyses; (2) NUMB $\times$ polarity coupling should approach zero in cancer ($|\rho| < 0.05$ at protein level); (3) at least two inter-layer sign reversals at $|\Delta\rho| > 0.3$; and (4) asymmetric-layer coupling should decrease while symmetric-layer coupling is preserved.

9.7 Software and Reproducibility

All analyses were implemented in Python 3.12.3 using the following packages with version numbers at time of final execution: numpy 2.4.4 (array operations and linear algebra), scipy 1.17.1 (scipy.stats.spearmanr for Spearman correlation and scipy.stats.mannwhitneyu for Mann-Whitney U tests), h5py 3.16.0 (HDF5 file access for standard operations), requests 2.32.3 (HTTP API access for UCSC Xena and cBioPortal), and the Python standard library modules zlib (raw decompression), mmap (memory-mapped file access), gzip (compressed file reading), csv (metadata parsing), and json (results serialization). All bootstrap and permutation analyses used `numpy.random.seed(42)` for reproducibility. No proprietary software, commercial databases, or non-public data were used.

The analysis scripts and their output files are as follows: `03_class_separation_validation.py` produces `final_validation_results.json` (28+8 pair Δp values, class separation statistics, Mann-Whitney and bootstrap); `01_tcga_paired_normal_correlations.py` produces `tcga_paired_normal_results.json` (TCGA Blocks A through C); `02_tcga_extended_battery.py` produces `tcga_extended_battery_results.json` (Blocks D and E plus CLSPN correction); `04_layer_decoupling_analysis.py` produces `criterion4_entanglement.json` (layer-level coupling analysis); `06_pelka_preprocessing.py` produces the cell-to-group mapping files; `07_pelka_pseudobulk_mmr.py` produces `pelka_pseudobulk_results.json` and `pelka_mmr_stratification.json`; `08_census_singlecell_validation.py` produces `census_singlecell_validation.json`, `census_pseudobulk_validation.json`, and `hsc_hierarchy_routing.json`; `05_mode_test_lihc_prad_gbm.py` produces `mode_test_results.json`; `09_g4_tfbs_analysis.py` produces `s2_g4_tfbs_analysis.json` and `s2_g4_tfbs_reproduced.json` (Supplementary Section S2 four-arm G4 TFBS enrichment); and the mouse cross-species

analysis produces mouse_cross_species_results.json. All JSON files contain the complete numerical results with sufficient precision for independent verification.

9.8 Analytical Limitations

Several analytical choices warrant explicit acknowledgment. The fifty control pairs use five model genes (NUMB, MFN1, OPA1, BAX, APEX1) crossed with ten housekeeping genes, meaning the controls are not fully independent of the model. An independent control analysis using 22 non-model regulatory pairs and 15 non-model execution pairs (§6.2) confirms that the Class I/II separation is not an artifact of gene selection: independent replication-essential execution pairs (mean $|\Delta\rho| = 0.065$) show the same preservation as the model's Class II (0.0396), while independent regulatory pairs (0.170) match the model's Class I (0.1815). The 22 independent regulatory pairs are kinase-substrate cascades and transcription factor programs with zero gene overlap with the model panel: EGFR $\times$ SOS1, BRAF $\times$ MAP2K1, AKT1 $\times$ MTOR, JAK2 $\times$ STAT3, MAPK1 $\times$ ELK1, PIK3CA $\times$ AKT1, KRAS $\times$ BRAF, ABL1 $\times$ CRK, FGFR1 $\times$ FRS2, PDGFRA $\times$ PLCG1, IGF1R $\times$ IRS1, RUNX1 $\times$ CBFB, FOXO3 $\times$ CDKN1A, TP53 $\times$ CDKN1A, NOTCH1 $\times$ HES1, SMAD2 $\times$ SMAD4, TGFB1 $\times$ SMAD2, HIF1A $\times$ VEGFA, NFE2L2 $\times$ HMOX1, RELA $\times$ NFKBIA, JUN $\times$ FOS, SOX2 $\times$ NANOG. The 6 strict replication-essential execution pairs are: MCM2 $\times$ MCM5, MCM3 $\times$ MCM7, RFC2 $\times$ RFC5, POLR2A $\times$ POLR2B, NDC80 $\times$ NUF2, SMC1A $\times$ SMC3. Nine additional tissue-biology-affected physical complexes (proteasome, nuclear pore, chaperonin, adherens junction, COPI coat, translocon, FACT complex, ATP synthase subunits) served as a third comparison class. This independent analysis uses a single cohort (BRCA TCGA), while the model's values are pooled across nine cohorts; multi-cohort pooling of the independent pairs would provide a more rigorous comparison. The Fisher z-transform pooling assumes cohort independence, which is

approximately true for different cancer types but not perfectly so given the shared TCGA platform and processing pipeline. The 30-cell minimum for pseudo-bulk construction is a practical threshold; sensitivity analyses at different thresholds (20, 50, 100 cells) would strengthen the finding's robustness. The 190 individual correlation tests are interpreted as a class-level pattern rather than as individually significant findings, and no multiple-testing correction is applied to individual $\Delta\rho$ values; the primary statistical test is the single Mann-Whitney comparison of Class I versus Class II absolute $\Delta\rho$ distributions. Four of nine Class II pairs involve a single hub gene (CLSPN); however, CLSPN pairs (mean $|\Delta\rho| = 0.037$) and non-CLSPN Class II pairs (mean $|\Delta\rho| = 0.042$) show virtually identical disruption levels, and the finding remains significant when all CLSPN pairs are removed ($p = 1.1 \times 10^{-3}$ for 28 Class I versus 4 non-CLSPN Class II).

9.9 G4 Transcription Factor Binding Site Analysis

The transcription factor binding site enrichment analysis reported in §3.3.3 was conducted on 30 G4-containing promoters selected from genes with experimentally validated G4 structures (defined as sequences with G4Hunter score > 1.5 within 1 kb upstream of the annotated transcription start site). For each G4-containing sequence, proto-binding sites for six transcription factor families (p53, NF- κ B, ETS, YY1, KLF4, MAZ) were identified using published consensus motifs from the JASPAR database. GC-matched null models were constructed by generating 10,000 random sequences per gene, matched for length, GC content, and dinucleotide frequency, and scanning each for the same TF motifs. Enrichment was computed as the ratio of observed proto-site density in real G4 sequences to the mean density in null sequences, with p-values derived from the empirical null distribution. All analysis scripts and intermediate data are deposited with this manuscript.

9.10 Summary of Key Statistical Results

The primary statistical finding — the Mann-Whitney U test comparing Class I versus Class II absolute $\Delta\rho$ — is reported with the following complete statistics: $U = 212$, $p = 8.53 \times 10^{-6}$ (conservative, 28 versus 8), effect size (mean difference) = 0.117 (Class I mean $|\Delta\rho| = 0.1815$, Class II mean $|\Delta\rho| = 0.0396$). For the core-only analysis: $U = 160$, $p = 3.22 \times 10^{-7}$ (20 versus 8), effect size = 0.139. Bootstrap 95% CI on mean difference: [0.1035, 0.1796] (conservative scenario). Leave-one-out: worst-case $p = 1.1 \times 10^{-5}$. Independent control validation (BRCA TCGA, single cohort): independent regulatory pairs mean $|\Delta\rho| = 0.170$ ($n = 22$), independent replication-essential execution pairs mean $|\Delta\rho| = 0.065$ ($n = 6$), tissue-biology-affected complexes mean $|\Delta\rho| = 0.315$ ($n = 9$; $p = 8 \times 10^{-4}$, Mann-Whitney U one-sided, versus replication-essential). CLSPN dominance: CLSPN Class II mean $|\Delta\rho| = 0.037$, non-CLSPN Class II mean $|\Delta\rho| = 0.042$, finding survives complete CLSPN removal ($p = 1.1 \times 10^{-3}$). All individual Spearman ρ values, per-cohort sample sizes, and $\Delta\rho$ values for all 61 model pairs and 50 control pairs are provided in the deposited JSON files (§9.7). No multiple-testing correction is applied to individual $\Delta\rho$ values because the primary test is the single class-level Mann-Whitney comparison; individual pair results are reported as pattern components, not as independently significant findings.

9.11 Data Availability

All data used in this study are publicly available from the following repositories. TCGA HiSeqV2 RNA-seq expression data: UCSC Xena (<https://tcga.xenahubs.net>), dataset TCGA..sampleMap/HiSeqV2 for each of nine cancer types. CPTAC protein quantification: cBioPortal for Cancer Genomics (<https://www.cbioportal.org/api>), studies brca_cptac_2020, coad_cptac_2019, gbm_cptac_2021, luad_cptac_2020, lusc_cptac_2021, paad_cptac_2021,

ucec_cptac_2020, molecular profile suffix _protein_quantification. Single-cell RNA-seq: NCBI GEO accession GSE132465 (Lee et al. 2020), GSE178341 (Pelka et al. 2021), GSE92332 (Haber et al. 2017), GSE116256 (van Galen et al. 2019), GSE103322 (Puram et al. 2017). Cross-tissue and hematopoietic hierarchy pseudo-bulk data: CZI CELLxGENE Census (version 2024-07-01; <https://chanzuckerberg.github.io/cellxgene-census/>), accessed via the cellxgene-census Python package (v1.17.0). No proprietary data, commercial databases, or non-public datasets were used.

All quantitative values in Parts 6–8 trace to deposited JSON files produced by deposited scripts.

The pipeline is: Script → JSON → Manuscript value. The mapping is as follows. §6.1

permutation null: permutation_null_test.json. §6.2 primary results (28 Class I + 8 Class II Δp

values): final_validation_results.json. §6.3 pre-specified criteria: final_validation_results.json

(class_level_stats). §6.3 layer decoupling: criterion4_entanglement.json. §6.4 BER collapse and

fork protection: final_validation_results.json and insulator_arm_results.json. §6.5.1 HNSC

single-cell: hnscc_puram_results.json. §6.5.2 AML quiescent evasion:

aml_singlecell_cycling.json (single-cell) and aml_laml_bulk_correlations.json (bulk). §6.5.3–

6.5.5 mode tests (LIHC, PRAD, GBM): mode_test_results.json. §6.6 cross-tissue:

census_singlecell_validation.json and census_pseudobulk_validation.json. §6.7 hematopoietic

hierarchy: hsc_hierarchy_routing.json. §6.8 PIDD1 network: pidd1_network_results.json. §6.9

residualized: residualized_results.json. Pelka pseudo-bulk: pelka_pseudobulk_results.json. Pelka

MMR stratification: pelka_mmr_stratification.json. Mouse cross-species:

mouse_cross_species_results.json. §6.2 independent controls: independent_control_results.json.

§6.3 CPTAC protein polarity: cptac_protein_polarity_dissolution.json. §6.3 Lee 2020 epithelial

pseudo-bulk: lee2020_epithelial_pseudobulk.json. §6.5.2 AML hierarchy CEBPA correlations:

aml_hierarchy_correlations.json. §6.6 additional Census tests: census_additional_tests.json. §6.8 tumor purity proxy: purity_proxy_and_controls.json. All deposited JSON files use a data freeze date of 2026-05-22 (residualized_results.json: 2026-05-23). Source data for all data figures are provided as CSV files (supplementary_class_separation.csv, figure3_fidelity_index.csv, figure5_mode_tests.csv, figure5_aml_cycling.csv, figure4_gradient.csv, figure4_trajectory.csv, figure4_routing.csv).

9.12 Code Availability

All analysis scripts and their output JSON files are available at

<https://zenodo.org/records/20402773>. The repository includes:

01_tcga_paired_normal_correlations.py through 09_g4_tfbs_analysis.py (primary analyses), 10_independent_control_analysis.py through 16_aml_cycling_analysis.py (extended validations), 17_census_additional_tests.py through 22_pidd1_network_analysis.py (network and cross-species analyses), and verify_manuscript_values.py (264-check programmatic verification), along with all 25 JSON output files sufficient for independent verification of every quantitative claim in the manuscript. The code requires only Python 3.12 with numpy, scipy, h5py, and requests; no proprietary software is needed. A README documents the execution order and expected outputs.

9.13 Statistical Methods Summary

All correlations are Spearman rank (scipy.stats.spearmanr). The primary statistical test is the class-level Mann-Whitney comparison; individual pair correlations are descriptive. Where multiple testing applies across the 190-test battery, the Bonferroni threshold is $\alpha = 2.6 \times 10^{-4}$.

Sample sizes are reported at each point of use. No data points were excluded. Gene symbols follow HGNC nomenclature.

Competing Interests

The author declares no competing interests.

Author Contributions

D.M. conceived the model, designed and executed all computational analyses, and wrote the manuscript.

Ethics Statement

This study uses only publicly available, de-identified data from established repositories (TCGA, CPTAC, GEO, CZI CELLxGENE Census). No new human or animal data were generated. No institutional review board approval was required.

