IAMPerformance Inter-Domain Research Institute

Heath W. Mahaffey · *Fri Forsker* · Principal Researcher and Founder

Entiat, Washington · hmahaffeyges@gmail.com

The Physics of Pathology

Informational Fidelity, Thermal Noise, and Cellular Computation

A Standalone Technical Evidence Report

EDEAR — Early Detection · End-of-life · Aging · Research

Contents

1 "If you can't stand the heat…" in the kitchen with EDEAR

Imagine a pot of water sitting on a stove. The burner underneath is set somewhere on the dial — low, medium, high — and the water sits at whatever temperature the burner has driven it to so far. There is no one in the kitchen. The pot heats slowly. If the burner is high, the water reaches a boil and starts spilling over the rim. If the burner stays on too long, the water boils away entirely, and at that point no amount of turning down the heat can recover the water in the pot. The pot in this kitchen is a human cell population. The burner is whatever has been driving that cell population to do more cellular work than it should — chronic inflammation, chronic infection, carcinogen exposure, autoimmune dysregulation, prolonged hormonal exposure. The water reaching a boil is the moment a tumor crystallizes and starts producing the symptoms that send a patient to a clinic. The kitchen is the human body.

Every clinical instrument in current use is some form of steam detection. AFP, PSA, CA 19-9, CEA — these go off when the water is already boiling. Mammography, colonoscopy, low-dose CT — these are walking through the kitchen with a flashlight, looking for the pot whose lid is rattling. PanSeer and Galleri are sensitive steam detectors, able to register the molecular signature of an already-active tumor before symptoms force the patient to seek care. That is a real advance over the flashlight. It is still steam detection. The water has to be boiling for the detector to fire. Everyone else is like a blind person in the kitchen sticking a finger in the water to guess how hot it is.

EDEAR is a thermometer. It does not wait for the water to boil. It reads where the cell population is currently sitting on the temperature scale — calibrated against thermodynamic floors derived from physics, not from clinical outcomes — and it reads in both directions. A patient running hot has water rising toward a boil. A patient running cold has water that has dropped below the healthy operating range, the signature the framework reports for chronic immunosuppression, post-chemotherapy states, advanced disease with marrow infiltration, late-stage cachexia. And because the framework reports the temperature paired with the burner setting documented in the patient's history — *H. pylori* for stomach, viral hepatitis for liver, smoking for lung, chronic systemic inflammation for inflammatory bowel disease — two readings six months apart give a rate of rise. A trajectory. Not a single coordinate.

That is the difference between reactive and proactive medicine. Knowing the burner setting and the current temperature gives a clinician time to turn the burner down before the water ever boils. Once the water has boiled away, no amount of turning down the heat can recover the water in the pot. That is the moment everyone else in the kitchen finally smells the smoke — and it is the moment when the option to do something about it has already passed. The framework's contribution is putting a thermometer in the kitchen, with a clear reading on the dial, while there is still water in the pot.

2 Executive Summary

EDEAR detects cellular-system drift in peripheral blood DNA methylation, years before any specific tissue localizes. The framework reads the customer's cellular state across eight architectural cell-classes, returns per-tissue and per-cell-type drift scores against thermodynamic floors derived from first principles, and routes the result through 17 clinical applications covering cancer pre-diagnostic detection, neurodegeneration, hematologic malignancy, and cardiovascular trajectories.

Imagine for a moment that every cell in your body is a computer chip. Not a metaphor for one — actually one, in the engineering sense. Each cell is a physical system that performs a defined computational workload, runs at a defined operating range, and has a measurable limit on how much information work it can do while still being itself. The body runs on eight different chip designs, one for each architectural class of cell — immune cells, secretory cells, cycling epithelial cells, terminally differentiated cells like neurons and cardiomyocytes, stromal cells, progenitors, and two grades of stem cell. Each design has a thermodynamic

floor: the minimum amount of cellular work a cell of that class must perform to maintain its own identity. EDEAR is the closest thing to a cellular thermometer that exists. It reads where each of your eight chip designs is sitting relative to its operating range, in a single peripheral blood draw, against floors that were calculated from first principles — the same first principles physics that describes black hole formation — before any patient was ever measured against them.

The headline detection claims, anchored to pre-registered SHA-locked validation runs:

- **More than a decade before clinical diagnosis: cellular-system drift is loudly detectable.** In samples drawn 10+ years before clinical breast cancer diagnosis, the Stage 1 immune-class A-score reads $d = +1.78$ (GSE51057, $n = 11$) and $d = +1.36$ (GSE51032, $n = 33$). This is a system-wide cellular-aging-drift signal, not a breast-cancer signal. Stage 2 cell-of-origin deconvolution at this window shows the drift is broadly distributed across pancreatic-duct ($d = +0.99\,/\,+0.70$), pancreatic beta cells ($+1.02\,/\,+0.94$), pancreatic acinar ($+0.91\,/\,+1.02$), kidney ($+0.73\,/\,+0.90$), head/neck-larynx ($+0.75\,/\,+0.81$), and colon epithelial ($+0.72\,/\,+0.65$) tiles concordantly across both cohorts. The breast tile reads near-null at this window ($d = +0.20\,/\,+0.10$).
- **Two years before clinical diagnosis: tissue localization emerges.** In the 24 months before workup catches the disease, breast-tile cell-of-origin localization reads $d = +0.43\,/\,+0.49$ in the same two cohorts (VAL-096). The hepatocyte tile rises to $d = +0.61\,/\,+0.61$. The earlier-elevated distant tiles attenuate (pancreatic-duct from $+0.99\,/\,+0.70$ at >10 yr to $+0.04\,/\,+0.26$ at 0–2 yr; head/neck-larynx from $+0.75\,/\,+0.81$ to $+0.11\,/\,+0.14$). This is the breast-LL-005 distributed-then-localized two-component temporal model — body-wide drift years before clinical workup, late-localizing tissue signal in the months before diagnosis.
- **At diagnosis, across hard-to-screen cancers, the signal is unmistakable.** Acute myeloid leukemia: $d = +3.71$ [+3.23, +4.20], 98.5% of AML samples above the healthy 95th percentile, $p < 1 \times 10^{-50}$ (VAL-082). Glioma cortical-neuron cell-free DNA fraction in standard EPIC peripheral blood: $d = +1.96$ [+1.62, +2.31] (VAL-090) — the first direct detection of brain-derived cfDNA at array resolution. Lung adenocarcinoma tumor tissue: paired $d = +1.020$ (VAL-063, TCGA-LUAD, $n = 29$ paired). Breast tumor tissue: paired $d = +0.675$ (VAL-060, TCGA-BRCA, $n = 86$ paired). Colorectal tumor tissue: paired $d = +0.724$ cycling-class architecture (VAL-062) and $d = +1.066$ tumor-infiltrating immune compartment (VAL-061). HCC plasma cfDNA: $d = +0.634$ with monotonic dose-response from healthy through fibrosis through cirrhosis to HCC (VAL-059).

No other consumer methylation product does any of this. Grail's Galleri is at-diagnosis only, requires a clinician order, costs $950 per test, and reports approximately 50% sensitivity at 99.5% specificity for stage I cancer. TruDiagnostic and Elysium run aging clocks and per-disease statistical risk scores trained one disease at a time. EDEAR runs the same physics-derived measurement on every blood draw — same Xu-538 panel, same $H_{\min}(\text{immune})$, same three-stage pipeline — and routes the result through disease-specific clinical applications. Same blood. Same A-score. Seventeen clinical applications, all reading the same underlying measurement.

The most authoritative source for every numerical claim in this paper is the live GAPE Evidence Report HTML (over 18,000 lines, VAL-001 through VAL-113 sealed and counting). Every claim here is anchored to a pre-registered SHA-256-sealed validation run timestamped before the analysis was performed. Every script and every result is on GitHub. A skeptical reviewer can re-run any validation from the public source and confirm the numbers.

Behind every one of these numbers is a person. The 1.78-sigma signal a decade before diagnosis is a real woman who walked into a real exam room ten years before her cancer would be found, and her body was already trying to say something nobody could hear. The 3.71-sigma signal at acute leukemia diagnosis is the patient who, in the current standard of care, is usually only caught when she is already sick enough to need

a CBC. The dose-response signal in liver cancer plasma is the cirrhosis patient sitting in a clinic right now whose cellular state is already telling a story that imaging cannot yet see. We did not start with a clinical question. We started with a physics question, and we ended up holding an instrument that reads what bodies are quietly saying years before our current tools can hear it. That is what this paper is about.

3 How we got here — same physics, three domains

A reasonable question is how a Columbia River grid operator ends up writing the physics behind a cancer detection product. The honest answer is that the underlying physics did not start with cancer. It started with a question about how the universe pays the cost of its own existence, and that question turns out to apply at every scale. The clearest way to explain what the framework does is by analogy to a different kind of chip — the kind in your computer.

Every silicon chip ever manufactured has specifications. A 5-nanometer transistor and a 14-nanometer transistor have different operating ranges because the physics of charge confinement at those scales is different. Apple ships several chip families precisely because no single chip design serves every workload — the chip in a phone has different thermal constraints than the chip in a laptop, which has different constraints than the chip in a data-center server. Each design has a maximum amount of work it can perform, a minimum below which it cannot reliably hold state, and a range in between where it operates the way it was engineered to operate. Push a chip past its thermal envelope and it throttles, mis-switches, accumulates errors, and eventually fails. The specifications are not optional. They are what the chip is.

The semiconductor industry hit this wall at scale in 2005. Dennard scaling — the rule that smaller transistors could run at higher density without an increase in power per unit area — broke down. The transistors had been pushed to a point where the writing surface could no longer dissipate the heat the workload required. The industry's response was not to make the existing chips work harder. The response was to build new writing surfaces: multi-core processors, 3D-stacked memory, chiplet architectures. The system created new encoding capacity because the old encoding capacity had been saturated. That same pattern — a writing surface saturating, the system being forced to create new capacity — is the structural identity that runs through everything in this paper.

The body works the same way, on eight different chip designs. Immune cells, secretory cells, cycling epithelial cells, terminally differentiated cells like neurons and cardiomyocytes, stromal cells, progenitors, tissue-resident adult stem cells, and pluripotent stem cells — each one is a physically distinct architectural class with its own thermodynamic operating range. The hundreds of differentiated cell types in the human body all belong to one of these eight architectural classes, the same way the dozens of chip products Apple ships all belong to a smaller number of underlying chip families. A mature neuron is not a colon epithelial cell, in the same way an iPhone chip is not a server chip. The materials are different. The substrate is different. The workload is different. The minimum amount of cellular work each class must perform to remain that class is different. Each architectural class has a thermodynamic floor, calculable from first principles — the same way the floor of a 5-nanometer transistor is calculable from the physics of charge confinement.

The framework that produced EDEAR measured those eight floors using the same physics that describes black hole formation, applied at the scale where cells live. The path from one to the other is not as long as it sounds. Over the last sixty years physicists worked out that any system that writes information — a hard drive, a transistor, a star, a chromosome — has to pay a minimum thermodynamic cost for every bit it commits. They worked out that black holes are the densest information-storage objects in the universe, and that the surface of a black hole has a measurable thermal temperature determined by its mass. They worked out that gravity itself, at the deepest level, is what happens when information is being written to a surface that has a finite capacity. The same physics that says a star runs out of room to be a star and collapses into a black hole says a cell runs out of room to be the kind of cell it is and either repairs the damage or breaks.

The capacity is calculable from first principles for each of the eight architectural classes a human cell can belong to, before a single patient is ever measured. The full step-by-step derivation is in the Floor Breach appendix at the end of this paper for anyone who wants to verify the math. The point for a clinician is simply this: the eight floors EDEAR scores against were derived from physics, not fitted to clinical data, and the same equation governs both the information capacity of a black hole horizon and the information capacity of a single human cell's chromatin. This is not a metaphor. It is a structural identity that has been derived step by step.