AI Use Disclosure

This manuscript was developed with the assistance of large language models (Anthropic Claude) used as interactive research tools throughout the project. The AI assisted with literature synthesis, computational analysis design and execution, data verification, manuscript drafting and revision, and citation auditing. All scientific hypotheses, model architecture, interpretive claims, and analytical decisions were made by the author. All quantitative values were computed from public data sources using deposited scripts and verified against deposited JSON output files. The AI did not independently generate any empirical claims; all values trace to the Script → JSON → Manuscript pipeline documented in §9.11. The author takes full responsibility for the accuracy and integrity of all content.

Acknowledgments

This work was conducted without institutional affiliation or external funding. The author acknowledges the TCGA Research Network (<https://www.cancer.gov/tcga>) for generating the bulk RNA-seq data; the Clinical Proteomic Tumor Analysis Consortium (CPTAC) for generating the proteomic data; the UCSC Xena project for hosting TCGA expression data; cBioPortal for Cancer Genomics for providing API access to TCGA and CPTAC datasets; the Chan Zuckerberg Initiative and the CZI CELLxGENE Census team for providing unified access to single-cell RNA-seq atlases; the NCBI Gene Expression Omnibus for hosting single-cell datasets; the HUGO Gene Nomenclature Committee (HGNC) for standardized gene nomenclature; and the original investigators of the single-cell datasets — Lee et al. (2020), Pelka et al. (2021), Haber et al. (2017), van Galen et al. (2019), and Puram et al. (2017) — for generating and depositing the publicly available data that made this analysis possible. Figures 1–6 were generated with the assistance of ChatGPT (OpenAI) and refined through iterative accuracy review. Computational analysis assistance, manuscript preparation, and verification tooling were provided by Claude (Anthropic).

References

- Agger, K., Cloos, P. A., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A. E., & Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature*, *449*(7163), 731–734. <https://doi.org/10.1038/nature06145>
- Aghili, L., Foo, J., DeGregori, J., & De, S. (2014). Patterns of somatically acquired amplifications and deletions in apparently normal tissues of ovarian cancer patients. *Cell reports*, *7*(4), 1310–1319. <https://doi.org/10.1016/j.celrep.2014.03.071>
- Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J. A., Greaves, M., ... Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415–421. <https://doi.org/10.1038/nature12477>
- Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature reviews. Genetics*, *17*(8), 487–500. <https://doi.org/10.1038/nrg.2016.59>
- Aspinwall, R., Rothwell, D. G., Roldan-Arjona, T., Anselmino, C., Ward, C. J., Cheadle, J. P., Sampson, J. R., Lindahl, T., Harris, P. C., & Hickson, I. D. (1997). Cloning and characterization of a functional human homolog of Escherichia coli endonuclease III. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(1), 109–114. <https://doi.org/10.1073/pnas.94.1.109>
- Ba, X., Bacsí, A., Luo, J., Aguilera-Aguirre, L., Zeng, X., Radak, Z., Brasier, A. R., & Boldogh, I. (2014). 8-oxoguanine DNA glycosylase-1 augments proinflammatory gene expression by facilitating the recruitment of site-specific transcription factors. *Journal of immunology (Baltimore, Md. : 1950)*, *192*(5), 2384–2394. <https://doi.org/10.4049/jimmunol.1302472>
- Barbot, M., Jans, D. C., Schulz, C., Denkert, N., Kroppen, B., Hoppert, M., Jakobs, S., & Meinecke, M. (2015). Mic10 oligomerizes to bend mitochondrial inner membranes at cristae junctions. *Cell metabolism*, *21*(5), 756–763. <https://doi.org/10.1016/j.cmet.2015.04.006>
- Barker, N., van Es, J. H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P. J., & Clevers, H. (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature*, *449*(7165), 1003–1007. <https://doi.org/10.1038/nature06196>
- Beard, W. A., Horton, J. K., Prasad, R., & Wilson, S. H. (2019). Eukaryotic Base Excision Repair: New Approaches Shine Light on Mechanism. *Annual review of biochemistry*, *88*, 137–162. <https://doi.org/10.1146/annurev-biochem-013118-111315>
- Bedeckovics, T., Hussain, S., Zhang, Y., Ali, A., Jeon, Y. J., & Galaray, P. J. (2021). *USP24* Is a Cancer-Associated Ubiquitin Hydrolase, Novel Tumor Suppressor, and Chromosome Instability Gene Deleted in Neuroblastoma. *Cancer research*, *81*(5), 1321–1331. <https://doi.org/10.1158/0008-5472.CAN-20-1777>
- Bell, R. J., Rube, H. T., Kreig, A., Mancini, A., Fouse, S. D., Nagarajan, R. P., Choi, S., Hong, C., He, D., Pekmezci, M., Wiencke, J. K., Wensch, M. R., Chang, S. M., Walsh, K. M., Myong, S., Song, J. S., & Costello, J. F. (2015). Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science (New York, N.Y.)*, *348*(6238), 1036–1039. <https://doi.org/10.1126/science.aab0015>
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., & Lander, E. S. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, *125*(2), 315–326. <https://doi.org/10.1016/j.cell.2006.02.041>
- Betschinger, J., & Knoblich, J. A. (2004). Dare to be different: asymmetric cell division in Drosophila, C. elegans and vertebrates. *Current biology : CB*, *14*(16), R674–R685. <https://doi.org/10.1016/j.cub.2004.08.017>
- 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>

- Bonnet, D., & Dick, J. E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature medicine*, *3*(7), 730–737. <https://doi.org/10.1038/nm0797-730>
- Brosh, R. M., Jr, & Bohr, V. A. (2007). Human premature aging, DNA repair and RecQ helicases. *Nucleic acids research*, *35*(22), 7527–7544. <https://doi.org/10.1093/nar/gkm1008>
- Bunel, B., Leclercq, R., Goïame, R., Gautier, A., Morin, X., & Fischer, E. (2025). Unequal mitochondrial segregation promotes asymmetric fates during neurogenesis. *Nature communications*, *16*(1), 11049. <https://doi.org/10.1038/s41467-025-66932-0>
- Bunting, S. F., Callén, E., Kozak, M. L., Kim, J. M., Wong, N., López-Contreras, A. J., Ludwig, T., Baer, R., Faryabi, R. B., Malhowski, A., Chen, H. T., Fernandez-Capetillo, O., D'Andrea, A., & Nussenzweig, A. (2012). BRCA1 functions independently of homologous recombination in DNA interstrand crosslink repair. *Molecular cell*, *46*(2), 125–135. <https://doi.org/10.1016/j.molcel.2012.02.015>
- 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>
- Burns, M. B., Temiz, N. A., & Harris, R. S. (2013). Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nature genetics*, *45*(9), 977–983. <https://doi.org/10.1038/ng.2701>
- Cadet, J., & Wagner, J. R. (2013). DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. *Cold Spring Harbor perspectives in biology*, *5*(2), a012559. <https://doi.org/10.1101/cshperspect.a012559>
- Cairns, J., Overbaugh, J., & Miller, S. (1988). The origin of mutants. *Nature*, *335*(6186), 142–145. <https://doi.org/10.1038/335142a0>
- Candelas, A., Vianay, B., Gelin, M., Faivre, L., Larghero, J., Blanchoin, L., Théry, M., & Brunet, S. (2024). Heterotypic interaction promotes asymmetric division of human hematopoietic progenitors. *Development (Cambridge, England)*, *151*(17), dev203088. <https://doi.org/10.1242/dev.203088>
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., & Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science (New York, N.Y.)*, *298*(5595), 1039–1043. <https://doi.org/10.1126/science.1076997>
- Carroll S. B. (2005). Evolution at two levels: on genes and form. *PLoS biology*, *3*(7), e245. <https://doi.org/10.1371/journal.pbio.0030245>
- Causinus, E., & Gonzalez, C. (2005). Induction of tumor growth by altered stem-cell asymmetric division in *Drosophila melanogaster*. *Nature genetics*, *37*(10), 1125–1129. <https://doi.org/10.1038/ng1632>
- Ceccaldi, R., Sarangi, P., & D'Andrea, A. D. (2016). The Fanconi anaemia pathway: new players and new functions. *Nature reviews. Molecular cell biology*, *17*(6), 337–349. <https://doi.org/10.1038/nrm.2016.48>
- Charlton, S. J., Flury, V., Kanoh, Y., Genzor, A. V., Kollenstart, L., Ao, W., Brögger, P., Weisser, M. B., Adamus, M., Alcaraz, N., Delvaux de Fenffe, C. M., Mattioli, 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>
- Chen, H., Detmer, S. A., Ewald, A. J., Griffin, E. E., Fraser, S. E., & Chan, D. C. (2003). Mitofusins Mfn1 and Mfn2 coordinately regulate mitochondrial fusion and are essential for embryonic development. *The Journal of cell biology*, *160*(2), 189–200. <https://doi.org/10.1083/jcb.200211046>
- Cimmino, L., Dolgalev, I., Wang, Y., Yoshimi, A., Martin, G. H., Wang, J., Ng, V., Xia, B., Witkowski, M. T., Mitchell-Flack, M., Grillo, I., Bakogianni, S., Ndiaye-Lobry, D., Martín, M. T., Guillaumot, M., Banh, R. S., Xu, M., Figueroa, M. E., Dickins, R. A., Abdel-Wahab, O., ... Aifantis, I. (2017). Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. *Cell*, *170*(6), 1079–1095.e20. <https://doi.org/10.1016/j.cell.2017.07.032>
- 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>

- Cipolat, S., Martins de Brito, O., Dal Zilio, B., & Scorrano, L. (2004). OPA1 requires mitofusin 1 to promote mitochondrial fusion. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(45), 15927–15932. <https://doi.org/10.1073/pnas.0407043101>
- Clevers, H., & Nusse, R. (2012). Wnt/ β -catenin signaling and disease. *Cell*, *149*(6), 1192–1205. <https://doi.org/10.1016/j.cell.2012.05.012>
- Cogliati, S., Frezza, C., Soriano, M. E., Varanita, T., Quintana-Cabrera, R., Corrado, M., Cipolat, S., Costa, V., Casarin, A., Gomes, L. C., Perales-Clemente, E., Salviati, L., Fernandez-Silva, P., Enriquez, J. A., & Scorrano, L. (2013). Mitochondrial cristae shape determines respiratory chain supercomplexes assembly and respiratory efficiency. *Cell*, *155*(1), 160–171. <https://doi.org/10.1016/j.cell.2013.08.032>
- Cogoi, S., & Xodo, L. E. (2006). G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic acids research*, *34*(9), 2536–2549. <https://doi.org/10.1093/nar/gkl286>
- Colaluca, I. N., Tosoni, D., Nuciforo, P., Senic-Matuglia, F., Galimberti, V., Viale, G., Pece, S., & Di Fiore, P. P. (2008). NUMB controls p53 tumour suppressor activity. *Nature*, *451*(7174), 76–80. <https://doi.org/10.1038/nature06412>
- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., Abramowitz, L. K., Bartolomei, M. S., Rambow, F., Bassi, M. R., Bruno, T., Fanciulli, M., Renner, C., Klein-Szanto, A. J., Matsumoto, Y., Kobi, D., ... Bellacosa, A. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, *146*(1), 67–79. <https://doi.org/10.1016/j.cell.2011.06.020>
- Crasta, K., Ganem, N. J., Dagher, R., Lantermann, A. B., Ivanova, E. V., Pan, Y., Nezi, L., Protopopov, A., Chowdhury, D., & Pellman, D. (2012). DNA breaks and chromosome pulverization from errors in mitosis. *Nature*, *482*(7383), 53–58. <https://doi.org/10.1038/nature10802>
- Croteau, D. L., Popuri, V., Opresko, P. L., & Bohr, V. A. (2014). Human RecQ helicases in DNA repair, recombination, and replication. *Annual review of biochemistry*, *83*, 519–552. <https://doi.org/10.1146/annurev-biochem-060713-035428>
- Curtis, P. D., & Brun, Y. V. (2010). Getting in the loop: regulation of development in *Caulobacter crescentus*. *Microbiology and molecular biology reviews : MMBR*, *74*(1), 13–41. <https://doi.org/10.1128/MMBR.00040-09>
- Dahal, S., Siddiqua, H., Sharma, S., Babu, R. K., Rathore, D., Sharma, S., & Raghavan, S. C. (2022). Unleashing a novel function of Endonuclease G in mitochondrial genome instability. *eLife*, *11*, e69916. <https://doi.org/10.7554/eLife.69916>
- Dancy, B. M., & Cole, P. A. (2015). Protein lysine acetylation by p300/CBP. *Chemical reviews*, *115*(6), 2419–2452. <https://doi.org/10.1021/cr500452k>
- Dang, L., White, D. W., Gross, S., Bennett, B. D., Bittinger, M. A., Driggers, E. M., Fantin, V. R., Jang, H. G., Jin, S., Keenan, M. C., Marks, K. M., Prins, R. M., Ward, P. S., Yen, K. E., Liao, L. M., Rabinowitz, J. D., Cantley, L. C., Thompson, C. B., Vander Heiden, M. G., & Su, S. M. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*, *462*(7274), 739–744. <https://doi.org/10.1038/nature08617>
- de Bono, J., Mateo, J., Fizazi, K., Saad, F., Shore, N., Sandhu, S., Chi, K. N., Sartor, O., Agarwal, N., Olmos, D., Thiery-Vuillemin, A., Twardowski, P., Mehra, N., Goessl, C., Kang, J., Burgents, J., Wu, W., Kohlmann, A., Adelman, C. A., & Hussain, M. (2020). Olaparib for Metastatic Castration-Resistant Prostate Cancer. *The New England journal of medicine*, *382*(22), 2091–2102. <https://doi.org/10.1056/NEJMoa1911440>
- Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & development*, *25*(10), 1010–1022. <https://doi.org/10.1101/gad.2037511>
- Delettre, C., Lenaers, G., Griffoin, J. M., Gigarel, N., Lorenzo, C., Belenguer, P., Pelloquin, L., Grosgeorge, J., Turc-Carel, C., Perret, E., Astarie-Dequeker, C., Lasquelléc, L., Arnaud, B., Ducommun, B., Kaplan, J., & Hamel, C. P. (2000). Nuclear gene OPA1, encoding a mitochondrial dynamin-related protein, is mutated in dominant optic atrophy. *Nature genetics*, *26*(2), 207–210. <https://doi.org/10.1038/79936>
- Denecker, G., Hoste, E., Gilbert, B., Hocheplé, T., Ovaere, P., Lippens, S., Van den Broecke, C., Van Damme, P., D'Herde, K., Hachem, J. P., Borgonie, G., Presland, R. B., Schoonjans, L., Libert, C., Vandekerckhove, J.,

- Gevaert, K., Vandenabeele, P., & Declercq, W. (2007). Caspase-14 protects against epidermal UVB photodamage and water loss. *Nature cell biology*, *9*(6), 666–674. <https://doi.org/10.1038/ncb1597>
- Di Noia, J. M., & Neuberger, M. S. (2007). Molecular mechanisms of antibody somatic hypermutation. *Annual review of biochemistry*, *76*, 1–22. <https://doi.org/10.1146/annurev.biochem.76.061705.090740>
- DiNardo, C. D., Stein, E. M., de Botton, S., Roboz, G. J., Altman, J. K., Mims, A. S., Swords, R., Collins, R. H., Mannis, G. N., Pollyea, D. A., Donnellan, W., Fathi, A. T., Pigneux, A., Erba, H. P., Prince, G. T., Stein, A. S., Uy, G. L., Foran, J. M., Traer, E., Stuart, R. K., ... Kantarjian, H. M. (2018). Durable Remissions with Ivosidenib in IDH1-Mutated Relapsed or Refractory AML. *The New England journal of medicine*, *378*(25), 2386–2398. <https://doi.org/10.1056/NEJMoa1716984>
- Durrett, R., & Schmidt, D. (2008). Waiting for two mutations: with applications to regulatory sequence evolution and the limits of Darwinian evolution. *Genetics*, *180*(3), 1501–1509. <https://doi.org/10.1534/genetics.107.082610>
- Döhla, J., Kuuluvainen, E., Gebert, N., Amaral, A., Englund, J. I., Gopalakrishnan, S., Konovalova, S., Nieminen, A. I., Salminen, E. S., Torregrosa Muñumer, R., Ahlqvist, K., Yang, Y., Bui, H., Otonkoski, T., Käkälä, R., Hietakangas, V., Tyynismaa, H., Ori, A., & Katajisto, P. (2022). Metabolic determination of cell fate through selective inheritance of mitochondria. *Nature cell biology*, *24*(2), 148–154. <https://doi.org/10.1038/s41556-021-00837-0>
- Eddy, J., & Maizels, N. (2006). Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic acids research*, *34*(14), 3887–3896. <https://doi.org/10.1093/nar/gkl529>
- Eminli, S., Foudi, A., Stadtfeld, M., Maherali, N., Ahfeldt, T., Mostoslavsky, G., Hock, H., & Hochedlinger, K. (2009). Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nature genetics*, *41*(9), 968–976. <https://doi.org/10.1038/ng.428>
- Ertürk, A., Wang, Y., & Sheng, M. (2014). Local pruning of dendrites and spines by caspase-3-dependent and proteasome-limited mechanisms. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *34*(5), 1672–1688. <https://doi.org/10.1523/JNEUROSCI.3121-13.2014>
- Evans, L. T., Anglen, T., Scott, P., Lukasik, K., Loncarek, J., & Holland, A. J. (2021). ANKRD26 recruits PIDD1 to centriolar distal appendages to activate the PIDDosome following centrosome amplification. *The EMBO journal*, *40*(4), e105106. <https://doi.org/10.15252/embj.2020105106>
- 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](https://doi.org/10.1016/0092-8674(90)90186-i)
- Fernandez-Vidal, A., Vignard, J., & Mirey, G. (2017). Around and beyond 53BP1 Nuclear Bodies. *International journal of molecular sciences*, *18*(12), 2611. <https://doi.org/10.3390/ijms18122611>
- Fernando, P., Brunette, S., & Megeney, L. A. (2005). Neural stem cell differentiation is dependent upon endogenous caspase 3 activity. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, *19*(12), 1671–1673. <https://doi.org/10.1096/fj.04-2981fje>
- Fernando, P., Kelly, J. F., Balazsi, K., Slack, R. S., & Megeney, L. A. (2002). Caspase 3 activity is required for skeletal muscle differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(17), 11025–11030. <https://doi.org/10.1073/pnas.162172899>
- Fisher, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron*, *1*, 3–32.
- Fleming, A. M., & Burrows, C. J. (2017). 8-Oxo-7,8-dihydroguanine, friend and foe: Epigenetic-like regulator versus initiator of mutagenesis. *DNA repair*, *56*, 75–83. <https://doi.org/10.1016/j.dnarep.2017.06.009>
- Foti, R., Gnan, S., Cornacchia, D., Dileep, V., Bulut-Karslioglu, A., Diehl, S., Buness, A., Klein, F. A., Huber, W., Johnstone, E., Loos, R., Bertone, P., Gilbert, D. M., Manke, T., Jenuwein, T., & Buonomo, S. C. (2016). Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing Program. *Molecular cell*, *61*(2), 260–273. <https://doi.org/10.1016/j.molcel.2015.12.001>

- Frezza, C., Cipolat, S., Martins de Brito, O., Micaroni, M., Beznoussenko, G. V., Rudka, T., Bartoli, D., Polishuck, R. S., Danial, N. N., De Strooper, B., & Scorrano, L. (2006). OPA1 controls apoptotic cristae remodeling independently from mitochondrial fusion. **Cell**, **126**(1), 177–189. <https://doi.org/10.1016/j.cell.2006.06.025>
- Galluzzi, L., Vitale, I., Aaronson, S. A., Abrams, J. M., Adam, D., Agostinis, P., Alnemri, E. S., Altucci, L., Amelio, I., Andrews, D. W., Annicchiarico-Petruzzelli, M., Antonov, A. V., Arama, E., Baehrecke, E. H., Barlev, N. A., Bazan, N. G., Bernassola, F., Bertrand, M. J. M., Bianchi, K., Blagosklonny, M. V., ... Kroemer, G. (2018). Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018. **Cell death and differentiation**, **25**(3), 486–541. <https://doi.org/10.1038/s41418-017-0012-4>
- Gehring, K., Leroy, J. L., & Guéron, M. (1993). A tetrameric DNA structure with protonated cytosine-cytosine base pairs. **Nature**, **363**(6429), 561–565. <https://doi.org/10.1038/363561a0>
- Giansanti, M. G., Gatti, M., & Bonaccorsi, S. (2001). The role of centrosomes and astral microtubules during asymmetric division of *Drosophila* neuroblasts. **Development (Cambridge, England)**, **128**(7), 1137–1145. <https://doi.org/10.1242/dev.128.7.1137>
- Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., & Chédin, F. (2012). R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. **Molecular cell**, **45**(6), 814–825. <https://doi.org/10.1016/j.molcel.2012.01.017>
- Gurard-Levin, Z. A., Quivy, J. P., & Almouzni, G. (2014). Histone chaperones: assisting histone traffic and nucleosome dynamics. **Annual review of biochemistry**, **83**, 487–517. <https://doi.org/10.1146/annurev-biochem-060713-035536>
- Gönczy P. (2008). Mechanisms of asymmetric cell division: flies and worms pave the way. **Nature reviews. Molecular cell biology**, **9**(5), 355–366. <https://doi.org/10.1038/nrm2388>
- Götz, M., & Huttner, W. B. (2005). The cell biology of neurogenesis. **Nature reviews. Molecular cell biology**, **6**(10), 777–788. <https://doi.org/10.1038/nrm1739>
- Haber, A. L., Biton, M., Rogel, N., Herbst, R. H., Shekhar, K., Smillie, C., Burgin, G., Delorey, T. M., Howitt, M. R., Katz, Y., Tirosh, I., Beyaz, S., Dionne, D., Zhang, M., Raychowdhury, R., Garrett, W. S., Rozenblatt-Rosen, O., Shi, H. N., Yilmaz, O., Xavier, R. J., ... Regev, A. (2017). A single-cell survey of the small intestinal epithelium. **Nature**, **551**(7680), 333–339. <https://doi.org/10.1038/nature24489>
- Hahn, M. A., Qiu, R., Wu, X., Li, A. X., Zhang, H., Wang, J., Jui, J., Jin, S. G., Jiang, Y., Pfeifer, G. P., & Lu, Q. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. **Cell reports**, **3**(2), 291–300. <https://doi.org/10.1016/j.celrep.2013.01.011>
- 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>
- Haradhvala, N. J., Polak, P., Stojanov, P., Covington, K. R., Shinbrot, E., Hess, J. M., Rheinbay, E., Kim, J., Maruvka, Y. E., Braunstein, L. Z., Kamburov, A., Hanawalt, P. C., Wheeler, D. A., Koren, A., Lawrence, M. S., & Getz, G. (2016). Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. **Cell**, **164**(3), 538–549. <https://doi.org/10.1016/j.cell.2015.12.050>
- Hardeland, U., Kunz, C., Focke, F., Szadkowski, M., & Schär, P. (2007). Cell cycle regulation as a mechanism for functional separation of the apparently redundant uracil DNA glycosylases TDG and UNG2. **Nucleic acids research**, **35**(11), 3859–3867. <https://doi.org/10.1093/nar/gkm337>
- He, H., Wang, Y., Zhang, X., Li, X., Liu, C., Yan, D., Deng, H., Sun, W., Yi, C., & Wang, J. (2024). Age-related noncanonical TRMT6-TRMT61A signaling impairs hematopoietic stem cells. **Nature aging**, **4**(2), 213–230. <https://doi.org/10.1038/s43587-023-00556-1>
- He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C., & Xu, G. L. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. **Science (New York, N.Y.)**, **333**(6047), 1303–1307. <https://doi.org/10.1126/science.1210944>