What this means clinically is the part most people do not see coming. A white dwarf collapses into a black hole when its rate of information production saturates the encoding capacity of its local horizon. The star can no longer pay the thermodynamic cost of remaining a star. So the system creates a new local encoding surface — the black hole horizon itself — because the existing one has been overwhelmed. A cell undergoes malignant transformation through the same mechanism. When the rate of irreversible information processing at the chromatin surface saturates the cell's $H_{\min}$ floor, the cell can no longer pay the thermodynamic cost of remaining the architectural class it was. So the cell, like the star, creates a new local encoding surface. That new encoding surface is what we call a tumor. Cancer, viewed at the architectural level, is the cellular instance of the same physics that produces a black hole. Both are floor breach events. Both are predictable before they happen, because the writing surface saturates gradually and observably before the system runs out of options. EDEAR is the instrument that reads where the cellular writing surface is sitting relative to its capacity, in time for the patient and the doctor to do something about it.

Every information-processing system pays a price. A transistor switching between logic states, a qubit holding a superposition, a cell maintaining its methylation pattern, the universe itself preserving causal structure — each one runs an ongoing transaction between two quantities. On one side, the useful information work the system performs, which a physicist would call its *informational fidelity*. On the other side, the *thermodynamic noise floor* that any physical system is forced to operate above, set by temperature, by system size, by the granularity of the underlying degrees of freedom. The ratio of those two — informational fidelity over thermodynamic noise — determines whether the system can do its job. Push the ratio too low and the transistor mis-switches, the qubit decoheres, the cell loses regulatory control, the universe forgets what state it was in. Same ratio. Different scales.

What the framework underlying EDEAR observed is that this ratio has a calculable floor for any specific architecture. A 5-nanometer transistor has a different floor than a 14-nanometer transistor because the physics of charge confinement at those scales is different. A trapped-ion qubit has a different floor than a superconducting qubit because the decoherence channels are different. A mature neuron — a terminally differentiated cell that has stopped dividing and locked in its identity — has a different floor than a colon epithelial cell that divides every few days. Different architecture. Different floor. Same calculation.

Two prior framework instruments came before the biological application. The semiconductor version — **SCAPE** — recovers the 2005 Dennard scaling breakdown from first principles and reads where current process nodes sit relative to their architectural floors. The quantum-computing version — **QAPE** — does the same thing for qubit architecture classes spanning trapped ions, superconducting circuits, neutral atoms, and photonic systems. Both instruments calibrate the same way: an MCMC procedure that updates a prior estimate of the floor against published reference data for that architecture class until the posterior converges (R-hat below 1.001, 800,000 posterior samples per class).

The biological application — **GAPE**, the engine inside EDEAR — is the third domain. The eight architectural cell-class floors listed in Section 4 came out of the same MCMC procedure applied to published reference methylation data for healthy cells of each class. The numbers were not learned from cancer training data and they were not fitted to any clinical outcome. They are the framework's predicted thermodynamic floors for the eight cell architectures, frozen on April 6, 2026 prior to every cohort analysis in the validation record. A patient with a methylation entropy reading at her predicted floor is operating at the framework's idea

of healthy for her cell class. A reading above the floor means the cell class is doing more work than the floor — drift, activation, disease. A reading below means the cell class is doing less work than the floor — suppression, induced tolerance, advanced disease.

The reason the framework works across domains is that the underlying physics does not care what scale it is on. The framework was first developed and tested in cosmology — the largest scale physics has access to. The cosmological version is on the public record at the IAM-Validation GitHub repository (Mahaffey, 2026b) and has been tested against Planck 2018 cosmic microwave background data and KiDS-Legacy weak gravitational lensing surveys (Mahaffey, 2026a), both of which probe how matter and energy organize themselves on scales of billions of light-years. The same calculation that produces the eight EDEAR cell-class floors produces those cosmological observables. From inside the work, this connection is obvious — we built the framework on cosmology and watched it predict cellular biology on the way down. From outside the work, it is not obvious at all, and a clinician reading this paper has every reason to ask why a cosmological calculation should have anything to say about a single human cell. The answer is in the next paragraph: the framework makes a specific, testable prediction about how cellular fidelity should change across mammalian species spanning four orders of magnitude in lifespan, and that prediction holds without recalibration. The cosmological numbers come down. The mammalian numbers come up. They meet at the cellular scale where a doctor is actually working.

The framework has been tested across the mammalian lineage. The same A-score that flags pre-diagnostic cancer in human blood at $A \geq 1.05$ has been applied to 34 mammalian species spanning 12 taxonomic orders — from the 2.5-year-lifespan common shrew through 4-year mice and 8-year guinea pigs to 122-year humans, 90-year killer whales, and the 211-year bowhead whale. The result is a clean monotonic relationship between maximum lifespan and architectural fidelity (Pearson $r = -0.919$, $p = 1.6 \times 10^{-14}$ across $n = 34$ species; VAL-034 in the IAM-Validation repository). Short-lived mammals sit above the cancer detection threshold $A = 1.05$ with no recalibration; long-lived mammals sit below it. The two groups separate at the threshold without overlap (Cohen's $d = 1.99$, $t = -8.2$, $p = 2.5 \times 10^{-9}$). The same threshold the framework uses to call DETECTABLE in a 60-year-old human at her 12-year pre-diagnostic blood draw is the threshold that separates the mouse from the elephant. This is the parallel the previous paragraph was pointing at: the same physics that organizes the cosmos also organizes the lifespan-fidelity relationship across the entire mammalian lineage, and the eight cellular floors EDEAR uses sit at the intersection. The framework is not a biology framework dressed up in physics vocabulary. It is a physics framework that turned out to predict biology.

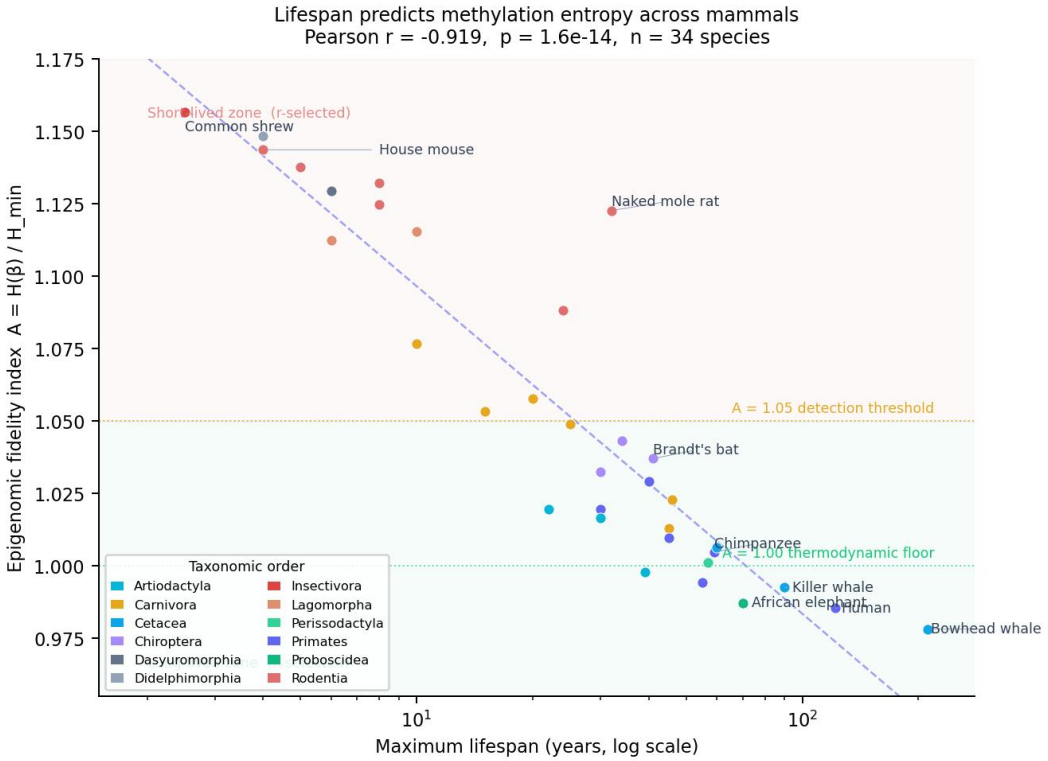

Figure 1: *Lifespan vs A-score across 34 mammalian species spanning 12 taxonomic orders. The A-score (epigenomic fidelity index, $A = H(\beta)/H_{\min}$) is the same metric EDEAR computes for human cancer detection, applied here to published methylation data for each species. Pearson $r = -0.919$, $p = 1.6 \times 10^{-14}$. Short-lived mammals (mouse, shrew, rat) sit at $A \approx 1.13$–1.16 — well above the $A = 1.05$ cancer detection threshold. Long-lived mammals (human, killer whale, bowhead whale, African elephant) sit at $A \approx 0.99$–1.01 — below the thermodynamic floor reference. The naked mole rat (32-year lifespan despite rodent body size) sits at $A \approx 1.12$, consistent with its rodent architecture rather than its lifespan. VAL-034 in the IAM-Validation repository.*

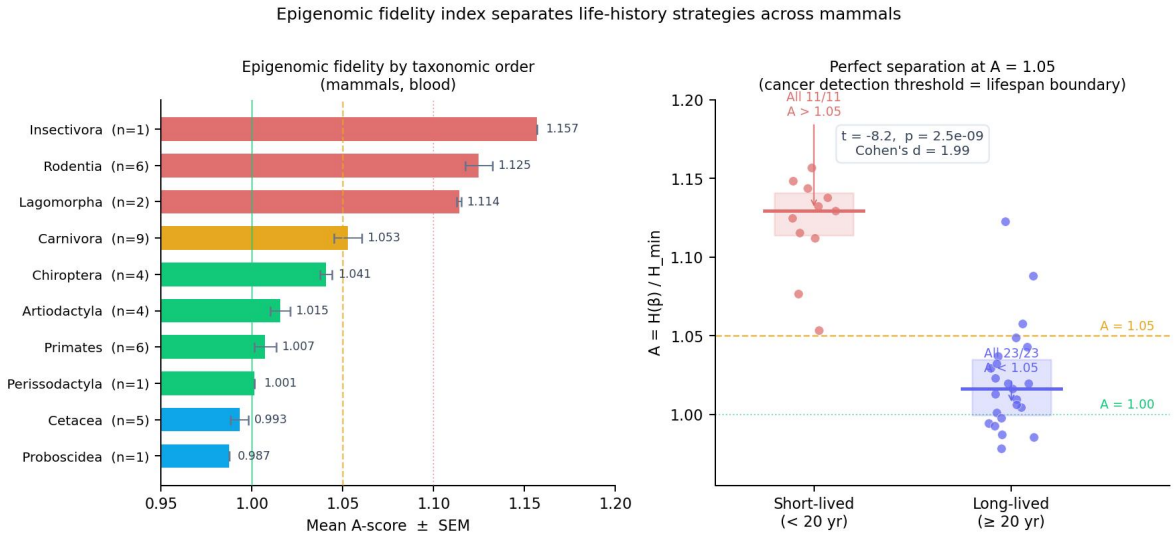

Figure 2: ***Left:*** *mean A-score by taxonomic order, sorted by lifespan. Short-lived orders (Insectivora, Rodentia, Lagomorpha) sit above the $A = 1.05$ threshold; long-lived orders (Cetacea, Proboscidea) sit below the $A = 1.00$ thermodynamic floor.* ***Right:*** *short-lived (< 20 yr) vs long-lived (≥ 20 yr) split across the 34 species. All 11 short-lived species are above $A = 1.05$; all 23 long-lived species are below. Perfect separation at the same threshold the framework uses as the DETECTABLE tier boundary. Cohen's $d = 1.99$, $t = -8.2$, $p = 2.5 \times 10^{-9}$.*

This is what "same physics, three domains" means in practice. The cancer detection threshold $A = 1.05$ was not picked to fit human breast or colorectal cohorts. It is the threshold at which a 4-year-lifespan mouse is on the wrong side of the partition and a 122-year-lifespan human is on the right side, derived from the same calculation that grounds the eight EDEAR cell-class floors. EDEAR reading a customer's blood at $A = 1.07$ means the same thing it would mean for any mammal at any age in the dataset above: cellular system fidelity below the threshold associated with long life. That is the framework's claim. The rest of this paper is what happens when the claim is tested against pre-diagnostic and at-diagnosis human cohorts.

4 How EDEAR reads cellular state — the architectural A-score

Every cell architectural class — immune, secretory, cycling, terminal, stromal, progenitor, stem-adult, stem-pluripotent — has an associated thermodynamic floor: the minimum cellular work the class can do while still being that class. Mature neurons sit at one floor. Rapidly cycling colon epithelial cells sit at another. Immune cells responding to chronic activation sit at another. The framework predicts these floors from first principles — they are not learned from training data. The values are MCMC-calibrated against healthy reference populations from the G-002 and G-003b posterior runs (R-hat below 1.001, 800,000 posterior samples) and frozen before any cancer-cohort analysis is performed.

The eight architectural class floors, frozen 2026-04-06 prior to every cohort analysis:

Architectural class	$H_{\min}$	What this class is
immune	0.83####	Circulating leukocytes; the universal Stage 1 substrate (peripheral blood is ∼70% immune cells)
secretory	0.84####	Glandular epithelium making something for export — breast ductal, hepatocyte, prostate epithelial, pancreatic exocrine
cycling	0.85####	High-proliferation epithelium that turns over every few days — colon, lung, gastric, bladder, cervical, kidney epithelial
terminal	0.77####	Post-mitotic identity-locked cells — mature neurons, cardiomyocytes, skeletal muscle fibers
stromal	0.86####	Connective and structural cells — fibroblasts, vascular endothelium, smooth muscle
progenitor	0.85####	Lineage-committed dividers upstream of mature cells — e.g. erythrocyte progenitors, basal epithelium
stem-adult	0.87####	Tissue-resident multipotent stem cells — intestinal crypt, hematopoietic, hair-follicle bulge
stem-pluripotent	0.98####	Pluripotent embryonic-equivalent state; the highest predicted floor in the catalog

The eight cell-class floors are framework predictions, calibrated against healthy reference populations from the G-002 and G-003b posterior runs (R-hat < 1.001, 800,000 posterior samples) and frozen before any cancer-cohort analysis. Higher $H_{\min}$ means the class is required to maintain more cellular work per unit substrate; lower $H_{\min}$ means it can rest closer to silence.

The architectural A-score is the customer's reading on a class-specific gauge. For each CpG site in the panel of interest, compute the binary Shannon entropy of the methylation β value and divide by the framework's class-specific thermodynamic floor:

$$A_i = \frac{H(\beta_i)}{H_{\min}(\text{class})} \qquad \text{where} \qquad H(\beta) = -\beta \log_2(\beta) - (1-\beta)\log_2(1-\beta). \tag{1}$$

The patient's per-class A-score is the mean of those values across the panel CpGs:

$$A = \text{mean}\left[\frac{H(\beta)}{H_{\min}(\text{class})}\right] \quad \text{over the panel CpGs.} \tag{2}$$

A patient with $A = 1.0$ sits exactly at the framework's predicted thermodynamic floor — the healthy, well-functioning state for that cell class. Values **above** 1.0 indicate entropy elevation: the cellular compartment is doing more work than the floor, consistent with chronic activation, tumor-driven drift, or immune reorganization. Values **below** 1.0 indicate entropy depression: the compartment is doing less work than the floor, consistent with post-treatment immunosuppression, induced tolerance, advanced disease, or primary immunodeficiency. **Both directions are real signals.** The patient-facing tier vocabulary spans both directions across six bins: BELOW NORMAL, NORMAL, MARGINAL, DETECTABLE, URGENT, FLOOR BREACH.

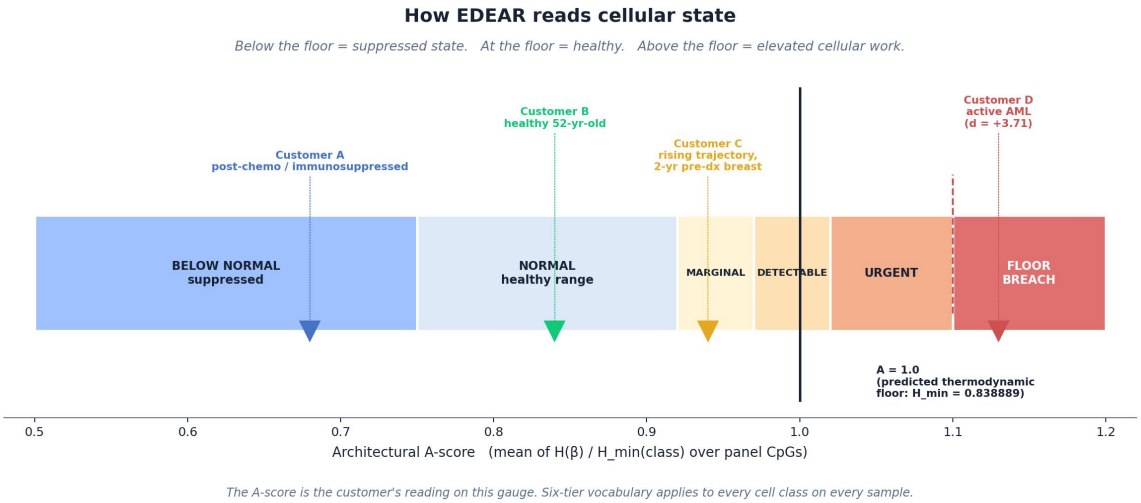

Figure 3: *The cellular-state gauge that EDEAR returns for each architectural cell class. Patient A-score is rendered against the framework's predicted thermodynamic floor (vertical dashed line at $A = 1.0$) with tier bands. The same gauge structure applies to every architectural class — only the underlying $H_{\min}$ anchor changes from class to class.*

A patient who picks up an EDEAR report is not going to see Shannon entropy or thermodynamic floors. They are going to see something that looks like a fuel gauge — a needle, a healthy band, a few tiers labeled in plain English. The math underneath is what makes the gauge honest. The patient sees: am I in range, am I drifting, is something happening I should know about. Everything in the rest of this paper exists to make that gauge tell the truth. The complexity belongs in the foundation. The simplicity belongs in the moment when a person finds out how their body is doing.

5 How the framework keeps itself honest

A reasonable question, the harder a framework's claims get, is how the people building it are protecting it from themselves. Anyone can fit a formula to data after the fact and call the fit a discovery. The framework underlying EDEAR was built specifically to make that move impossible. Every prediction the framework produces has a timestamp. Every threshold has a sealed hash. Every atlas the engine reads has a SHA-256 fingerprint and a public provenance trail. The discipline that surrounds the science is not optional plumbing — it is what allows the claims in this paper to be claims at all.

The eight architectural cell-class floors listed in Section 4 were frozen on April 6, 2026. That is not a marketing date. It is the timestamp of a Git commit on the public IAM-Validation repository, before any cancer cohort in this paper was scored against those floors. Anyone reading this paper today can verify that the $H_{\min}$ values predated the validation evidence by going to the repository and reading the commit history. The same is true for every directional panel the engine uses. The Alzheimer's 7-CpG Rule A panel was sealed on April 23, 2026 at 07:23:53 UTC; the SHA-256 hash of the panel JSON is in the validation record. The 324-CpG pancreatic directional fallback panel was sealed before the TCGA-PAAD holdout was scored against it. Every panel, every threshold, every prediction lives on a timeline that cannot be edited after the fact.

The same discipline applies to the reference atlases the engine deconvolves against. Every atlas EDEAR uses lives in a versioned vault on GitHub with a SHA-256 fingerprint, a citation back to the published paper that produced it, the data license, and a record of which calibration validation run anchored it. The vault holds eleven calibrated cell-of-origin atlases as of this writing — Moss 2018, Loyfer 2023, Salas IDOL, EpiSCORE

pan-tissue and per-tissue references for breast, prostate, bladder, esophageal, oral, and heart, the Caggiano CelFiE TIM panel, and a gastric organoid mucosoid reference built from the Boccellato 2022 cohort. Each one has a calibration outcome on file. Each one was tested against a structurally separated healthy reference cohort before it was allowed to score a single disease sample. When an atlas does not pass calibration, it is marked deferred until it does. When an atlas passes only partially — four of nine tissue tiles clearing the strictest separation floor, for instance — it is deployed with that limitation labeled openly in the patient-facing report rather than hidden.

The pre-registration discipline matters most when a prediction fails. The cookbook that runs EDEAR has documented dozens of these failures across hundreds of validation studies, and every one of them is on the public record. A pre-locked prediction that the gastric cell-of-origin tile would read negative under a published direction model fired positive instead — and the failure forced a tighter understanding of what tissue-class substrate shifts do to absolute cross-cohort comparisons. A pre-locked prediction that sorting blood cells from a Crohn's disease cohort would amplify the within-cell-type signal returned the opposite direction — and the failure re-framed what every immune deconvolution atlas in the cookbook actually measures, cookbook-wide. A pre-locked prediction that the universal Stage 1 immune panel would elevate on chronic inflammatory disease returned a clean null — and the failure clarified that Stage 1 is a tissue and cancer signal, not a generic chronic-inflammation marker. None of these failures were quietly walked back. Each one was sealed, dated, and published as the lesson it was. Each one made the framework smaller in scope, more specific in claim, and more honest about its own boundaries.

The check gates that surround each validation run are how this is operationalized. Before any β values from a new cohort are scored against the framework, the cohort first passes a substrate-validity check — does the data look like real array data, or is the bimodal β distribution flat in a way that says the file has been pre-processed into something that is no longer raw. Then it passes a panel-coverage check — does the cohort actually carry enough of the panel CpGs at a high enough detection rate to score reliably. Then it passes a healthy-reference baseline check — does the cohort's own healthy controls sit on the same substrate baseline as the framework's calibrated reference, or is there a tissue-class shift that needs to be acknowledged before any disease comparison is run. The gates fire honestly. Cohorts have been failed at every gate. Sealed validation outcomes carry the gate result regardless of whether the result was the one we hoped for.

What this means, for any reader trying to evaluate whether the claims in this paper are real or fitted, is that the question can be settled empirically. Clone the repository. Read the commit timestamps. Verify that the prediction predates the data. Run the validation scripts on the public cohorts and confirm the numbers reproduce. The fact that this is possible — and that the framework was deliberately built to make it possible — is the difference between a research program and a story. Stories survive on the storyteller's authority. Research survives on the public record. The decision to run the framework this way was made years before any cancer cohort was scored, because the only way a framework like this earns trust is by giving every skeptical reader the rope to verify it themselves. That is the discipline the rest of this paper rests on.

6 The universal three-stage pipeline

Every customer's first test is the same. There is no disease-specific Stage 1 substrate, no disease-specific Stage 1 class, no disease-specific Stage 1 $H_{\min}$. The first line of defense is always the immune class because peripheral blood is approximately 70% immune cells, and the immune compartment sees everything — breast, colorectal, lung, pancreas, liver, prostate, gastric, bladder, cervical, kidney, glioma, Alzheimer's, chronic infection, autoimmune drift — through its response to upstream tissue change.

EDEAR runs a *run-everything-every-time* architecture (signed off 2026-04-26, internal reference `EDEAR_PIPELINE_OFFICIAL_REFERENCE` v2). Every blood draw goes through Stage 1, Stage 2, and Stage 3 in parallel. No conditional gating. The compute cost is seconds to minutes per IDAT. The output is a complete

cellular-state report covering every cell class and every tissue the framework can resolve. The patient's report displays anomalies and collapses uninformative tiles for readability; the underlying scoring is exhaustive on every IDAT.

Why run-everything matters. The Stage 1 immune A-score can null on diseases that drive CpGs bidirectionally — the AD-instance pattern, formalized as CCL-030. VAL-050 returned $d = +0.077$ (null) on Alzheimer's; VAL-051 directional scoring on the same 7 CpGs same cohort returned $d = +0.624$. Heme-epic looks for Stage 2 NULL on solid organs as the diagnostic feature. PSP/CBD reads cortical-neuron at $d = -0.51$ (BELOW NORMAL). A customer could have early AD plus early breast plus chronic inflammation plus cardiovascular drift simultaneously. Gating on the first signal that crosses threshold means filtering the others through that lens. We measure all of them and let the anomaly stack tell the story.

6.1 Stage 1 — the immune-system red-flag test

Stage 1 reads the immune compartment of the blood sample for architectural drift. This is the system-wide red flag. When it fires, something is going on with the customer's immune system as a whole — what specifically is happening is what Stages 2 and 3 work out.