- Hermann, A., Goyal, R., & Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *The Journal of biological chemistry*, *279*(46), 48350–48359. <https://doi.org/10.1074/jbc.M403427200>
- Hodgkinson, A., & Eyre-Walker, A. (2011). Variation in the mutation rate across mammalian genomes. *Nature reviews. Genetics*, *12*(11), 756–766. <https://doi.org/10.1038/nrg3098>
- Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G., & Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature*, *419*(6903), 135–141. <https://doi.org/10.1038/nature00991>
- Horn, S., Figl, A., Rachakonda, P. S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K., Schadendorf, D., & Kumar, R. (2013). TERT promoter mutations in familial and sporadic melanoma. *Science (New York, N.Y.)*, *339*(6122), 959–961. <https://doi.org/10.1126/science.1230062>
- Huang, S., Eichler, G., Bar-Yam, Y., & Ingber, D. E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Physical review letters*, *94*(12), 128701. <https://doi.org/10.1103/PhysRevLett.94.128701>
- Huppert, J. L., & Balasubramanian, S. (2005). Prevalence of quadruplexes in the human genome. *Nucleic acids research*, *33*(9), 2908–2916. <https://doi.org/10.1093/nar/gki609>
- Hänsel-Hertsch, R., Di Antonio, M., & Balasubramanian, S. (2017). DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nature reviews. Molecular cell biology*, *18*(5), 279–284. <https://doi.org/10.1038/nrm.2017.3>
- Ichim, G., Lopez, J., Ahmed, S. U., Muthalagu, N., Giampazolias, E., Delgado, M. E., Haller, M., Riley, J. S., Mason, S. M., Athineos, D., Parsons, M. J., van de Kooij, B., Bouchier-Hayes, L., Chalmers, A. J., Rooswinkel, R. W., Oberst, A., Blyth, K., Rehm, M., Murphy, D. J., & Tait, S. W. G. (2015). Limited mitochondrial permeabilization causes DNA damage and genomic instability in the absence of cell death. *Molecular cell*, *57*(5), 860–872. <https://doi.org/10.1016/j.molcel.2015.01.018>
- Inoue, A., Shen, L., Dai, Q., He, C., & Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell research*, *21*(12), 1670–1676. <https://doi.org/10.1038/cr.2011.189>
- Ito, K., Hirao, A., Arai, F., Matsuoka, S., Takubo, K., Hamaguchi, I., Nomiyama, K., Hosokawa, K., Sakurada, K., Nakagata, N., Ikeda, Y., Mak, T. W., & Suda, T. (2004). Regulation of oxidative stress by ATM is required for self-renewal of haematopoietic stem cells. *Nature*, *431*(7011), 997–1002. <https://doi.org/10.1038/nature02989>
- Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., & Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, *466*(7310), 1129–1133. <https://doi.org/10.1038/nature09303>
- Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., & Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (New York, N.Y.)*, *333*(6047), 1300–1303. <https://doi.org/10.1126/science.1210597>
- Iyer, L. M., Tahiliani, M., Rao, A., & Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell cycle (Georgetown, Tex.)*, *8*(11), 1698–1710. <https://doi.org/10.4161/cc.8.11.8580>
- 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>
- Katajisto, P., Döhla, J., Chaffer, C. L., Pentimikko, N., Marjanovic, N., Iqbal, S., Zoncu, R., Chen, W., Weinberg, R. A., & Sabatini, D. M. (2015). Stem cells. Asymmetric apportioning of aged mitochondria between daughter cells is required for stemness. *Science (New York, N.Y.)*, *348*(6232), 340–343. <https://doi.org/10.1126/science.1260384>
- Kayano, M., Shiga, M., & Mamitsuka, H. (2014). Detecting Differentially Coexpressed Genes from Labeled Expression Data: A Brief Review. *IEEE/ACM transactions on computational biology and bioinformatics*, *11*(1), 154–167. <https://doi.org/10.1109/TCBB.2013.2297921>

- King, M. C., & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science (New York, N.Y.)*, *188*(4184), 107–116. <https://doi.org/10.1126/science.1090005>
- Knoblich J. A. (2010). Asymmetric cell division: recent developments and their implications for tumour biology. *Nature reviews. Molecular cell biology*, *11*(12), 849–860. <https://doi.org/10.1038/nrm3010>
- 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>
- Kohli, R. M., & Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, *502*(7472), 472–479. <https://doi.org/10.1038/nature12750>
- Kollmann, K., Heller, G., Schneckenleithner, C., Warsch, W., Scheicher, R., Ott, R. G., Schäfer, M., Fajmann, S., Schleeder, M., Schiefer, A. I., Reichart, U., Mayerhofer, M., Hoeller, C., Zöchbauer-Müller, S., Kerjaschki, D., Bock, C., Kenner, L., Hoefler, G., Freissmuth, M., Green, A. R., ... Sexl, V. (2013). A kinase-independent function of CDK6 links the cell cycle to tumor angiogenesis. *Cancer cell*, *24*(2), 167–181. <https://doi.org/10.1016/j.ccr.2013.07.012>
- Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S. A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Wong, W. S., Sigurdsson, G., Walters, G. B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D. F., Helgason, A., Magnusson, O. T., Thorsteinsdottir, U., ... Stefansson, K. (2012). Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, *488*(7412), 471–475. <https://doi.org/10.1038/nature11396>
- Krasovec, G., Horkan, H. R., Quéinnec, É., & Chambon, J. P. (2023). Intrinsic apoptosis is evolutionarily divergent among metazoans. *Evolution letters*, *8*(2), 267–282. <https://doi.org/10.1093/evlett/qrad057>
- Krokan, H. E., & Bjørås, M. (2013). Base excision repair. *Cold Spring Harbor perspectives in biology*, *5*(4), a012583. <https://doi.org/10.1101/cshperspect.a012583>
- 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>
- Lange, C., Huttner, W. B., & Calegari, F. (2009). Cdk4/cyclinD1 overexpression in neural stem cells shortens G1, delays neurogenesis, and promotes the generation and expansion of basal progenitors. *Cell stem cell*, *5*(3), 320–331. <https://doi.org/10.1016/j.stem.2009.05.026>
- Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, *9*, 559. <https://doi.org/10.1186/1471-2105-9-559>
- Larsen, B. D., Rampalli, S., Burns, L. E., Brunette, S., Dilworth, F. J., & Megeney, L. A. (2010). Caspase 3/caspase-activated DNase promote cell differentiation by inducing DNA strand breaks. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(9), 4230–4235. <https://doi.org/10.1073/pnas.0913089107>
- Lawrence C. W. (2004). Cellular functions of DNA polymerase zeta and Rev1 protein. *Advances in protein chemistry*, *69*, 167–203. [https://doi.org/10.1016/S0065-3233\(04\)69006-1](https://doi.org/10.1016/S0065-3233(04)69006-1)
- Lee, H. O., Hong, Y., Etioglu, H. E., Cho, Y. B., Pomella, V., Van den Bosch, B., Vanhecke, J., Verbandt, S., Hong, H., Min, J. W., Kim, N., Eum, H. H., Qian, J., Boeckx, B., Lambrechts, D., Tsantoulis, P., De Hertogh, G., Chung, W., Lee, T., An, M., ... Park, W. Y. (2020). Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nature genetics*, *52*(6), 594–603. <https://doi.org/10.1038/s41588-020-0636-z>
- Lens, S. M., Voest, E. E., & Medema, R. H. (2010). Shared and separate functions of polo-like kinases and aurora kinases in cancer. *Nature reviews. Cancer*, *10*(12), 825–841. <https://doi.org/10.1038/nrc2964>
- Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, *69*(6), 915–926. [https://doi.org/10.1016/0092-8674\(92\)90611-f](https://doi.org/10.1016/0092-8674(92)90611-f)
- Lin, Y. F., Shih, H. Y., Shang, Z. F., Kuo, C. T., Guo, J., Du, C., Lee, H., & Chen, B. P. C. (2018). PIDD mediates the association of DNA-PKcs and ATR at stalled replication forks to facilitate the ATR signaling pathway. *Nucleic acids research*, *46*(4), 1847–1859. <https://doi.org/10.1093/nar/gkx1298>

- Lin, Y. F., Shih, H. Y., Shang, Z., Matsunaga, S., & Chen, B. P. (2014). DNA-PKcs is required to maintain stability of Chk1 and Claspin for optimal replication stress response. **Nucleic acids research**, **42*(7)*, 4463–4473. <https://doi.org/10.1093/nar/gku116>
- Lin, Y., Ma, W., & Benchimol, S. (2000). Pidd, a new death-domain-containing protein, is induced by p53 and promotes apoptosis. **Nature genetics**, **26*(1)*, 122–127. <https://doi.org/10.1038/79102>
- Lindahl T. (1993). Instability and decay of the primary structure of DNA. **Nature**, **362*(6422)*, 709–715. <https://doi.org/10.1038/362709a0>
- Lindahl, T., & Nyberg, B. (1974). Heat-induced deamination of cytosine residues in deoxyribonucleic acid. **Biochemistry**, **13*(16)*, 3405–3410. <https://doi.org/10.1021/bi00713a035>
- Loeffler, D., Schneiter, F., Wang, W., Wehling, A., Kull, T., Lengerke, C., Manz, M. G., & Schroeder, T. (2022). Asymmetric organelle inheritance predicts human blood stem cell fate. **Blood**, **139*(13)*, 2011–2023. <https://doi.org/10.1182/blood.2020009778>
- Loeffler, D., Wehling, A., Schneiter, F., Zhang, Y., Müller-Böttcher, N., Hoppe, P. S., Hilsenbeck, O., Kokkaliaris, K. D., Ende, M., & Schroeder, T. (2019). Asymmetric lysosome inheritance predicts activation of haematopoietic stem cells. **Nature**, **573*(7774)*, 426–429. <https://doi.org/10.1038/s41586-019-1531-6>
- Logette, E., Schuepbach-Mallepell, S., Eckert, M. J., Leo, X. H., Jaccard, B., Manzl, C., Tardivel, A., Villunger, A., Quadroni, M., Gaide, O., & Tschopp, J. (2011). PIDD orchestrates translesion DNA synthesis in response to UV irradiation. **Cell death and differentiation**, **18*(6)*, 1036–1045. <https://doi.org/10.1038/cdd.2011.19>
- London, T. B., Barber, L. J., Mosedale, G., Kelly, G. P., Balasubramanian, S., Hickson, I. D., Boulton, S. J., & Hiom, K. (2008). FANCD1 is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. **The Journal of biological chemistry**, **283*(52)*, 36132–36139. <https://doi.org/10.1074/jbc.M808152200>
- Lukas, C., Savic, V., Bekker-Jensen, S., Doil, C., Neumann, B., Pedersen, R. S., Grøfte, M., Chan, K. L., Hickson, I. D., Bartek, J., & Lukas, J. (2011). 53BP1 nuclear bodies form around DNA lesions generated by mitotic transmission of chromosomes under replication stress. **Nature cell biology**, **13*(3)*, 243–253. <https://doi.org/10.1038/ncb2201>
- Madabhushi, R., Gao, F., Pfenning, A. R., Pan, L., Yamakawa, S., Seo, J., Rueda, R., Phan, T. X., Yamakawa, H., Pao, P. C., Stott, R. T., Gjoneska, E., Nott, A., Cho, S., Kellis, M., & Tsai, L. H. (2015). Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. **Cell**, **161*(7)*, 1592–1605. <https://doi.org/10.1016/j.cell.2015.05.032>
- Magusto, J., Beaupère, C., Afonso, M. B., Auclair, M., Delaunay, J. L., Soret, P. A., Courtois, G., Aït-Slimane, T., Housset, C., Jéru, I., Fève, B., Ratzu, V., Rodrigues, C. M. P., & Gautheron, J. (2022). The necroptosis-inducing pseudokinase mixed lineage kinase domain-like regulates the adipogenic differentiation of pre-adipocytes. **iScience**, **25*(10)*, 105166. <https://doi.org/10.1016/j.isci.2022.105166>
- Maio, N., & Rouault, T. A. (2015). Iron-sulfur cluster biogenesis in mammalian cells: New insights into the molecular mechanisms of cluster delivery. **Biochimica et biophysica acta**, **1853*(6)*, 1493–1512. <https://doi.org/10.1016/j.bbamcr.2014.09.009>
- Maiti, A., & Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. **The Journal of biological chemistry**, **286*(41)*, 35334–35338. <https://doi.org/10.1074/jbc.C111.284620>
- Malumbres, M., & Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. **Nature reviews. Cancer**, **9*(3)*, 153–166. <https://doi.org/10.1038/nrc2602>
- Manthei, K. A., & Keck, J. L. (2013). The BLM dissolvosome in DNA replication and repair. **Cellular and molecular life sciences : CMLS**, **70*(21)*, 4067–4084. <https://doi.org/10.1007/s00018-013-1325-1>
- Markkanen E. (2017). Not breathing is not an option: How to deal with oxidative DNA damage. **DNA repair**, **59**, 82–105. <https://doi.org/10.1016/j.dnarep.2017.09.007>
- Massagué, J., & Xi, Q. (2012). TGF- β control of stem cell differentiation genes. **FEBS letters**, **586*(14)*, 1953–1958. <https://doi.org/10.1016/j.febslet.2012.03.023>