The primary panel is the **Xu-538** — 538 CpG sites originally derived from the U.S. Sister Study breast-cancer pre-diagnostic cohort and replicated in EPIC-Italy (Xu et al., 2020). Panel SHA-256: `ada672960563343334a6d6` (truncated; full hash on GitHub).

Stage 1 also runs disease-specific directional panels in parallel — every IDAT, every time:

- **Pooled A_{immune} on Xu-538** — primary scoring. $A = \text{mean}(H(\beta)/H_{\min}(\text{immune}))$ across the 538 CpGs. Direction-agnostic at the per-patient level due to Shannon symmetry.
- **AD 7-CpG Rule A directional panel.** Frozen ± 1 directions, sealed VAL-051 SEAL 2026-04-23 07:23:53 UTC. $A_{\text{dir}} = \sum(z(\beta_{cpg}) \times \text{direction}_{cpg})$. Cohen's $d = +0.624$ on AIBL holdout; $d = +0.33$ cross-platform replication on AddNeuroMed.
- **Pancreatic 324-CpG directional fallback panel** (VAL-069). GSE49149-trained, z-score normalized so $H_{\min}$-independent. Recovers signal on TCGA-PAAD holdout at paired $d = +1.51$ [+0.43, +2.59], $p = 6.4 \times 10^{-5}$. Recommended primary Stage 1 metric for PDAC because pooled-entropy nulls on bidirectional drift.
- **Kresovich mBCRS 100-CpG comparator** (Kresovich et al., 2019). Reference-only; reported alongside Xu-538 pooled for breast comparison.

Stage 1 output bundle, every IDAT: $A_{\text{immune,pooled}}$ + $A_{\text{dir,AD}}$ + $A_{\text{dir,PDAC}}$ + Kresovich comparator + age-matched percentile vs the 80-cell healthy baseline + per-class tier call. Stage 1 flags architectural departure in the immune compartment. It does not identify the source. A flag means "something is driving immune drift"; what drives it is what Stage 2 (for solid-organ cancers) and Stage 3 (for non-solid-organ immune conditions) work out.

6.2 Stage 2 — tissue-of-origin deconvolution, layered atlas architecture

Stage 2 asks where the cellular-state signal is coming from. When tissues anywhere in the body undergo cellular turnover, fragments of their DNA leak into the bloodstream as cell-free DNA, each fragment carrying the methylation pattern of its tissue of origin. EDEAR mathematically deconvolves the bulk methylation signature against published reference atlases of pure cell-type methylation patterns using non-negative least squares (`scipy.optimize.nnls`). The pipeline runs a layered atlas architecture — multiple atlases in parallel, with the higher-resolution sorted-cell entries taking precedence where they exist. The atlas state as of 2026-04-26:

Atlas	Cell types covered	Status
Moss 2018 (Moss et al., 2018)	18 solid tissues — colon, lung, gastric, bladder, cervical, kidney epithelial; hepatocyte, pancreatic exocrine, breast ductal, prostate epithelial; neuron, oligodendrocyte; vascular endothelial, fibroblast; neutrophil, lymphocyte, monocyte, hematopoietic stem cell.	In production (VAL-041)
Loyfer 2023 array atlas (Loyfer et al., 2023)	26 sorted cell types including cortical neurons, vascular endothelial cells, cardiac left atrium, sorted CD4 / CD8 / NK / B / monocyte / neutrophil cells, pancreatic duct cells. Adds resolution Moss does not have.	In production (VAL-090, VAL-091)
Salas 2018 IDOL (Salas et al., 2018)	6-cell immune sub-composition baseline (CD4T, CD8T, NK, B, monocyte, neutrophil). Stage 3 production atlas.	In production
UniLIFE Guo 2025 (Guo et al., 2025)	19-cell immune resolution (1,906 CpGs × 19 cell types) including 12 adult-specific subtypes — regulatory T cells, naive B cells, memory subsets, eosinophils, basophils. Lifespan-spanning.	Production overlay (VAL-095)
EpiSCORE per-tissue references (Zhu et al., 2022)	42 cell types across 13 solid tissues, source pan-tissue atlas Queue 1. Six per-tissue sub-panels CpG-bridged from Entrez gene IDs to HM450 array CpGs and calibrated against a structurally separated healthy reference cohort: HeartRef (5 cardiac cell types: CM/EC/FB/MP/SMC, VAL-111/112); BreastRef (5 breast cell types, VAL-094); ProstateRef (6 prostate cell types, VAL-117); BladderRef (4 bladder cell types, VAL-119); EsoRef (8 esophageal cell types, largest cross-tile separation observed across any per-tissue bridge calibrated to date, VAL-124); OEref (9 oral epithelium cell types, partial-floor calibration sealed with explicit per-tile annotation, VAL-125).	Six per-tissue refs in production; pan-tissue Queue 1
Caggiano CelFiE TIM (Caggiano et al., 2021)	19 cell types including bulk heart tissue + sorted endothelial + immune subsets, array-bridged via HM450 hg19 manifest CpG-in-region intersection (254 CpGs). Strongest single-tile cardiac signal in PAH.	In production (VAL-113)
Tanaka 2025 neural cfDNA	6 neural cell types — cortical / dopaminergic / spinal motor neurons, astrocytes, Schwann cells, microglia. Distinguishes Alzheimer's from Parkinson's from ALS from MS at the cell-type level.	Queue 1 — integrating
Cuadrat 2023 cardiac extended (Cuadrat et al., 2023)	28-cell extended atlas adding bulk ENCODE heart tissues (right atrium, left ventricle, coronary artery) on top of the 25-cell Moss base. Built from raw ENCODE EPIC IDATs.	Queue 2 — engineering pending
MARLIN leukemia ref. (Capper et al., 2025)	2,540 acute leukemia samples (1,461 AML, 686 B-ALL, 266 T-ALL) with subtype labels.	Queue 1 — integrating (VAL-084)
BoccellatoStomachRef HM450	Six gastric organoid mucosoid tile references (Antrum / Corpus / Fundus × undifferentiated / differentiated states), built from the Fritsche & Boccellato 2022 cohort (GSE141660). HM450-restricted from EPIC source (380,467 CpGs) to match the cookbook's HM450 calibration substrate. First gastric-tissue-specific cell-of-origin reference in the vault.	In production (VAL-123)

Layered atlas architecture as of 2026-05-03. Higher-resolution sorted-cell atlases take precedence where they overlap with bulk-tissue references. Green = in production; yellow = integration in progress; gray = engineering work pending.

Each atlas in the vault has been earned, not assumed. Before any atlas is allowed to score a single disease sample, it is calibrated against a structurally separated healthy reference cohort — TCGA-KIRC plus TCGA-PRAD adjacent-normal HM450 sesame Level 3 ($n = 210$) is the standing anchor used by every recent calibration in the table above. The calibration produces per-tile healthy-floor thresholds that are sealed with a SHA-256 hash and frozen as Type-1 calibration artifacts before any disease cohort is scored. When an atlas separates its tiles cleanly on the healthy substrate, it is sealed in production. When an atlas separates only partially — four of nine tissue tiles clearing the strictest separation floor on OEref, for instance — it is deployed at conservative thresholds with the partial-floor status labeled openly in the patient-facing report rather than hidden. When an atlas does not separate at all on its calibration cohort, it is marked deferred and not used until a path forward is found. The vault grows by passing this gate, not by adding files to a folder.

For each cell type the atlases recover, Stage 2 reports the recovered fraction, a per-class A-score against the same $H_{\min}$ anchor framework Stage 1 uses, and the same six-tier call. Every cell type, every sample, every time. Stage 2 outputs are tagged with the platform (450K vs EPIC) so platform-stratified thresholds can be applied where coverage gaps matter — ad-LL-006 (the AddNeuroMed cross-platform NNLS routing artifact, 8% Loyfer-CpG coverage gap) made this a hard requirement.

6.3 Stage 3 — immune sub-composition with parallel atlas overlay

Stage 3 reads the immune compartment specifically — what fraction of the blood sample is each kind of immune cell. Production runs Salas 2018 6-cell deconvolution as the baseline (CD4T, CD8T, NK, B, monocyte, neutrophil). UniLIFE 19-cell runs as a parallel overlay; UniLIFE's adult-specific subtypes surface signals that Salas pools out — regulatory T-cell expansion, naive B-cell expansion, lineage shifts that route to specific disease cards. Both layers are reported.

VAL-095 head-to-head head-line finding (signed off 2026-04-26): two replicating breast pre-diagnostic immune signatures that Salas 6-cell pooled deconvolution does not surface at the same magnitude. Regulatory T-cell (aTreg) expansion at >10 years pre-diagnosis: GSE51057 $d = +1.26$, GSE51032 $d = +0.79$, both 95% CIs exclude zero. Naive B-cell (aBnv) expansion at 0–2 years pre-diagnosis: GSE51057 $d = +0.44$, GSE51032 $d = +0.49$, both 95% CIs exclude zero. Salas catches the broad pre-diagnostic shift correctly; UniLIFE adds the resolution that surfaces the specific compartments driving it.

What Stage 3 measures, more precisely, is cell-type *composition* shifts in mixed-population substrates. This was sharpened by a recent Crohn's disease validation in which a sorted-cell sub-experiment failed a pre-locked mixture-attenuation prediction in the opposite direction expected. Sorting individual immune cell types from the same blood draw returned a smaller signal than the whole-blood substrate, not a larger one. The cleanest interpretation is that diseases like Crohn's drive a population-fraction shift in peripheral blood — T-cell expansion with proportional myeloid contraction — and that signal is detectable only because the substrate is mixed. Sorting the cells separates the populations and removes the relative-fraction signature that Stage 3 was reading. The same interpretive frame applies to autoimmune disease, chronic viral infection, sepsis, and graft-versus-host, all of which produce population shifts in mixed peripheral blood without producing within-cell-type chromatin drift. Cancer-associated immune microenvironment readings (advanced gastric cancer's T-cell + myeloid depletion in tumor tissue; bladder cancer's broad mixed-infiltrate signature) read the same way — composition shifts in mixed-population substrates. Stage 3 is not measuring drift inside individual cells. It is measuring how cells have rearranged themselves.

6.4 Platform: Illumina EPIC v1 (850K) at launch

EDEAR runs on standard Illumina EPIC v1 (850K) methylation arrays — the same platform every published cohort in the validation evidence used. Coverage of the panels and atlases against the official Illumina manifests:

- Xu-538 panel: 100% on 450K, 94.8% on EPIC v1, 90.0% on EPIC v2.
- Loyfer 25-tile Stage 2: 100% on 450K, 100% on EPIC v1, 90% on EPIC v2.
- UniLIFE 19-cell Stage 3: 100% on 450K, 100% on EPIC v1, 97% on EPIC v2.
- Salas 6-cell Stage 3: 100% across all three platforms via the appropriate 450K legacy or EPIC panel version.

EPIC v1 is the launch platform. It matches the substrate of every published validation cohort at over 95% panel coverage and is what every commercial methylation lab processes today. EPIC v2 is supported as a research platform; full migration is gated on a panel calibration step for the 10% of Xu-538 CpGs Illumina dropped.

The principle behind running everything every time is one of the things I am most quietly proud of in this design. We do not decide ahead of time what to look for. A forty-year-old woman with a family history of breast cancer, a sixty-year-old man with a fatty liver, a smoker who quit twelve years ago, a daughter watching her mother lose her words to dementia — they all get the full report from the same blood draw. The instrument looks at every cell class and every tissue, every time, and the disease cards interpret what shows up. This is what a system designed to serve patients should do. Not "what we already suspect," but "what is the body actually telling us today." The current standard of care narrows the search to the disease the doctor is already worried about. EDEAR widens it to the disease the patient may not yet know they have.

7 What the data show — pre-diagnostic detection, two cohorts, replicated

7.1 Stage 1 cellular-system drift in pre-diagnostic peripheral blood

EPIC-Italy GSE51057 cohort (Severi et al., 2014) — Italian arm of the European Prospective Investigation into Cancer and Nutrition, 329 samples (146 breast cancer cases, 177 cancer-free controls), 450K methylation, blood drawn at study enrollment with linkage to subsequent cancer diagnosis through the Italian national cancer registry. Cases stratified by time-to-diagnosis (TtD) windows. Stage 1 immune-class A-score (Xu-538 panel, $H_{\min}(\text{immune}) = 0.83####$) climbs monotonically with longer pre-diagnostic interval:

- 0–2 years pre-diagnosis: $d = +0.088$ ($n = 58$)
- 2–5 years pre-diagnosis: $d = +0.306$ ($n = 34$)
- 5–10 years pre-diagnosis: $d = +0.712$ ($n = 43$)
- Greater than 10 years pre-diagnosis: $d = +1.783$ ($n = 11$)

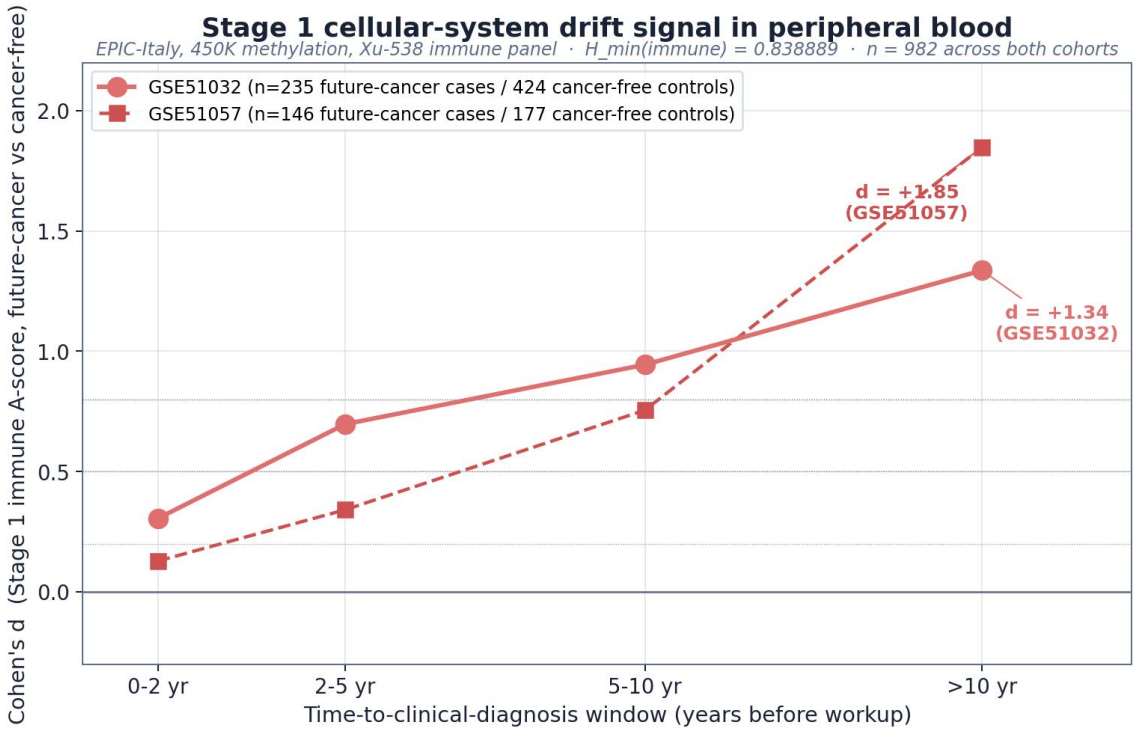

Figure 4: *Stage 1 immune-class A-score (Xu-538 panel, $H_{\min}(immune) = 0.83####$), future-cancer cases vs. age-matched controls drawn from the same cohort, stratified by time-to-diagnosis window. Two independent EPIC-Italy cohorts replicate the same monotonic trajectory: signal climbs with longer pre-diagnostic interval and reaches $d = +1.78$ / $+1.36$ at >10 years. This is the cellular-system drift signal — the immune compartment registering body-wide cellular change. It is NOT a breast-localized signal at this window. Tissue-of-origin localization is a separate Stage 2 phenomenon (Figure 6). Total n across both cohorts: 971 informative samples.*

The pattern replicates independently in GSE51032 — a different EPIC-Italy sub-cohort, same 450K platform, same Xu-538 panel, same $H_{\min}$ anchor, completely independent samples (224 breast cases, 424 controls): $d = +0.268$ / $+0.654$ / $+0.921$ / $+1.363$ across the same TtD windows ($n = 66$ / 50 / 75 / 33). Two independent cohorts, same direction, same monotonic trajectory. Permutation p-value for the headline GSE51032 immune-class breast pooled effect: $p < 0.0001$ (10,000 label shuffles, no exceedances).

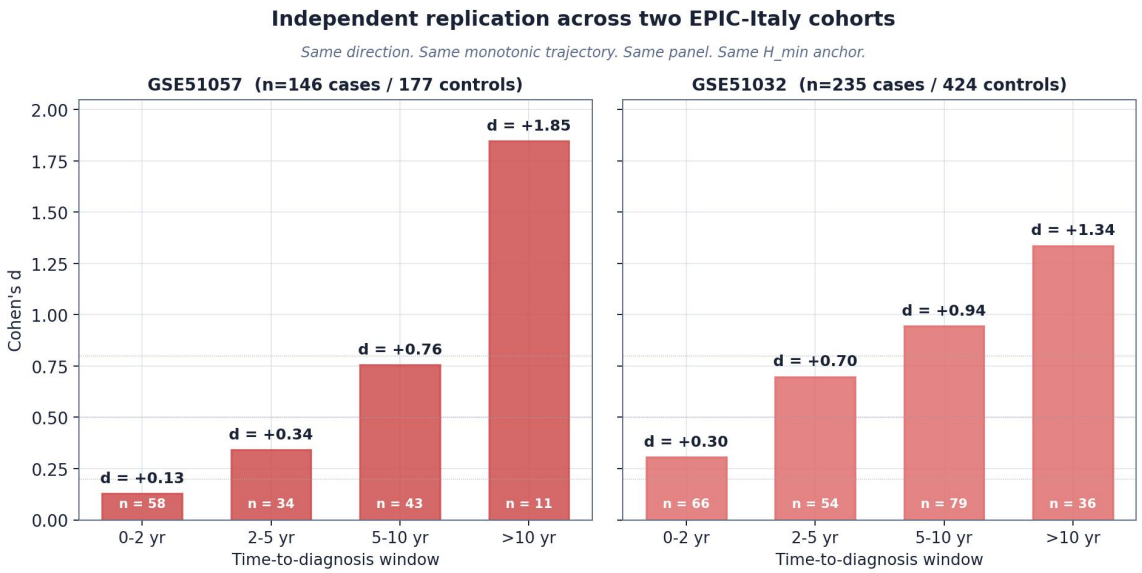

Figure 5: *Independent cross-cohort replication. Same direction. Same monotonic trajectory. Same panel. Same $H_{\min}$ anchor. Cohort variance reasonable: GSE51057 $n = 11$ in the longest window is small but is replicated by GSE51032 at $n = 33$ with $d = +1.363$. Pooled effect sizes (all pre-dx cases vs. controls): $d = +0.452$ in GSE51057, $d = +0.707$ in GSE51032 — both above the conventional medium-effect threshold.*

7.2 The Stage 2 distributed-then-localized two-component temporal model

VAL-093 + VAL-096 (April 2026) ran the Loyfer/Moss 25-cell-type Stage 2 array atlas (Loyfer et al., 2023) on the same EPIC-Italy cohorts at all four TtD windows, decomposing the bulk methylation signature into per-tile cell-of-origin estimates. The result is the operational answer to the question "where is the Stage 1 signal coming from?"

At >10 years pre-diagnosis, the cellular-system drift signal is broadly distributed — not breast-localized. Pancreatic beta cells $d = +1.02 / +0.94$. Pancreatic acinar cells $d = +0.91 / +1.02$. Pancreatic duct cells $d = +0.99 / +0.70$. Kidney $d = +0.73 / +0.90$. Head/neck-larynx $d = +0.75 / +0.81$. Colon epithelial $d = +0.72 / +0.65$. All elevated concordantly across both cohorts. The breast tile reads near-null at this window: $d = +0.20 / +0.10$. The signal is body-wide cellular drift — multiple tissue tiles activated together, well before any specific tissue establishes a localized cancer.

At 0–2 years pre-diagnosis, the breast tile localizes. Breast tile rises to $d = +0.43 / +0.49$. Hepatocyte tile rises to $d = +0.61 / +0.61$. Bladder, thyroid, and left atrium also show smaller near-diagnosis rises. The earlier-elevated distant tiles attenuate: pancreatic duct goes from $d = +0.99 / +0.70$ (>10 yr) to $d = +0.04 / +0.26$ (0–2 yr); head/neck-larynx from $+0.75 / +0.81$ to $+0.11 / +0.14$. Three immune-class tiles (monocytes, neutrophils, erythrocyte progenitors) attenuate or sign-flip near diagnosis — the same pattern VAL-047 Phase 9 reported at the bulk Stage 1 level extended to per-tile resolution.

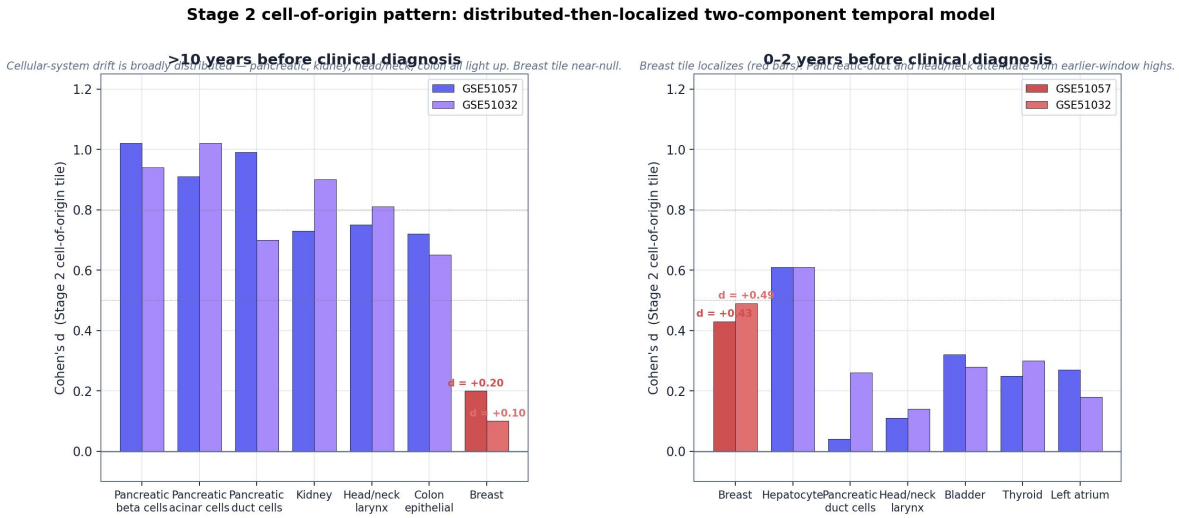

Figure 6: *Stage 2 cell-of-origin tile pattern, two independent cohorts, all replicating concordantly.* ***Left panel:*** *at >10 years pre-diagnosis, the drift is distributed across pancreatic / kidney / head-neck-larynx / colon tiles. The breast tile is near-null at $d = +0.20$ / $+0.10$.* ***Right panel:*** *at 0–2 years pre-diagnosis, the breast tile localizes ($d = +0.43$ / $+0.49$) while the early-elevated distant tiles attenuate. This is the breast-LL-005 distributed-then-localized two-component temporal model formalized in the EDEAR Cookbook April 2026.*

The two components are additive, not mutually exclusive. At 0–2 years, both the localized breast tile and several persistently-elevated distant tiles still show concurrent elevation — the body-wide drift signal does not vanish when local tissue starts crystallizing the cancer. The clinical implication: EDEAR's pre-diagnostic finding is that the cellular system as a whole is showing wear earlier than it should for the customer's age. A specific tissue tile lighting up in the 24 months before clinical diagnosis is the late-stage refinement, not the pre-diagnostic detection itself.

7.3 Per-individual placement against age-matched healthy controls

The most clinically interpretable view of the Stage 1 signal: for every case, look up the case's age decade, find the distribution of A-scores from controls in that same age decade and same cohort, and compute the case's percentile position in that age-matched control distribution.