- Masuda, Y., Mitsuyuki, S., Kanao, R., Hishiki, A., Hashimoto, H., & Masutani, C. (2018). Regulation of HLTf-mediated PCNA polyubiquitination by RFC and PCNA monoubiquitination levels determines choice of damage tolerance pathway. *Nucleic acids research*, *46*(21), 11340–11356. <https://doi.org/10.1093/nar/gky943>
- Masutani, C., Kusumoto, R., Iwai, S., & Hanaoka, F. (2000). Mechanisms of accurate translesion synthesis by human DNA polymerase η . *The EMBO journal*, *19*(12), 3100–3109. <https://doi.org/10.1093/emboj/19.12.3100>
- McFaline-Figueroa, J. R., Vevea, J., Swayne, T. C., Zhou, C., Liu, C., Leung, G., Boldogh, I. R., & Pon, L. A. (2011). Mitochondrial quality control during inheritance is associated with lifespan and mother-daughter age asymmetry in budding yeast. *Aging cell*, *10*(5), 885–895. <https://doi.org/10.1111/j.1474-9726.2011.00731.x>
- McGill, M. A., & McGlade, C. J. (2003). Mammalian numb proteins promote Notch1 receptor ubiquitination and degradation of the Notch1 intracellular domain. *The Journal of biological chemistry*, *278*(25), 23196–23203. <https://doi.org/10.1074/jbc.M302827200>
- Merino, M. M., Levayer, R., & Moreno, E. (2016). Survival of the Fittest: Essential Roles of Cell Competition in Development, Aging, and Cancer. *Trends in cell biology*, *26*(10), 776–788. <https://doi.org/10.1016/j.tcb.2016.05.009>
- Monroe, J. G., Srikant, T., Carbonell-Bejerano, P., Becker, C., Lensink, M., Exposito-Alonso, M., Klein, M., Hildebrandt, J., Neumann, M., Kliebenstein, D., Weng, M. L., Imbert, E., Ågren, J., Rutter, M. T., Fenster, C. B., & Weigel, D. (2022). Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature*, *602*(7895), 101–105. <https://doi.org/10.1038/s41586-021-04269-6>
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., & Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell*, *102*(5), 553–563. [https://doi.org/10.1016/S0092-8674\(00\)00078-7](https://doi.org/10.1016/S0092-8674(00)00078-7)
- Nelson, J. R., Lawrence, C. W., & Hinkle, D. C. (1996). Deoxycytidyl transferase activity of yeast REV1 protein. *Nature*, *382*(6593), 729–731. <https://doi.org/10.1038/382729a0>
- Nieto, M. A., Huang, R. Y., Jackson, R. A., & Thiery, J. P. (2016). EMT: 2016. *Cell*, *166*(1), 21–45. <https://doi.org/10.1016/j.cell.2016.06.028>
- Ochs, F., Somyajit, K., Altmeyer, M., Rask, M. B., Lukas, J., & Lukas, C. (2016). 53BP1 fosters fidelity of homology-directed DNA repair. *Nature structural & molecular biology*, *23*(8), 714–721. <https://doi.org/10.1038/nsmb.3251>
- Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3), 247–257. [https://doi.org/10.1016/S0092-8674\(00\)81656-6](https://doi.org/10.1016/S0092-8674(00)81656-6)
- Onodera, A., González-Avalos, E., Lio, C. J., Georges, R. O., Bellacosa, A., Nakayama, T., & Rao, A. (2021). Roles of TET and TDG in DNA demethylation in proliferating and non-proliferating immune cells. *Genome biology*, *22*(1), 186. <https://doi.org/10.1186/s13059-021-02384-1>
- Oppenheim R. W. (1991). Cell death during development of the nervous system. *Annual review of neuroscience*, *14*, 453–501. <https://doi.org/10.1146/annurev.ne.14.030191.002321>
- Orkin, S. H., & Zon, L. I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*, *132*(4), 631–644. <https://doi.org/10.1016/j.cell.2008.01.025>
- Pan, L., Zhu, B., Hao, W., Zeng, X., Vlahopoulos, S. A., Hazra, T. K., Hegde, M. L., Radak, Z., Bacsi, A., Brasier, A. R., Ba, X., & Boldogh, I. (2016). Oxidized Guanine Base Lesions Function in 8-Oxoguanine DNA Glycosylase-1-mediated Epigenetic Regulation of Nuclear Factor κ B-driven Gene Expression. *The Journal of biological chemistry*, *291*(49), 25553–25566. <https://doi.org/10.1074/jbc.M116.751453>
- Pastushenko, I., & Blanpain, C. (2019). EMT Transition States during Tumor Progression and Metastasis. *Trends in cell biology*, *29*(3), 212–226. <https://doi.org/10.1016/j.tcb.2018.12.001>
- Pelka, K., Hofree, M., Chen, J. H., Sarkizova, S., Pirl, J. D., Jorgji, V., Bejnood, A., Dionne, D., Ge, W. H., Xu, K. H., Chao, S. X., Zollinger, D. R., Lieb, D. J., Reeves, J. W., Fuhrman, C. A., Hoang, M. L., Delorey, T., Nguyen, L. T., Waldman, J., Klapholz, M., ... Hacohen, N. (2021). Spatially organized multicellular

- immune hubs in human colorectal cancer. **Cell**, **184*(18)*, 4734–4752.e20. <https://doi.org/10.1016/j.cell.2021.08.003>
- Perillo, B., Ombra, M. N., Bertoni, A., Cuzzo, C., Sacchetti, S., Sasso, A., Chiariotti, L., Malorni, A., Abbondanza, C., & Avvedimento, E. V. (2008). DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. **Science (New York, N.Y.)**, **319*(5860)*, 202–206. <https://doi.org/10.1126/science.1147674>
- Petrakis, T. G., Komseli, E. S., Papaioannou, M., Vougas, K., Polyzos, A., Myrianthopoulos, V., Mikros, E., Trougakis, I. P., Thanos, D., Branzi, D., Townsend, P., & Gorgoulis, V. G. (2016). Exploring and exploiting the systemic effects of deregulated replication licensing. **Seminars in cancer biology**, **37-38**, 3–15. <https://doi.org/10.1016/j.semcancer.2015.12.002>
- Petryk, N., Dalby, M., Wenger, A., Stromme, C. B., Strandsby, A., Andersson, R., & Groth, A. (2018). MCM2 promotes symmetric inheritance of modified histones during DNA replication. **Science (New York, N.Y.)**, **361*(6409)*, 1389–1392. <https://doi.org/10.1126/science.aau0294>
- Phillips, J. E., & Corces, V. G. (2009). CTCF: master weaver of the genome. **Cell**, **137*(7)*, 1194–1211. <https://doi.org/10.1016/j.cell.2009.06.001>
- Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C. L., Mroz, E. A., Emerick, K. S., Deschler, D. G., Varvares, M. A., Mylvaganam, R., Rozenblatt-Rosen, O., Rocco, J. W., Faquin, W. C., Lin, D. T., Regev, A., & Bernstein, B. E. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. **Cell**, **171*(7)*, 1611–1624.e24. <https://doi.org/10.1016/j.cell.2017.10.044>
- Quyn, A. J., Appleton, P. L., Carey, F. A., Steele, R. J., Barker, N., Clevers, H., Ridgway, R. A., Sansom, O. J., & Näthke, I. S. (2010). Spindle orientation bias in gut epithelial stem cell compartments is lost in precancerous tissue. **Cell stem cell**, **6*(2)*, 175–181. <https://doi.org/10.1016/j.stem.2009.12.007>
- Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. **Cell**, **159*(7)*, 1665–1680. <https://doi.org/10.1016/j.cell.2014.11.021>
- Rizzotto, D., Englmaier, L., & Villunger, A. (2021). At a Crossroads to Cancer: How p53-Induced Cell Fate Decisions Secure Genome Integrity. **International journal of molecular sciences**, **22*(19)*, 10883. <https://doi.org/10.3390/ijms221910883>
- Roberts, S. A., Lawrence, M. S., Klimczak, L. J., Grimm, S. A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G. V., Carter, S. L., Saksena, G., Harris, S., Shah, R. R., Resnick, M. A., Getz, G., & Gordenin, D. A. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. **Nature genetics**, **45*(9)*, 970–976. <https://doi.org/10.1038/ng.2702>
- Rosa, F., Dray, N., Bedu, S., & Bally-Cuif, L. (2024). Non-apoptotic caspase events and Atf3 expression underlie direct neuronal differentiation of adult neural stem cells. **Development (Cambridge, England)**, **151*(22)*, dev204381. <https://doi.org/10.1242/dev.204381>
- Rosselin, M., Santo-Domingo, J., Bermont, F., Giacomello, M., & Demarex, N. (2017). L-OPA1 regulates mitoflash biogenesis independently from membrane fusion. **EMBO reports**, **18*(3)*, 451–463. <https://doi.org/10.15252/embr.201642931>
- Sale J. E. (2013). Translesion DNA synthesis and mutagenesis in eukaryotes. **Cold Spring Harbor perspectives in biology**, **5*(3)*, a012708. <https://doi.org/10.1101/cshperspect.a012708>
- Sale, J. E., Lehmann, A. R., & Woodgate, R. (2012). Y-family DNA polymerases and their role in tolerance of cellular DNA damage. **Nature reviews. Molecular cell biology**, **13*(3)*, 141–152. <https://doi.org/10.1038/nrm3289>
- Sarkies, P., Murat, P., Phillips, L. G., Patel, K. J., Balasubramanian, S., & Sale, J. E. (2012). FANCD1 coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. **Nucleic acids research**, **40*(4)*, 1485–1498. <https://doi.org/10.1093/nar/gkr868>
- Sarkies, P., Reams, C., Simpson, L. J., & Sale, J. E. (2010). Epigenetic instability due to defective replication of structured DNA. **Molecular cell**, **40*(5)*, 703–713. <https://doi.org/10.1016/j.molcel.2010.11.009>

- Sato, T., van Es, J. H., Snippert, H. J., Stange, D. E., Vries, R. G., van den Born, M., Barker, N., Shroyer, N. F., van de Wetering, M., & Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature*, *469*(7330), 415–418. <https://doi.org/10.1038/nature09637>
- 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>
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and analgesia*, *126*(5), 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>
- Schuster-Böckler, B., & Lehner, B. (2012). Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature*, *488*(7412), 504–507. <https://doi.org/10.1038/nature11273>
- Scorrano, L., & Korsmeyer, S. J. (2003). Mechanisms of cytochrome c release by proapoptotic BCL-2 family members. *Biochemical and biophysical research communications*, *304*(3), 437–444. [https://doi.org/10.1016/s0006-291x\(03\)00615-6](https://doi.org/10.1016/s0006-291x(03)00615-6)
- Sethi, A., Fernández-Casañas, M., Delpino, B., & Coster, G. (2025). How DNA secondary structures drive replication fork instability. *DNA repair*, *156*, 103913. <https://doi.org/10.1016/j.dnarep.2025.103913>
- Shah, P., Hill, R., Dion, C., Clark, S. J., Abakir, A., Willems, J., Arends, M. J., Garaycochea, J. I., Leitch, H. G., Reik, W., & Crossan, G. P. (2024). Primordial germ cell DNA demethylation and development require DNA translesion synthesis. *Nature communications*, *15*(1), 3734. <https://doi.org/10.1038/s41467-024-47219-2>
- Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., & Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, *119*(7), 941–953. <https://doi.org/10.1016/j.cell.2004.12.012>
- Simons, B. D., & Clevers, H. (2011). Strategies for homeostatic stem cell self-renewal in adult tissues. *Cell*, *145*(6), 851–862. <https://doi.org/10.1016/j.cell.2011.05.033>
- Smirnova, E., Griparic, L., Shurland, D. L., & van der Bliek, A. M. (2001). Dynamin-related protein Drp1 is required for mitochondrial division in mammalian cells. *Molecular biology of the cell*, *12*(8), 2245–2256. <https://doi.org/10.1091/mbc.12.8.2245>
- Smith, C. A., Lau, K. M., Rahmani, Z., Dho, S. E., Brothers, G., She, Y. M., Berry, D. M., Bonneil, E., Thibault, P., Schweisguth, F., Le Borgne, R., & McGlade, C. J. (2007). aPKC-mediated phosphorylation regulates asymmetric membrane localization of the cell fate determinant Numb. *The EMBO journal*, *26*(2), 468–480. <https://doi.org/10.1038/sj.emboj.7601495>
- Snippert, H. J., van der Flier, L. G., Sato, T., van Es, J. H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A. M., van Rheenen, J., Simons, B. D., & Clevers, H. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell*, *143*(1), 134–144. <https://doi.org/10.1016/j.cell.2010.09.016>
- Song, J., Teplova, M., Ishibe-Murakami, S., & Patel, D. J. (2012). Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science (New York, N.Y.)*, *335*(6069), 709–712. <https://doi.org/10.1126/science.1214453>
- Spencer, S. L., Cappell, S. D., Tsai, F. C., Overton, K. W., Wang, C. L., & Meyer, T. (2013). The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. *Cell*, *155*(2), 369–383. <https://doi.org/10.1016/j.cell.2013.08.062>
- Spies, J., Lukas, C., Somyajit, K., Rask, M. B., Lukas, J., & Neelsen, K. J. (2019). 53BP1 nuclear bodies enforce replication timing at under-replicated DNA to limit heritable DNA damage. *Nature cell biology*, *21*(4), 487–497. <https://doi.org/10.1038/s41556-019-0293-6>
- Stein, E. M., DiNardo, C. D., Pollyea, D. A., Fathi, A. T., Roboz, G. J., Altman, J. K., Stone, R. M., DeAngelo, D. J., Levine, R. L., Flinn, I. W., Kantarjian, H. M., Collins, R., Patel, M. R., Frankel, A. E., Stein, A., Sekeres, M. A., Swords, R. T., Medeiros, B. C., Willekens, C., Vyas, P., ... Tallman, M. S. (2017). Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. *Blood*, *130*(6), 722–731. <https://doi.org/10.1182/blood-2017-04-779405>