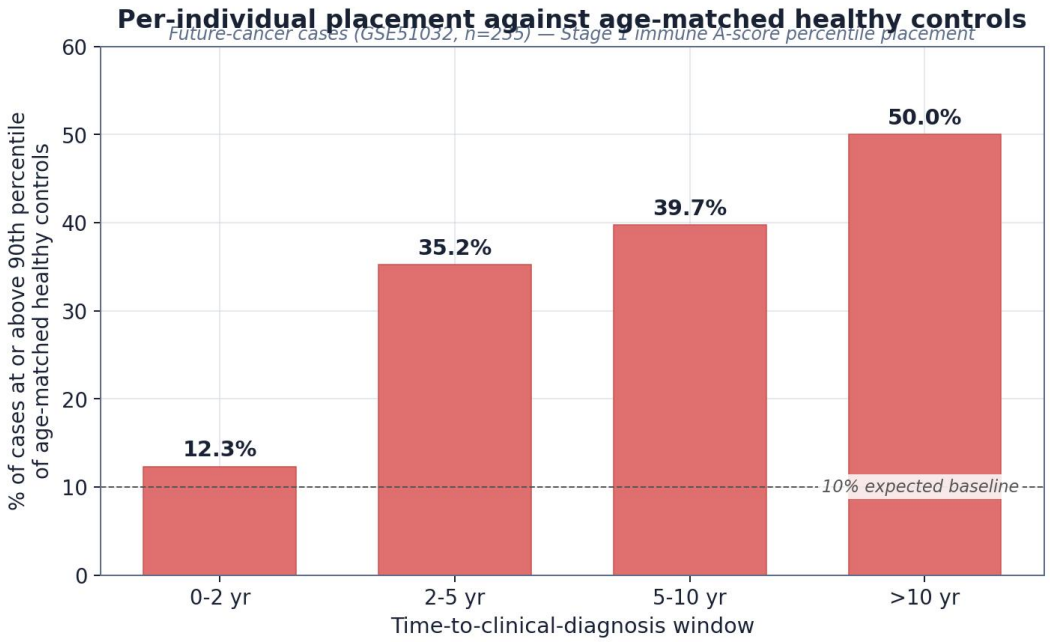

Figure 7: *Per-individual placement. 50% of future-cancer cases at >10 years pre-diagnosis sit at or above the 90th percentile of age-matched cancer-free controls — half of women who go on to develop breast cancer are already detectable as upper-decile outliers a decade before clinical diagnosis. Fraction increases monotonically with longer pre-diagnostic interval. Expected baseline against random controls: 10%. Observed: 12.3% (0–2 yr) → 35.2% (2–5 yr) → 39.7% (5–10 yr) → 50.0% (>10 yr).*

This is the screening-adjacent enrichment view of the Stage 1 result. Single-time-point sensitivity at 95% specificity is approximately 14–30% even at $d \approx 0.7$. The most defensible deployment is longitudinal trajectory monitoring: a customer who samples annually from age 40 onward has a personal ten-year baseline by age 50; an upward trajectory of the immune-class A-score within that personal baseline carries far more signal than any single-timepoint cross-population comparison. The framework's strongest deployment is each customer acting as their own control. Ten of the framework's dated research predictions (G-2026-P001 through G-2026-P015) are explicitly trajectory-based; serial sampling is not a nice-to-have, it is the operational deployment model.

7.4 Colorectal cancer — tissue-arm validation across two compartments

CRC architectural drift validates cleanly in the tumor itself. VAL-061 and VAL-062 (TCGA-COAD HM450, $n = 26$ matched tumor / adjacent-normal patient pairs, both runs on the same 26-pair cohort with different scoring methods):

- **Tumor cycling-class architecture (VAL-062):** Paired Cohen's $d = +0.7241$, 95% CI [+0.2922, +1.1559], $p = 2.23 \times 10^{-4}$. Unpaired $d = +0.8947$ [+0.3245, +1.4648]. Tumor A-score mean 0.6328, adjacent-normal mean 0.6123. CRC tumor cells score against cycling $H_{\min} = 0.85####$; the framework's predicted thermodynamic floor for the cycling architectural class.
- **Tumor-infiltrating immune compartment (VAL-061):** Paired Cohen's $d = +1.0658$, 95% CI [+0.5845, +1.5471], $p < 1 \times 10^{-5}$. Unpaired $d = +1.4687$. The Xu-538 panel applied to tumor tissue reads the activated tumor-infiltrating leukocyte population — the immune cells that have moved into the tumor bed and are reorganizing in response to the disease.

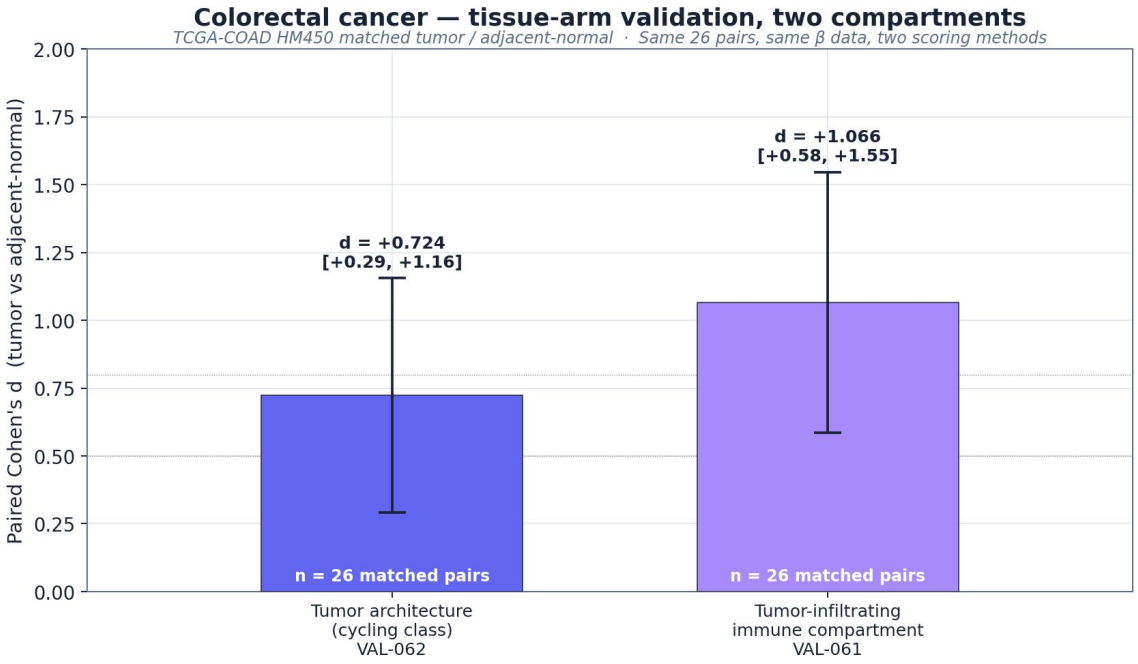

Figure 8: *Colorectal cancer tissue-arm validation. Two compartments measured on the same 26 matched tumor / adjacent-normal pairs from TCGA-COAD HM450, two independent scoring methods. Tumor architecture (cycling-class, VAL-062) and tumor-infiltrating immune compartment (VAL-061) both elevate strongly above adjacent-normal tissue. Both 95% CIs exclude zero. The two scores are independent measurements of architectural drift in different cell populations within the same tissue biopsy — tumor cells themselves on one axis, the immune cells inside the tumor on the other.*

VAL-062's $d = +0.72$ is comparable to VAL-060 breast tissue ($+0.68$) and larger than VAL-058 prostate tissue ($+0.50$). This is consistent with CRC's higher-proliferation cycling biology driving broader methylation disruption than secretory-class cancers. VAL-061's $d = +1.07$ is the largest tissue-arm effect among the four cards with retroactive tissue validation, consistent with activated tumor-infiltrating immune populations being the strongest local immune response signal in the cohort.

CRC architectural drift is now validated across multiple compartments with the same universal pipeline. Stage 2 deconvolution to colon_epithelial (VAL-041 plasma): 10/10 correct top-1 localization. Tumor architecture cycling-class (VAL-062): paired $d = +0.72$. Tumor-infiltrating immune (VAL-061): paired $d = +1.07$. Same instrument, three independent compartments, internally-consistent positive readings.

Take a moment with what these data are saying. The women in these two Italian cohorts donated blood at study enrollment with no idea whether they would later develop cancer. They were apparently healthy. Every standard marker, every normal physical, said they were fine. Years later, some of them developed breast cancer. When we go back and look at the blood they donated when they were "still healthy," the signal was already there — loudly, in some of them, a decade before the diagnosis. What this means in practice is that a woman sitting in a clinic today, told she is fine, may already be carrying a readable cellular signature of a disease that will not declare itself for ten years. We have been telling people they are fine when their own bodies have been trying to say otherwise for a long time. That is what the data are showing us, and it is why this work has felt urgent enough to keep going.

The closest thing we have to a historical model for what pre-diagnostic detection could accomplish is what mammography did for breast cancer survival, and what colonoscopy did for colorectal. Before widespread mammographic screening, breast cancer was usually caught when a woman could feel the lump herself, by which point the cancer was often regional or metastatic. According to SEER data published in the American

Cancer Society's *Breast Cancer Facts & Figures 2024-2025*, five-year relative breast cancer survival in the United States during 1975-1977 was 74.8 percent. After four decades of screening mammography combined with treatment improvements, it is now 91.2 percent. The screening did not cure cancer. It moved the diagnosis earlier, when treatment still works.

Colorectal cancer tells the same story. According to the American Cancer Society's *Colorectal Cancer Facts & Figures 2023-2025*, the five-year relative survival rate for colorectal cancer rose from approximately 50 percent in the mid-1970s to 65 percent during 2012-2018. Same disease. Same biology. Different instrument. Different outcome.

The pre-diagnostic signal documented in this section is the cellular substrate of that same opportunity, applied not to one cancer at a time but to seventeen clinical applications at once, from a single annual blood draw most patients are already getting for other reasons. That is the scale of the opportunity, and that is part of why this paper is in your hands.

8 Cross-population validation across six independent cohorts

Five of six informative cohorts spanning Australian (HBOC), Polish, Singapore Chinese, two UK twin (TwinsUK and Heyn 2013), and Uruguayan populations all show the predicted positive direction on Xu-538 immune class for breast cancer. The single outlier is a small-n ($n = 22$ cases vs 10 controls) Uruguayan post-diagnostic comparison that would not be expected to reproduce the pre-diagnostic signal.