- Stein, M., Hile, S. E., Weissensteiner, M. H., Lee, M., Zhang, S., Kejnovský, E., Kejnovská, I., Makova, K. D., & Eckert, K. A. (2022). Variation in G-quadruplex sequence and topology differentially impacts human DNA polymerase fidelity. **DNA repair**, **119**, 103402. <https://doi.org/10.1016/j.dnarep.2022.103402>
- Stockwell, B. R., Friedmann Angeli, J. P., Bayir, H., Bush, A. I., Conrad, M., Dixon, S. J., Fulda, S., Gascón, S., Hatzios, S. K., Kagan, V. E., Noel, K., Jiang, X., Linkermann, A., Murphy, M. E., Overholtzer, M., Oyagi, A., Pagnussat, G. C., Park, J., Ran, Q., Rosenfeld, C. S., ... Zhang, D. D. (2017). Ferroptosis: A Regulated Cell Death Nexus Linking Metabolism, Redox Biology, and Disease. **Cell**, **171*(2)*, 273–285. <https://doi.org/10.1016/j.cell.2017.09.021>
- Strasser A. (2005). The role of BH3-only proteins in the immune system. **Nature reviews. Immunology**, **5*(3)*, 189–200. <https://doi.org/10.1038/nri1568>
- Supek, F., & Lehner, B. (2015). Differential DNA mismatch repair underlies mutation rate variation across the human genome. **Nature**, **521*(7550)*, 81–84. <https://doi.org/10.1038/nature14173>
- Susin, S. A., Lorenzo, H. K., Zamzami, N., Marzo, I., Snow, B. E., Brothers, G. M., Mangion, J., Jacotot, E., Costantini, P., Loeffler, M., Larochette, N., Goodlett, D. R., Aebersold, R., Siderovski, D. P., Penninger, J. M., & Kroemer, G. (1999). Molecular characterization of mitochondrial apoptosis-inducing factor. **Nature**, **397*(6718)*, 441–446. <https://doi.org/10.1038/17135>
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. **Science (New York, N.Y.)**, **324*(5929)*, 930–935. <https://doi.org/10.1126/science.1170116>
- Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. **Cell**, **126*(4)*, 663–676. <https://doi.org/10.1016/j.cell.2006.07.024>
- Tell, G., Quadrioglio, F., Tiribelli, C., & Kelley, M. R. (2009). The many functions of APE1/Ref-1: not only a DNA repair enzyme. **Antioxidants & redox signaling**, **11*(3)*, 601–620. <https://doi.org/10.1089/ars.2008.2194>
- 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>
- Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., 2nd, Treacy, D., Trombetta, J. J., Rothenberg, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J. R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G., Kazer, S. W., ... Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. **Science (New York, N.Y.)**, **352*(6282)*, 189–196. <https://doi.org/10.1126/science.aad0501>
- Tran, V., Lim, C., Xie, J., & Chen, X. (2012). Asymmetric division of Drosophila male germline stem cell shows asymmetric histone distribution. **Science (New York, N.Y.)**, **338*(6107)*, 679–682. <https://doi.org/10.1126/science.1226028>
- Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., & Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. **Nature**, **439*(7078)*, 811–816. <https://doi.org/10.1038/nature04433>
- Tuteja, N., & Tuteja, R. (2004). Unraveling DNA helicases. Motif, structure, mechanism and function. *European Journal of Biochemistry*, *271(10)*, 1849–1863. <https://doi.org/10.1111/j.1432-1033.2004.04094.x> PMID: 15128295
- van Galen, P., Hovestadt, V., Wadsworth II, M. H., Hughes, T. K., Griffin, G. K., Battaglia, S., Verga, J. A., Stephansky, J., Pastika, T. J., Lombardi Story, J., Pinkus, G. S., Pozdnyakova, O., Galinsky, I., Stone, R. M., Graubert, T. A., Shalek, A. K., Aster, J. C., Lane, A. A., & Bernstein, B. E. (2019). Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. **Cell**, **176*(6)*, 1265–1281.e24. <https://doi.org/10.1016/j.cell.2019.01.031>
- Vesela, B., Killinger, M., Rihova, K., Benes, P., Svandová, E., Kratochvilová, A., Trcka, F., Kleparnik, K., & Matalova, E. (2022). Caspase-8 Deficient Osteoblastic Cells Display Alterations in Non-Apoptotic

- Pathways. **Frontiers in cell and developmental biology**, **10**, 794407. <https://doi.org/10.3389/fcell.2022.794407>
- Veselá, B., Ševčíková, A., Holomková, K., Ramešová, A., Kratochvílová, A., Sharpe, P. T., & Matalová, E. (2024). Inhibition of caspase-8 cascade restrains the osteoclastogenic fate of bone marrow cells. **Pflugers Archiv : European journal of physiology**, **476*(8)*, 1289–1302. <https://doi.org/10.1007/s00424-024-02977-2>
- Vitale, I., Manic, G., De Maria, R., Kroemer, G., & Galluzzi, L. (2017). DNA Damage in Stem Cells. **Molecular cell**, **66*(3)*, 306–319. <https://doi.org/10.1016/j.molcel.2017.04.006>
- 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, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., Thurman, R. E., Kaul, R., Myers, R. M., & Stamatoyannopoulos, J. A. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. **Genome research**, **22*(9)*, 1680–1688. <https://doi.org/10.1101/gr.136101.111>
- Wang, X., Fu, H., Sun, Q., Huang, B., Xu, Z., Zhai, X., Deng, C., Peng, L., Zhang, M., Peng, T., Gong, A., Liu, J., Zou, Z., Pan, G., Chen, J., Wu, G., Zhang, M., & Min, M. (2026). Asymmetric division in a two-cell-like state rejuvenates embryonic stem cells. **Cell research**, **36*(3)*, 219–232. <https://doi.org/10.1038/s41422-026-01221-z>
- Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D., & Klein, A. M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. **Science (New York, N.Y.)**, **367*(6479)*, eaaw3381. <https://doi.org/10.1126/science.aaw3381>
- Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K., & Peters, J. M. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. **Nature**, **451*(7180)*, 796–801. <https://doi.org/10.1038/nature06634>
- Whitaker, A. M., & Freudenthal, B. D. (2018). APE1: A skilled nucleic acid surgeon. **DNA repair**, **71**, 93–100. <https://doi.org/10.1016/j.dnarep.2018.08.012>
- White M. F. (2009). Structure, function and evolution of the XPD family of iron-sulfur-containing 5'→3' DNA helicases. **Biochemical Society transactions**, **37*(Pt 3)*, 547–551. <https://doi.org/10.1042/BST0370547>
- Wojtaszek, J. L., Chatterjee, N., Najeeb, J., Ramos, A., Lee, M., Bian, K., Xue, J. Y., Fenton, B. A., Park, H., Li, D., Hemann, M. T., Hong, J., Walker, G. C., & Zhou, P. (2019). A Small Molecule Targeting Mutagenic Translesion Synthesis Improves Chemotherapy. **Cell**, **178*(1)*, 152–159.e11. <https://doi.org/10.1016/j.cell.2019.05.028>
- Wood, R. D., & Doublé, S. (2016). DNA polymerase θ (POLQ), double-strand break repair, and cancer. **DNA repair**, **44**, 22–32. <https://doi.org/10.1016/j.dnarep.2016.05.003>
- Wooten, M., Snedeker, J., Nizami, Z. F., Yang, X., Ranjan, R., Urban, E., Kim, J. M., Gall, J., Xiao, J., & Chen, X. (2019). Asymmetric histone inheritance via strand-specific incorporation and biased replication fork movement. **Nature structural & molecular biology**, **26*(8)*, 732–743. <https://doi.org/10.1038/s41594-019-0269-z>
- Wright, E. P., Abdelhamid, M. A. S., Ehiabor, M. O., Grigg, M. C., Irving, K., Smith, N. M., & Waller, Z. A. E. (2020). Epigenetic modification of cytosines fine tunes the stability of i-motif DNA. **Nucleic acids research**, **48*(1)*, 55–62. <https://doi.org/10.1093/nar/gkz1082>
- Wu, M. J., Chen, Y. S., Kim, M. R., Chang, C. C., Gampala, S., Zhang, Y., Wang, Y., Chang, C. Y., Yang, J. Y., & Chang, C. J. (2019). Epithelial-Mesenchymal Transition Directs Stem Cell Polarity via Regulation of Mitofusin. **Cell metabolism**, **29*(4)*, 993–1002.e6. <https://doi.org/10.1016/j.cmet.2018.11.004>
- Xie, J., Wooten, M., Tran, V., Chen, B. C., Pozmanter, C., Simbolon, C., Betzig, E., & Chen, X. (2015). Histone H3 Threonine Phosphorylation Regulates Asymmetric Histone Inheritance in the Drosophila Male Germline. **Cell**, **163*(4)*, 920–933. <https://doi.org/10.1016/j.cell.2015.10.002>
- 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>
- Yamaguchi, R., Lartigue, L., Perkins, G., Scott, R. T., Dixit, A., Kushnareva, Y., Kuwana, T., Ellisman, M. H., & Newmeyer, D. D. (2008). Opa1-mediated cristae opening is Bax/Bak and BH3 dependent, required for apoptosis, and independent of Bak oligomerization. **Molecular cell**, **31*(4)*, 557–569. <https://doi.org/10.1016/j.molcel.2008.07.010>
- Yamashita, Y. M., Jones, D. L., & Fuller, M. T. (2003). Orientation of asymmetric stem cell division by the APC tumor suppressor and centrosome. **Science (New York, N.Y.)**, **301*(5639)*, 1547–1550. <https://doi.org/10.1126/science.1087795>
- 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>
- Yang, W. S., SriRamaratnam, R., Welsch, M. E., Shimada, K., Skouta, R., Viswanathan, V. S., Cheah, J. H., Clemons, P. A., Shamji, A. F., Clish, C. B., Brown, L. M., Girotti, A. W., Cornish, V. W., Schreiber, S. L., & Stockwell, B. R. (2014). Regulation of ferroptotic cancer cell death by GPX4. **Cell**, **156*(1-2)*, 317–331. <https://doi.org/10.1016/j.cell.2013.12.010>
- Yu, J., Zhang, Y., Fang, Y., Paulo, J. A., Yaghoubi, D., Hua, X., Shipkovenska, G., Toda, T., Zhang, Z., Gygi, S. P., Jia, S., Li, Q., & Moazed, D. (2024). A replisome-associated histone H3-H4 chaperone required for epigenetic inheritance. **Cell**, **187*(18)*, 5010–5028.e24. <https://doi.org/10.1016/j.cell.2024.07.006>
- 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>

Supplementary Table S1: Gene Panel — Established Functions, Model Interpretations, and Falsification Tests

Panel Composition

69 genes across 4 functional tiers. Each gene's tier assignment describes a functional relationship within the model, not an intrinsic property of the protein.

Column Definitions

- **Gene:** HGNC-approved symbol
- **Full Name:** Official protein name
- **Entrez ID:** NCBI Gene ID for reproducibility
- **Established Function:** Published biochemical role (citations in main text)
- **Model Tier:** Tier 1 (environmental sensing), Tier 2 (chromatin state), Tier 3 (regulatory coordination), Tier 4 (execution machinery)
- **Model Role:** Specific function within the ARC architecture
- **Falsification Test:** What observation would refute the model's assignment of this gene