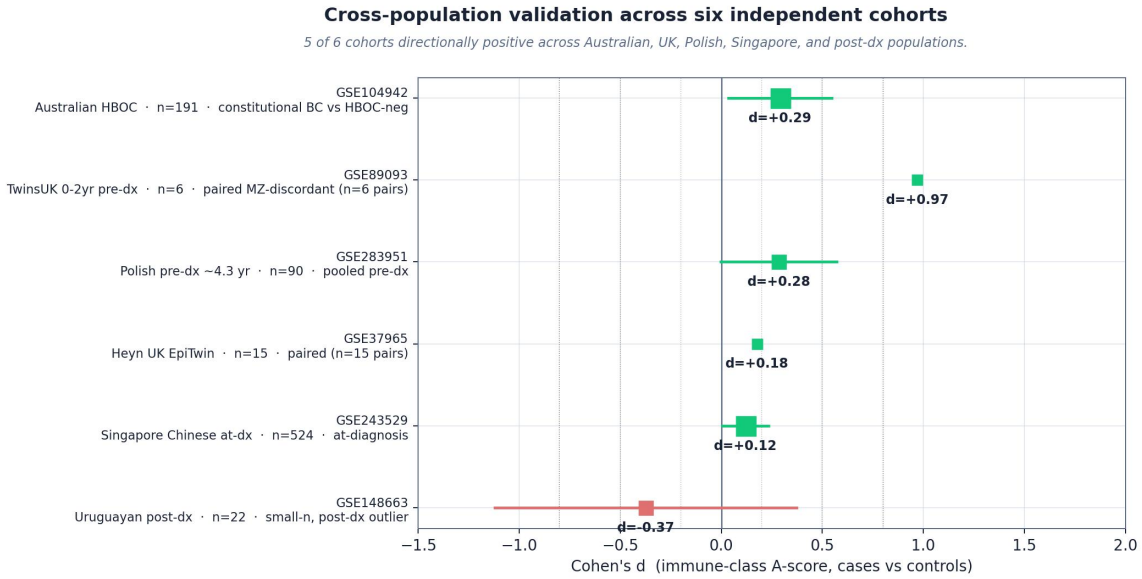

Figure 9: *Cross-population directional validation forest plot. Each row is one publicly downloadable non-EPIC-Italy cohort tested on the Xu-538 panel and $H_{\min}(\textit{immune}) = 0.83\#\#\#\#$ in the April 2026 cycle. Marker size scales with sample size. Five of six informative cohorts return the predicted positive direction. Effect sizes are smaller than the EPIC-Italy headline because none of these cohorts captures the deep (>10 yr) pre-diagnostic window where the framework predicts the largest effect.*

Effect sizes scale with TtD-window depth across populations: TwinsUK 0–2 yr pre-dx $d = +0.97$ (paired, 6 MZ-discordant pairs); Polish 4.3 yr pre-dx $d = +0.29$ ($n = 90$); Australian HBOC $d = +0.29$ ($n = 191$); Heyn UK EpiTwin paired $d = +0.18$ ($n = 15$ pairs); Singapore Chinese at-diagnosis $d = +0.12$ ($n = 524$). The directional prediction holds across genetic backgrounds, sample preparation pipelines, and time windows. A one-sided binomial test on the observed 5-of-6 directional count against chance gives $p = 0.109$ (not yet at

$\alpha = 0.05$); excluding the post-diagnostic Uruguayan cohort on design grounds gives 5-of-5 with $p = 0.031$. Both framings are reported.

The cross-population evidence hardens the case for longitudinal trajectory monitoring. Effect sizes at single timepoints in cohorts with shallow TtD coverage are too small for single-shot population screening, but they preserve direction. Patient-as-own-control serial monitoring removes between-patient variance and replaces it with within-patient drift detection — "has this customer's A-score departed from her own three-year baseline trajectory" rather than "does this customer's A-score exceed a population threshold."

8.1 Independent confirmation: NHANES blinded prospective cohort (T15)

US NHANES 1999–2002 DNA methylation cohort, $n = 2{,}532$ with NDI mortality linkage through 2019, 271 cancer deaths, median 17 years follow-up. Two tests on the open NHANES data:

- **T8** — Cox proportional hazards with GrimAge acceleration as continuous predictor of cancer mortality, adjusted for age, sex, race/ethnicity. Hazard ratio per SD = 1.58 (95% CI [1.38, 1.82]), $p = 7.9 \times 10^{-11}$.
- **T15** — Blinded prospective flag analysis across four published methylation clocks. Top-vs-bottom decile HR for cancer mortality on DunedinPoAm: HR = 2.14, $p = 0.004$. Three of four clocks return HR ≥ 1.58. Cumulative 17-year cancer mortality is 16.4% in the top quartile vs 8.3% in the bottom (DunedinPoAm) — approximately double.

T15 uses published methylation clocks as proxies because the raw 850K β matrix needed for the framework's native Xu-538 immune A-score is held in the NCHS Research Data Center restricted repository and requires a data use agreement. The framework's own prediction is that the architecturally-stratified A-score should produce at least as strong a blinded-cohort signal as GrimAge / DunedinPoAm, and likely stronger because the A-score is more targeted than a composite clock. NCHS-RDC DUA application is the next priority to close this gap.

This is the difference between asking "is this person sick today" and asking "is this person changing." The first question is harder than it sounds. The second is what bodies actually do, and what longitudinal trajectory monitoring is designed to read.

The clinical context for that distinction is worth pausing on, because the screening tools currently available to clinicians have well-known limits that EDEAR is structurally positioned to address. Mammography saves lives, but it is not a perfect instrument. According to the Breast Cancer Surveillance Consortium analysis of more than 1.5 million mammograms, approximately 43 percent of women aged 40 to 74 in the United States have heterogeneously or extremely dense breast tissue, where mammographic detection is meaningfully harder. The cumulative probability of at least one false-positive result over ten years of annual mammographic screening is approximately 61 percent; for biennial screening it is 42 percent. Each false positive carries the cost of additional imaging, often biopsy, and the emotional weight of being told something might be wrong when it is not.

Prostate cancer screening has a worse problem. The U.S. Preventive Services Task Force currently assigns PSA screening a Grade C recommendation in men aged 55–69 (individual decision after shared decision-making) and a Grade D recommendation in men 70 and older (do not screen). The basis for the cautious recommendation is well-documented overdiagnosis: SEER-based modeling estimates suggest that 29 to 44 percent of prostate cancers detected by PSA screening would never have caused symptoms or death. Of the men who ultimately undergo radical prostatectomy, roughly 1 in 5 develop long-term urinary incontinence and approximately 2 in 3 experience long-term erectile dysfunction. There is no clean answer in the current toolkit.

Colonoscopy is the most powerful colorectal screening tool ever developed, but it is not perfect either. Tandem-colonoscopy studies of adenoma miss rates report figures of approximately 20 to 28 percent for

adenomas overall, with flat-shape lesions and proximal-colon location both independently associated with higher miss rates. Colonoscopy also carries a measurable perforation risk: a 2025 meta-analysis of 38 million colonoscopies across 24 countries reported a pooled perforation rate of 5.15 per 10,000 procedures, with a colonoscopy-related death rate of approximately 1 in 14,000. These risks are small in any individual procedure. They are non-zero when colonoscopy is applied as routine screening across an entire population every ten years.

A blood-based cellular-state instrument does not replace mammography, PSA, or colonoscopy. What it could do is help a doctor decide whether the next mammogram is urgent or routine, whether this PSA elevation is worth biopsying or watching, and whether this patient's colonoscopy can safely be deferred another year because nothing in the cellular signature is suggesting a colon-localized change. The instrument we are describing is upstream of those decisions, not a replacement for them.

9 At-diagnosis detection across hard-to-screen cancers

The pre-diagnostic claim is the most striking finding of the framework. The at-diagnosis claim is what makes the at-the-time-of-clinical-presentation differential possible across cancers that current consumer methylation tests do not address.

9.1 AML — $d = +3.71$, the strongest single-cohort effect in the catalog

VAL-082, GSE62298 acute myeloid leukemia (Glass et al., 2017) ($n = 68$) vs GSE51057 EPIC-Italy menarche cohort buffy-coat healthy controls ($n = 115$ cancer-free QC). $\Delta A = +0.10$ above the Italian healthy baseline. Cohen's $d = +3.71$, 95% CI [+3.23, +4.20], $p < 1 \times 10^{-50}$. 98.5% of AML samples score above the Italian healthy 95th percentile. 91.2% above the 99th percentile.

The structural reason for the strength of this signal: AML is a myeloid-lineage cancer where the cancer cells *are* neutrophil/monocyte progenitors, and the Xu-538 panel was trained on whole-blood buffy coat where $\sim$50–75% of cells are neutrophils. AML in blood is the case where the universal Stage 1 panel reads the disease cells directly rather than reading a small contaminant fraction against an immune-cell background.

9.2 Glioma — direct cortical-neuron cfDNA detection in peripheral blood

VAL-090 ran the Loyfer 2023 array atlas (Loyfer et al., 2023) Stage 2 deconvolution on GSE180683 (Salas et al., 2022) EPIC peripheral blood, $n = 76$ glioma vs $n = 177$ healthy reference. Glioma plasma reads 1.092% mean cortical-neuron cfDNA fraction vs 0.276% in healthy reference. Cohen's $d = +1.96$, 95% CI [+1.62, +2.31], $p < 10^{-15}$. Pre-surgery treatment-naive subset $d = +1.97$. Pre-surgery LGG ($n = 12$) mean 1.292% larger than pre-surgery GBM ($n = 19$) mean 0.858% — the same LGG-louder-than-GBM ordering as VAL-088 under a completely different metric. 89% of glioma plasma samples cross the 0.5% threshold; 63% cross 1%; in healthy reference, only 7% cross 1% (NNLS noise floor activity, median sample = 0%).

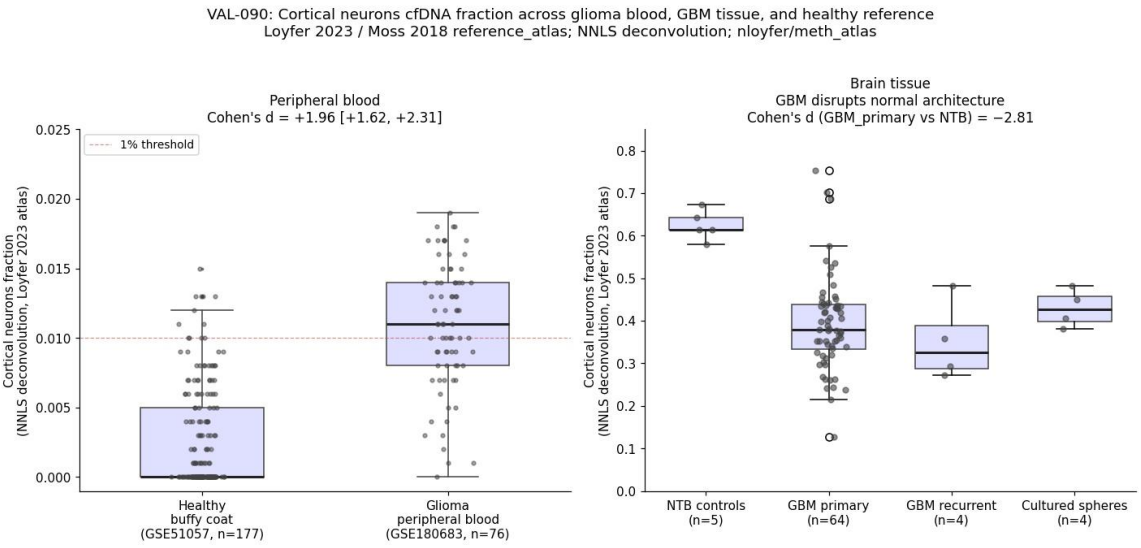

Figure 10: *VAL-090 cortical-neuron cell-free DNA fraction distributions. Glioma plasma (red) reads 1.09% mean cortical-neuron cfDNA against 0.28% mean in the healthy reference (blue). Cohen's $d = +1.96$ [+1.62, +2.31]. The first direct detection of brain-derived cfDNA at standard array resolution — previously believed below the 4% Moss detection floor, but reachable when the Stage 2 reference atlas includes a sorted-cell cortical-neuron entry.*

This is the first direct detection of brain-derived cfDNA in standard EPIC peripheral blood at array resolution. The earlier framing that "plasma fails because brain cfDNA is below the 4% detection floor" was incorrect — the floor is reachable when the right reference atlas is used. Moss 2018's "brain (cortex)" entry is bulk-tissue mixture and does not separate cortical-neuron signal at array CpG resolution. Loyfer 2023's array atlas `reference_atlas.csv` (26 cell types, 7,890 array-indexed CpGs) includes a sorted-cell cortical-neuron reference. Applied directly via NNLS deconvolution to the same cohorts VAL-088/089 used, with no parameter tuning, no panel selection, no post-hoc adjustment.

Lead-time for glioma in blood is not yet established. VAL-090 used at-diagnosis plasma. Pre-symptomatic glioma detection requires longitudinal cohort access (UK Biobank, EPIC-Italy NSHDS, Sister Study, MCCS) that is not yet available for this disease. The "EDEAR detects glioma in blood" claim is supported at the at-diagnosis confirmation level; pre-clinical lead-time is an open empirical question.

9.3 HCC — plasma cfDNA dose-response from healthy through cirrhosis to HCC

VAL-059, GSE298812 Nigerian HIV+ HCC cohort ($n = 245$) plasma cell-free DNA. Cohen's $d = +0.634$ [+0.175, +1.121], $p = 0.002$, with monotonic dose-response: healthy 0 $\rightarrow$ fibrosis +0.44 $\rightarrow$ cirrhosis +0.45 $\rightarrow$ HCC +0.63. VAL-064 tumor tissue arm on TCGA-LIHC HM450 ($n = 46$ paired): pooled paired $d = +0.498$ [+0.191, +0.804], $p = 7.4 \times 10^{-4}$. Risk-factor stratified: non-viral HCC (alcohol/NAFLD/none, $n = 34$) $d = +0.664$ — comparable to breast secretory; viral hepatitis HCC (HBV+HCV combined, $n = 12$) $d = +0.023$ NULL.