Gene	Full Name	Entrez	Established Function	Tier	Model Role	Falsification Test
NUMB	NUMB endocytic adaptor protein	8650	Notch inhibitor; asymmetric cell fate determinant; p53 stabilizer via MDM2	3	Partition system organizer; couples polarity to commitment	NUMB coupling preserved (not dissolved) in cancer
NUMBL	NUMB-like	9253	NUMB paralog; genome-integrity-associated functions	3	Genome-integrity comparator to NUMB	NUMBL tracks commitment rather than genome integrity
PRKCI	Protein kinase C iota	5584	Atypical PKC; apical polarity complex member	3	Polarity partner; NUMB coupling target	PRKCI×NUMB coupling preserved in cancer
PRKCZ	Protein kinase C zeta	5590	Atypical PKC; Par complex member	3	Polarity partner; NUMB coupling target	PRKCZ×NUMB coupling preserved in cancer
DLG1	Discs large homolog 1	1739	Scaffolding protein; basolateral polarity	3	Polarity partner; NUMB coupling target	DLG1×NUMB coupling preserved in cancer
PARD3	Par-3 family cell polarity regulator	56288	Apical polarity complex; tight junction formation	3	Polarity partner; NUMB coupling target	PARD3×NUMB coupling preserved in cancer
SCRIB	Scribble	23513	Basolateral polarity determinant	3	Polarity partner; NUMB coupling target	SCRIB×NUMB coupling preserved in cancer
EZH2	Enhancer of zeste	2146	H3K27 methyltransferase;	2	Chromatin writer;	EZH2×KDM6A

	homolog 2		PRC2 catalytic subunit		commitment-direction marker	coupling preserved in cancer
KDM6A	Lysine demethylase 6A	7403	H3K27 demethylase	2	Chromatin eraser; EZH2 counter-regulator	KDM6A×EZH2 coupling preserved in cancer
KDM6B	Lysine demethylase 6B	23135	H3K27 demethylase; inflammation-responsive	2	Chromatin eraser; EZH2 counter-regulator	KDM6B×EZH2 coupling preserved in cancer
KDM1A	Lysine-specific demethylase 1A (LSD1)	23028	H3K4/H3K9 demethylase	2	Chromatin eraser	KDM1A×KDM5A coupling preserved in cancer
KDM5A	Lysine demethylase 5A	5927	H3K4 demethylase	2	Chromatin eraser	KDM5A×KDM1A coupling preserved in cancer
TET2	Ten-eleven translocation 2	54790	Oxidizes 5mC → 5hmC/5fC/5caC; αKG-dependent	2	Write/erase demethylation; metabolic gate substrate	TET2×TDG coupling preserved in cancer
TET3	Ten-eleven translocation 3	200424	Oxidizes 5mC → 5hmC; neuronal enrichment	2	TET2 functional paralog	TET3×TET2 coupling preserved in cancer
TDG	Thymine DNA glycosylase	6996	Excises T from G:T mismatches; excises 5fC/5caC from DNA	2	Template-strand eraser; completes active demethylation	TDG×TET2 coupling preserved in cancer
OGG1	8-oxoguanine DNA glycosylase	4968	Excises 8-oxoguanine; G4-gated BER initiation	2	New-strand writer at G4 addresses; commitment chemistry	OGG1×MUTYH coupling preserved in cancer
MUTYH	MutY DNA glycosylase	4595	Excises A opposite 8-oxoG from DNA	2	BER coordination partner for OGG1	MUTYH×OGG1 coupling preserved in cancer
NEIL1	Nei-like DNA glycosylase 1	79661	Excises oxidized pyrimidines; replication fork associated	2	BER fork-coupled glycosylase	NEIL1×XRCC1 coupling preserved in cancer
XRCC1	X-ray repair cross-complementing protein 1	7515	BER scaffold; recruits Pol β and Lig III	2	BER scaffold partner for NEIL1	XRCC1×NEIL1 coupling preserved in cancer
APEX1	Apurinic/apyrimidinic endonuclease 1	328	BER endonuclease; redox factor for TF DNA binding	2	Death-coupled nuclease; purging quality control	APEX1×BAX coupling preserved in cancer
IDH1	Isocitrate dehydrogenase 1	3417	Cytoplasmic IDH; produces αKG; R132H produces 2-HG	1	Metabolic gate; αKG supply for TET/KDM dioxygenases	IDH1 couples to CDK4 (not CDK6) in non-GBM cancers
IDH2	Isocitrate dehydrogenase 2	3418	Mitochondrial IDH; produces αKG	1	Tissue-specific metabolic gate (myeloid isoform)	IDH2 couples to CDK6 (not CDK4) in AML
CDK6	Cyclin-dependent kinase 6	1021	G1 kinase; RB1 phosphorylation; kinase-independent TF	1	Metabolic sensor; couples to αKG/dioxygenase program	CDK6 regulatory couplings preserved in cancer
CDK4	Cyclin-dependent kinase 4	1019	G1 kinase; RB1 phosphorylation	4	Execution kinase; structural CDK partner	CDK4×CCND1 coupling dissolved in cancer
RB1	Retinoblastoma protein 1	5925	Pocket protein; E2F repressor; cell cycle brake	3	CDK6 downstream target; licensing gate	RB1×CDK6 coupling preserved in cancer
E2F1	E2F transcription factor 1	1869	S-phase gene transcription factor	3	CDK6 downstream target; S-phase licensing	E2F1×CDK6 coupling preserved in cancer
CCND1	Cyclin D1	595	CDK4/CDK6 activating partner	3/4	Dual-classified: CDK6 regulatory (Class I) + CDK4 structural (Class II)	CCND1×CDK6 preserved OR CCND1×CDK4 dissolved
SIRT1	Sirtuin 1	23411	NAD ⁺ -dependent deacetylase; H3K9ac,	3	Metabolic-	SIRT1×SIRT6 coupling dissolved

			H4K16ac, p53, PGC-1 α		epigenetic coupling	in cancer (unexpected)
SIRT6	Sirtuin 6	51548	NAD ⁺ -dependent deacetylase; H3K9ac, H3K56ac; DNA repair	3	Sirtuin coordination partner	SIRT6 $\times$ SIRT7 coupling preserved in cancer
SIRT7	Sirtuin 7	51547	NAD ⁺ -dependent deacetylase; H3K18ac; rDNA regulation	3	Sirtuin coordination partner	SIRT7 $\times$ SIRT6 coupling preserved in cancer
CLSPN	Claspin	63967	Fork protection complex; CHK1 activation; histone routing	4	Execution scaffold at replication fork	CLSPN $\times$ PCNA coupling dissolved in cancer
PCNA	Proliferating cell nuclear antigen	5111	Replication clamp; K164 ubiquitination substrate	4	Execution clamp; modification code platform	PCNA $\times$ CLSPN coupling dissolved in cancer
PARP1	Poly(ADP-ribose) polymerase 1	142	DNA damage signaling; chromatin relaxation	4	Fork protection complex member	PARP1 $\times$ CLSPN coupling dissolved in cancer
CHEK1	Checkpoint kinase 1	1111	ATR substrate; S/G2 checkpoint effector	4	Fork protection complex member	CHEK1 $\times$ CLSPN coupling dissolved in cancer
ATR	ATR serine/threonine kinase	545	Replication stress sensor; CHEK1 activator	4	Fork protection complex member	ATR $\times$ CLSPN coupling dissolved in cancer
FANCI	Fanconi anemia group I	55215	FA pathway; ICL repair; FANCD2 partner	4	Execution complex (ID complex)	FANCI $\times$ FANCD2 coupling dissolved in cancer
FANCD2	Fanconi anemia group D2	2177	FA pathway; ICL repair; monoubiquitination substrate	4	Execution complex (ID complex)	FANCD2 $\times$ FANCI coupling dissolved in cancer
PLK1	Polo-like kinase 1	5347	Mitotic kinase; centrosome maturation	4	Execution kinase (mitotic)	PLK1 $\times$ AURKA coupling dissolved in cancer
AURKA	Aurora kinase A	6790	Mitotic kinase; spindle assembly	4	Execution kinase (mitotic)	AURKA $\times$ PLK1 coupling dissolved in cancer
DNMT1	DNA methyltransferase 1	1786	Maintenance methyltransferase; replication-coupled	4	Fork-coupled maintenance	DNMT1 $\times$ PCNA coupling dissolved in cancer
BRCA1	BRCA1 DNA repair associated	672	HR; DSB signaling; PALB2 $\rightarrow$ BRCA2 $\rightarrow$ RAD51 loading	3	HR regulatory coordinator; Class I (partner-independent)	BRCA1 $\times$ RAD51 coupling dissolved in cancer
BRCA2	BRCA2 DNA repair associated	675	RAD51 loader; HR effector	3	Commitment-direction marker	BRCA2 decouples from lineage markers in cancer
BLM	Bloom syndrome helicase	641	RecQ helicase; HJ dissolution; G4 resolution	3	Fidelity-providing helicase	BLM couples to TLS (not fidelity)
WRN	Werner syndrome helicase	7486	RecQ helicase; exonuclease; G4 resolution	3	Fidelity-providing helicase	WRN couples to TLS (not fidelity)
BRIP1	BRCA1-interacting helicase 1 (FANCI)	83990	Fe-S helicase; G4 resolution; ICL repair	3	Fidelity-providing helicase	BRIP1 couples to TLS (not fidelity)
MFN1	Mitofusin 1	55669	Outer mitochondrial membrane fusion GTPase	3	Mitochondrial partition; OMM fusion	MFN1 $\times$ OPA1 coupling dissolved in cancer
OPA1	Optic atrophy protein 1	4976	IMM fusion; cristae junction width control	3	Mitochondrial partition; cristae gating	OPA1 $\times$ MFN1 coupling dissolved in cancer
FIS1	Mitochondrial fission protein 1	51024	Opposes OPA1; promotes cristae junction opening	3	Mitochondrial partition; IMS access	FIS1 promotes fusion (not fission) in cancer context
BAX	BCL-2-associated X protein	581	Pro-apoptotic; OMM pore oligomerization	3	Death-coupled quality control; purging	BAX $\times$ APEX1 coupling preserved in cancer
BAK1	BCL-2 antagonist/killer 1	578	Pro-apoptotic; OMM pore oligomerization	3	Death-coupled quality control; purging	BAK1 $\times$ APEX1 coupling preserved in cancer

BCL2	B-cell CLL/lymphoma 2	596	Anti-apoptotic; inhibits BAX/BAK1 pore formation at OMM (Kale et al., 2018)	3	Death-pathway anti-apoptotic component whose ratio with pro-apoptotic partners reflects commitment-gate threshold	Overexpression should raise the commitment threshold; if the model is correct, BCL2-high cancers should show delayed purging without altered regulatory coordination
MCL1	Myeloid cell leukemia 1	4170	Anti-apoptotic; short half-life allows dynamic regulation (Czabotar et al., 2014)	3	Rapidly tunable death-pathway component; MCL1 half-life provides temporal resolution to the commitment gate	MCL1 inhibition should lower the commitment threshold specifically in MCL1-dependent lineages; timing of MCL1 loss relative to cell cycle should determine whether purging or survival results
POLK	DNA polymerase kappa	51426	Y-family TLS polymerase; bulky adduct bypass	3	Error-prone TLS; commitment chemistry	POLK couples to helicases (not TLS) in cancer
POLH	DNA polymerase eta	5429	Y-family TLS; UV photoproduct bypass; error-prone at 8-oxoG	3	Error-prone TLS	POLH couples to helicases (not TLS)
REV1	REV1 DNA-directed polymerase	51455	dCMP insertase; TLS scaffold; chromatin maintenance	3	G4 gate activator; dual function	REV1 loses scaffold function in cancer
REV3L	REV3L catalytic subunit of Pol ζ	5980	B-family TLS polymerase; mutagenic extension	3	Error-prone extension; mark installation	REV3L couples to fidelity (not TLS)
HLTF	Helicase-like transcription factor	6563	G4 helicase; PCNA K63-polyUb E3 ligase	3	Resolution architecture; fork reversal	HLTF/PRIMPOL ratio same in cancer vs normal
TOP2B	DNA topoisomerase II-beta	7155	DSBs at enhancers for loop formation	3	Enhancer-commitment coupling	TOP2B×NUMB coupling preserved in cancer
RAD51	RAD51 recombinase	5888	Strand invasion; HR effector	3	HR execution; BRCA1 loading target	RAD51×BRCA1 coupling dissolved in cancer
PRIMPOL	Primase-polymerase	—	Leading-strand repriming at G4 structures	3	Gap creation → K164 code → TLS	PRIMPOL gaps filled faithfully (not by TLS)
GPX4	Glutathione peroxidase 4	2879	Prevents lipid peroxidation; ferroptosis suppressor	3	Stem-associated redox defense	GPX4 anti-coupled to stemness markers
MLKL	Mixed lineage kinase domain-like	197259	Necroptosis executor; membrane permeabilization	3	Death execution; necroptotic purging	MLKL couples to stemness (not commitment)
MKI67	Ki-67	4288	Proliferation marker; PCL chromatin function	—	Retired as primary proliferation control	—

Notes

- Tier assignments describe functional relationships within the ARC model, not intrinsic protein properties (§2.2)
- CCND1 is dual-classified: Tier 3 with CDK6 (regulatory), Tier 4 with CDK4 (structural)

- Falsification tests are stated as the observation that would refute the model's specific claim about each gene
- Entrez IDs enable programmatic data retrieval for independent verification
- All established functions are cited in the main text at first mention of each gene

MYC	MYC proto-oncogene	4609	Transcription factor; growth and proliferation; G4-containing promoter	4	Mark-installation readout; G4 address exemplar	MYC decoupled from commitment program in cancer
PIDD1	p53-induced death domain protein 1	55367	PIDDosome component; centrosome surveillance	3	Replication-quality sentinel; couples centrosome count to death pathway	PIDD1 network coupling preserved in cancer
CASP2	Caspase-2	835	Initiator caspase; PIDDosome effector; ancestral bilaterian caspase	3	Death-differentiation switch; centrosome-coupled elimination	CASP2 coupling to PIDD1 preserved
CRADD	CASP2 and RIPK1 domain-containing adaptor with death domain	8738	PIDDosome adaptor; bridges PIDD1 to CASP2	3	PIDDosome scaffold	CRADD×PIDD1 coupling preserved
ANKRD26	Ankyrin repeat domain-containing protein 26	22852	PIDDosome regulator; centrosome licensing	3	Centrosome number sensor	ANKRD26 coupling to PIDD1 preserved
MDM2	MDM2 proto-oncogene	4193	E3 ubiquitin ligase; p53 degradation; NUMB substrate	3	Partition readout; NUMB stabilizes p53 by blocking MDM2	MDM2×NUMB coupling dissolved in cancer

Table S2: Complete Pair Inventory with Classification and Coupling Change

Data freeze: 2026-05-22 | Source: UCSC Xena TCGA HiSeqV2 log2(RSEM+1)

Method: Spearman rank; Fisher z-transform pooling (n-3 inverse-variance weights)

Cohorts: BRCA, KIRC, LUAD, LUSC, THCA, LIHC, COAD, HNSC, STAD (n = 5269)

Class-level statistics

- Class I: n=28, mean $|\Delta\rho| = 0.1815$
- Class II: n=8, mean $|\Delta\rho| = 0.0396$
- Mann-Whitney U = 212, p = 8.5260e-06 (one-sided, greater)
- Bootstrap 95% CI: [0.1035, 0.1796] (seed=42, 10000 resamples)

Class I: Regulatory Coordination (n=28)

#	Pair	ρ_{normal}	ρ_{tumor}	$\Delta\rho$	$\Delta\rho$	Direction	n_cohorts
1	NUMB × PARD3	+0.3378	-0.0211	-0.3590	0.3590	weakening	9
2	BAK1 × APEX1	+0.3363	+0.0361	-0.3001	0.3001	weakening	9
3	CDK6 × CCND1	+0.2269	-0.0568	-0.2837	0.2837	weakening	9
4	BAX × APEX1	+0.5371	+0.2573	-0.2797	0.2797	weakening	9
5	NUMB × DLG1	+0.4567	+0.1888	-0.2679	0.2679	weakening	9
6	NUMB × PRKCZ	+0.2630	+0.0134	-0.2496	0.2496	weakening	9
7	CDK6 × E2F1	+0.1095	-0.1344	-0.2439	0.2439	weakening	9
8	NEIL1 × XRCC1	+0.4469	+0.2065	-0.2404	0.2404	weakening	9
9	TET2 × TET3	+0.7108	+0.4807	-0.2301	0.2301	weakening	9
10	MFN1 × OPA1	+0.4233	+0.6464	+0.2231	0.2231	strengthening	9
11	OGG1 × MUTYH	+0.4849	+0.2765	-0.2084	0.2084	weakening	9
12	TET2 × TDG	+0.4340	+0.2329	-0.2012	0.2012	weakening	9
13	EZH2 × KDM6A	+0.2036	+0.0109	-0.1927	0.1927	weakening	9
14	KDM1A × KDM5A	+0.2281	+0.0380	-0.1901	0.1901	weakening	9
15	NUMB × PRKCI	+0.2961	+0.1075	-0.1886	0.1886	weakening	9
16	EZH2 × KDM6B	+0.1066	-0.0802	-0.1868	0.1868	weakening	9
17	SIRT6 × SIRT7	+0.6792	+0.5047	-0.1744	0.1744	weakening	9
18	BRCA1 × RAD51	+0.3536	+0.5273	+0.1737	0.1737	strengthening	9
19	NUMB × SCRIB	-0.0641	-0.2084	-0.1443	0.1443	weakening	9
20	NUMB × MFN1	+0.2789	+0.1472	-0.1316	0.1316	weakening	9
21	NUMB × TOP2B	+0.2657	+0.1567	-0.1089	0.1089	weakening	9
22	FANCD2 × BRCA1	+0.5042	+0.6112	+0.1070	0.1070	strengthening	9
23	CDK4 × CDK6	-0.2013	-0.1040	+0.0972	0.0972	strengthening	9
24	NUMB × OPA1	+0.2648	+0.1720	-0.0928	0.0928	weakening	9
25	CDK6 × RB1	+0.2959	+0.2189	-0.0771	0.0771	weakening	9
26	BAX × OGG1	+0.2803	+0.2318	-0.0486	0.0486	weakening	9
27	NUMB × POLK	+0.2644	+0.3071	+0.0427	0.0427	strengthening	9
28	NUMB × REV3L	+0.2400	+0.2798	+0.0398	0.0398	strengthening	9

Class II: Execution (n=8)

#	Pair	ρ_{normal}	ρ_{tumor}	$\Delta\rho$	$\Delta\rho$	Direction	n_cohorts
1	FANCI × FANCD2	+0.8355	+0.7290	-0.1066	0.1066	weakening	9

2	CLSPN × CHEK1	+0.4743	+0.5477	+0.0734	0.0734	strengthening	9
3	CLSPN × PARP1	+0.3932	+0.4383	+0.0451	0.0451	strengthening	9
4	PLK1 × AURKA	+0.6911	+0.7319	+0.0409	0.0409	strengthening	9
5	CLSPN × PCNA	+0.4561	+0.4293	-0.0268	0.0268	weakening	9
6	DNMT1 × PCNA	+0.3383	+0.3532	+0.0149	0.0149	strengthening	9
7	CDK4 × CCND1	-0.1523	-0.1454	+0.0069	0.0069	strengthening	9
8	CLSPN × ATR	+0.3682	+0.3659	-0.0023	0.0023	weakening	9

Per-cohort sample sizes

Cohort	Tumor	Normal	Total
BRCA	1097	114	1211
KIRC	533	72	605
LUAD	515	59	574
LUSC	502	51	553
THCA	505	59	564
LIHC	371	50	421
COAD	286	41	327
HNSC	520	44	564
STAD	415	35	450
***Total**	***4744**	**525**	**5269**

Supplementary Table S3: C1–C5 Molecular Classification Scoring

Scoring Criteria (from Methods §9.6)

Each pair is scored on five molecular criteria for structural execution partnership:

- **C1:** Demonstrated direct physical binding (1 = yes, 0 = no)
- **C2:** Shared subcellular compartment during function (1 = yes, 0 = no)
- **C3:** Absence of partner-independent functions (1 = both partners lack independent roles, 0 = one or both have extensive independent functions)
- **C4:** Constitutive association (1 = constitutive complex, 0 = transient/damage-induced/cell-cycle-restricted)
- **C5:** Co-expression reflects stoichiometric complex (1) or shared regulatory programming (0)

Classification rule: Score $\geq 4 \rightarrow$ Class II (execution). Score $\leq 2 \rightarrow$ Class I (regulatory). Score 3 $\rightarrow$ evaluated case-by-case with literature justification.

CLASS II PAIRS (Score ≥ 4)

Pair	C1	C2	C3	C4	C5	Total	Class	Key evidence
CLSPN×PCNA	1	1	0	1	1	4	II	Direct binding at replication fork; CLSPN has independent CHK1 role (C3=0)
CLSPN×ATR	1	1	0	1	1	4	II	Fork protection complex; both have independent functions
CLSPN×CHEK1	1	1	0	1	1	4	II	Claspin mediates ATR $\rightarrow$ CHK1; CHEK1 has CLSPN-independent substrates
CLSPN×PARP1	1	1	0	1	1	4	II	Fork protection co-localization; PARP1 has extensive independent roles
DNMT1×PCNA	1	1	0	1	1	4	II	PCNA-interacting peptide (PIP box); DNMT1 has PCNA-independent functions
FANCI×FANCD2	1	1	1	1	1	5	II	Obligate ID2 heterodimer; no independent

								functions documented
PLK1×AURKA	1	1	0	1	1	4	II	Centrosome co-localization; both have independent mitotic substrates
CDK4×CCND1	1	1	1	1	1	5	II	Obligate kinase-cyclin pair; CDK4 inactive without CCND1

CLASS I PAIRS — Polarity dissolution (Score ≤ 2)

Pair	C1	C2	C3	C4	C5	Total	Class	Key evidence
NUMB×PRKCZ	0	1	0	0	0	1	I	Indirect: NUMB inhibits Notch; PRKCZ phosphorylates Par3; both have extensive independent roles
NUMB×PRKCI	0	1	0	0	0	1	I	Same logic as PRKCZ; co-regulation reflects shared polarity program
NUMB×DLG1	0	1	0	0	0	1	I	Both basolateral; no direct complex; regulatory co-expression
NUMB×PARD3	0	1	0	0	0	1	I	PARD3 is apical; NUMB is basolateral; complementary polarity
NUMB×SCRIB	0	1	0	0	0	1	I	Both basolateral polarity; no direct binding demonstrated
NUMB×POLK	0	0	0	0	0	0	I	Different compartments; no physical interaction; regulatory coupling only
NUMB×TOP2B	0	0	0	0	0	0	I	Different compartments; no physical interaction
NUMB×MFN1	0	0	0	0	0	0	I	Cytoplasmic vs OMM; no physical interaction
NUMB×OPA1	0	0	0	0	0	0	I	Cytoplasmic vs IMM; no physical interaction
NUMB×REV3L	0	0	0	0	0	0	I	Different compartments;

								mark-installation coupling
--	--	--	--	--	--	--	--	----------------------------

CLASS I PAIRS — BER/chromatin coordination (Score ≤ 2)

Pair	C1	C2	C3	C4	C5	Total	Class	Key evidence
TET2×TET3	0	1	0	0	0	1	I	Paralogous enzymes; independent substrates; shared regulatory program
TET2×TDG	0	1	0	0	0	1	I	Sequential but not complexed; TDG acts on TET2 product; both have independent roles
OGG1×MUTYH	0	1	0	0	0	1	I	Sequential BER on opposite lesions; no direct complex
NEIL1×XRCC1	1	1	0	0	0	2	I	Direct binding but transient; both have extensive independent roles
EZH2×KDM6A	0	1	0	0	0	1	I	Opposing enzymes on H3K27; no complex; counter-regulatory
EZH2×KDM6B	0	1	0	0	0	1	I	Same as KDM6A; inflammation-responsive variant
KDM1A×KDM5A	0	1	0	0	0	1	I	Both H3K4-related demethylases; different complexes; shared regulatory targets
BAX×APEX1	0	0	0	0	0	0	I	Different compartments (OMM vs nucleus); regulatory coupling to purging
BAK1×APEX1	0	0	0	0	0	0	I	Same as BAX×APEX1
BAX×OGG1	0	0	0	0	0	0	I	Different compartments; death-BER regulatory coupling

CLASS I PAIRS — CDK/sirtuin coordination (Score ≤ 2)

Pair	C1	C2	C3	C4	C5	Total	Class	Key evidence
CDK6×CCND1	1	1	0	0	0	2	I	Direct binding but CDK6 has kinase-independent TF functions (Kollmann 2013); CCND1 also activates CDK4
CDK6×E2F1	0	1	0	0	0	1	I	Indirect: CDK6 → RB1 → E2F1; both have extensive independent roles
CDK6×RB1	1	1	0	0	0	2	I	Direct substrate; but CDK6 has kinase-independent functions; RB1 has CDK-independent roles
CDK4×CDK6	0	1	0	0	0	1	I	Same family; different regulatory contexts; no heterodimer
SIRT6×SIRT7	0	1	0	0	0	1	I	Different substrates (H3K9ac vs H3K18ac); co-regulation reflects shared NAD ⁺ dependency

CLASS I PAIRS — Reclassified from Class II (Score 2-3, classified I on molecular grounds)

Pair	C1	C2	C3	C4	C5	Total	Class	Reclassification rationale
FANCD2×BRCA1	1	1	0	0	0	2	I	Damage-induced interaction (C4=0); both have extensive partner-independent functions (C3=0); Bunting 2012
MFN1×OPA1	0	0	0	1	0	1	I	Different membranes (OMM vs IMM, C2=0); OPA1 has MFN1-independent cristae function (C3=0); Frezza 2006
BRCA1×RAD51	0	1	0	0	0	1	I	Indirect loading (BRCA1 → PALB2 → BRCA2 → RAD51); both have extensive independent roles; compensatory strengthening

EXCLUDED PAIRS (Ambiguous paralogs)

Pair	C1	C2	C3	C4	C5	Total	Decision	Rationale
BLM×WRN	0	1	0	0	0	1	Excluded	Both RecQ helicases; different diseases; non-overlapping substrates; no heterodimer
TET1×TET3	0	1	0	0	0	1	Excluded	Both dioxygenases; different tissue distributions; no heterodimer

Summary Statistics

- Class II (score ≥ 4): 8 pairs (mean score 4.25)

- Class I (score ≤ 2): 28 pairs (mean score 0.68)
- Excluded: 2 pairs
- Score 3 (borderline): 0 pairs
- The gap between highest Class I score (2, for NEIL1×XRCC1 and CDK6×RB1) and lowest Class II score (4, for 6 pairs) is 2 points — no pair falls in the ambiguous zone.

Sensitivity to Classification Boundary

- Moving the boundary from ≥ 4 to ≥ 3 adds 0 pairs to Class II (no pairs score 3)
- The classification is robust because the C1–C5 criteria create a natural bimodal distribution with a 2-point gap between classes