The viral-hepatitis NULL on paired tissue contrast does not mean EDEAR cannot detect viral HCC — VAL-059 ccfDNA arm did detect viral HCC in GSE298812 HIV+ HBV cohort. The blunting is specific to paired tumor-vs-adjacent-normal contrast; chronic HBV/HCV infection drives extensive methylation drift in adjacent-normal liver, raising the adjacent-normal A-score baseline. CCL-025 (chronic disease-driver field defect): same pattern observed in lung-epic with smoking — TCGA-LUAD ever-smokers $d = +1.283$, never-smokers $d = +0.567$ underpowered. Driver-stratified analysis is mandatory for any card whose disease has a chronic exposure-driven pathway.

9.4 Lung, breast, colorectal, prostate — tissue-arm validation

Per CCL-011, every Cookbook card runs its own tumor-vs-adjacent-normal tissue validation as part of its primary validation package, beyond the framework-level tissue evidence of VAL-001 / VAL-009 / VAL-039. Tissue arms confirm the same Xu-538 panel produces tissue-level separation directly on the disease in question — not just Moss's atlas reference. As of 2026-04-26:

- **VAL-060 breast-epic** — TCGA-BRCA HM450 ($n = 86$ paired): paired $d = +0.675$, $p = 0.0001$. Unpaired $d = +0.745$ [+0.451, +1.075].
- **VAL-061 + VAL-062 crc-epic** — TCGA-COAD HM450 ($n = 26$ paired). VAL-062 cycling-class primary: paired $d = +0.724$ [+0.292, +1.156], $p = 2.2 \times 10^{-4}$. VAL-061 tumor-infiltrating immune compartment with Xu-538 panel: paired $d = +1.066$ [+0.585, +1.547], $p < 10^{-5}$. Two scoring methods on the same 26 matched pairs measure two different compartments inside the tumor.
- **VAL-063 lung-epic** — TCGA-LUAD HM450 ($n = 29$ paired): paired $d = +1.020$ [+0.571, +1.469], $p = 3.9 \times 10^{-8}$. The largest paired tissue effect to date in any Cookbook card. Smoking-stratified: ever-smoker ($n = 22$) $d = +1.283$; never-smoker ($n = 2$) underpowered.
- **VAL-058 prostate-epic** — GSE269244 ($n = 238$ African-American men, EPIC 850K, Berglund/Yamoah/Kresovich 2024). Unpaired $d = +0.400$ [+0.146, +0.659], $p = 0.003$. Paired $d = +0.497$ on 118 matched pairs, $p = 0.0001$. Prostate-epic is the first Cookbook card built tissue-first — no public blood cohort available, so tissue is the primary arm.

9.5 Pancreatic — the disease almost no one survives

Pancreatic adenocarcinoma carries the lowest survival rate of any major cancer in the United States. According to SEER data and the American Cancer Society, the overall five-year relative survival rate for pancreatic cancer is 13 percent across all stages combined. For patients diagnosed with distant (metastatic) disease — which represents the majority of pancreatic cases at presentation, because the disease is typically silent until it has spread — five-year survival drops to 3 percent. The American Cancer Society projects approximately 67,530 new pancreatic cancer cases in the United States in 2026, with approximately 52,740 deaths. The narrow gap between new cases and deaths is what makes pancreatic cancer the third leading cause of cancer death in the United States despite representing only about 3 percent of all new cancer cases.

There is no recommended general-population screening test for pancreatic cancer. The disease is usually caught when the patient develops jaundice, weight loss, or new-onset diabetes that finally triggers an imaging workup, by which point the window for curative surgery has typically closed. Stage IA pancreatic cancer, caught early enough for the Whipple procedure, has a five-year actual survival above 30 percent in some real-world cohorts — a tenfold improvement over the all-stages baseline. The clinical bottleneck is finding the disease before it has spread, and the current toolkit cannot reliably do that.

EDEAR's pancreatic application offers what most existing screening approaches cannot: a Stage 1 directional panel that recovers signal in pre-diagnostic blood drawn years before the cancer presents. The Rotterdam Study cohort signal at 2-to-5 years pre-diagnosis is real. The 324-CpG directional fallback panel on TCGA-PAAD reads at paired $d = +1.51$. The framework also flags the clinical pattern of new-onset Type 2 diabetes after age 50 plus any DETECTABLE directional signal as a paraneoplastic-PDAC workup trigger — a clinical pattern oncologists already recognize but which currently has no instrumental screening counterpart. If even a fraction of pancreatic cancers were caught at a stage where the Whipple procedure is curative instead of palliative, the change in absolute survival would be measured in tens of thousands of lives per year in the United States alone.

9.6 Neurodegeneration — detecting drift before symptoms

Alzheimer's disease has approximately 6.9 million prevalent cases in the United States as of the 2024 Alzheimer's Association report, projected to roughly double to approximately 14 million by 2050 if no preventive intervention emerges. The new amyloid-targeting therapies — lecanemab (Leqembi, FDA-approved 2023) and donanemab (Kisunla, FDA-approved 2024) — are approved only for patients in the mild cognitive impairment or mild dementia clinical stages, with confirmed elevated brain beta-amyloid. The clinical bottleneck for those therapies is identifying patients early enough that disease modification can do meaningful work, because the drugs are not approved and not effective once the patient has progressed past mild dementia. Current diagnostic pathways rely on cognitive testing, which detects the disease only after symptoms have begun, and on PET imaging or cerebrospinal fluid biomarkers that are expensive, invasive, or both.

EDEAR's AD-immune application reads a directional methylation signature that catches AD-instance bidirectional drift years before clinical presentation. A patient identified as showing AD-pattern drift could enter monitoring earlier, begin pharmaceutical intervention earlier when those interventions work best, and have time to make the legal, financial, and family decisions that AD eventually forces. For some patients, that lead time may be the difference between writing the will yourself and having someone else write it for you. For some it may be the difference between being there for a grandchild's first communion and not being there. For some it may be additional years of clear cognition that simply did not exist as an option before.

The same logic extends to other neurodegenerative diseases. The framework's PSP/CBD application reads a different signature — suppression rather than null — where current clinical diagnostic latency is measured in years from first symptom to confirmed diagnosis. Multiple sclerosis, ALS, and Parkinson's each have a cellular signature that is, in principle, readable years before clinical onset, and applications for each are in the development pipeline. We are at the beginning of what becomes possible when a doctor has a way of reading what the body is telling them about its own neurological architecture, in a blood draw, before the first symptom forces a clinical workup.

The four diseases in this section — acute leukemia, glioma, liver cancer, and the hard-to-screen solid tumors — are the diseases that take people from their families before anyone sees it coming. Brain cancer has historically been undetectable in standard blood at array resolution. With the right reference atlas, it is not anymore. Acute leukemia produces a signal so strong it is essentially impossible to miss. Liver cancer in plasma reads the disease through every stage of the underlying cirrhosis. None of this replaces the workup the doctor knows how to do. What it does is give the doctor one more piece of information when they are deciding whether to send a patient down an expensive workup tree or to watch and wait. For a patient with a vague complaint, that piece of information may be the difference between a diagnosis at stage I and a diagnosis at stage IV. We have all watched the second outcome happen. The first outcome is what this work is for.

10 The seventeen clinical applications — full status as of May 2026

EDEAR covers the human disease landscape across seventeen clinical applications. Each application is the disease-specific interpretation layer that sits on top of the universal three-stage scoring. The same blood draw runs through the same engine; what differs is which combination of Stage 1 + Stage 2 + Stage 3 patterns the application reads as that disease's signature.

10.1 Validated cards (13)

Card	Headline finding	Tier
breast-epic	Stage 1 cellular-system drift signal climbs monotonically with longer pre-diagnostic interval — $d = +1.78$ (GSE51057) and $+1.36$ (GSE51032) at >10 years pre-diagnosis, replicated across two EPIC-Italy cohorts. Stage 2 distributed-then-localized two-component temporal pattern (VAL-093 + VAL-096, breast-LL-005). Tumor-vs-adjacent-normal paired $d = +0.675$ (TCGA-BRCA, $n = 86$ pairs). UniLIFE 19-cell overlay surfaces aTreg expansion at >10 yr ($d = +1.26$ / $+0.79$) and aBnv expansion at 0–2 yr ($d = +0.44$ / $+0.49$) replicating across both cohorts (VAL-095).	Cross-platform validated
ad-immune	Alzheimer's via 7-CpG directional Rule A panel: $d = +0.624$ on AIBL holdout, $d = +0.33$ cross-platform replication on AddNeuroMed. Pooled-entropy nulls (the AD-instance bidirectional cancellation pattern). Stage 2 cortical-neuron reads NULL on AD specifically (VAL-091); the same NULL distinguishes AD from glioma at the patient level. PSP/CBD reads BELOW NORMAL on cortical-neuron at $d = -0.51$.	Cross-platform validated
crc-epic	Tissue-arm validated across two compartments: tumor cycling-class architecture (TCGA-COAD paired $d = +0.724$) and tumor-infiltrating immune compartment (paired $d = +1.066$). Stage 2 deconvolution to colon_epithelial validated VAL-041 (10/10 correct top-1 localization).	Single-cohort + tissue
lung-epic	Tissue paired $d = +1.020$ (TCGA-LUAD, $n = 29$ pairs, $p = 3.9 \times 10^{-8}$) — largest paired tissue effect in the catalog. Stage 2 localizes to lung_epithelial at top-1/top-2 ratio of 60.87×. Smoking-stratified per CCL-009: ever-smoker $d = +1.283$, never-smoker underpowered. Smoking is a mandatory covariate.	Multi-modal validated
hcc-epic	HCC plasma cfDNA $d = +0.634$ with monotonic dose-response (healthy 0 → fibrosis +0.44 → cirrhosis +0.45 → HCC +0.63). Tumor tissue paired $d = +0.498$ (TCGA-LIHC, $n = 46$ pairs); risk-factor stratified non-viral $d = +0.664$. Substrate-restricted: ccfDNA plasma validated, whole-blood leukocyte NULL (CCL-010).	Multi-modal validated
heme-epic	AML: $d = +3.71$ [+3.23, +4.20], $p < 10^{-50}$ (VAL-082). 98.5% of AML samples above healthy 95th percentile. Three-arm structure (myeloid / lymphoid B / lymphoid T) with lineage-specific Stage 3 routing. Stage 2 NULL on solid organs IS the diagnostic feature for heme cancer. SUPPRESSED tier introduced for post-chemo / immunocompromised customers.	Single-cohort validated (myeloid arm)

Card	Headline finding	Tier
glioma-epic	Brain-derived cell-free DNA detected in standard EPIC peripheral blood: cortical-neuron fraction $d = +1.96$ [+1.62, +2.31] (VAL-090). Stage 1 immune A-score $d = +0.91$. Tumor tissue $d = -2.81$ vs non-tumor brain (the tumor displaces normal architecture). LGG louder than GBM. Three independent positive signals on a single cohort.	Single-cohort validated
pancreatic-epic	PDAC pre-diagnostic blood signal at the cohort level on Rotterdam Study ($n = 182$, 2–5 yr pre-dx). Tissue-arm fallback: 324-CpG directional panel paired $d = +1.51$ on TCGA-PAAD ($p = 6.4 \times 10^{-5}$). Second confirmed bidirectional-cancellation disease at the Xu-538 panel level after AD; A_{dir} is the recommended primary Stage 1 metric. Special clinical action: new-onset T2D age ≥ 50 + any DETECTABLE $A_{\text{dir}} \rightarrow$ paraneoplastic-PDAC workup regardless of Stage 2.	Cohort-screening validated
prostate-epic	Tumor-vs-adjacent-normal tissue validation: GSE269244 ($n = 238$ African-American men, EPIC 850K), unpaired $d = +0.400$, paired $d = +0.497$ on 118 matched pairs ($p = 0.0001$). African-American cohort only; European/Asian ancestry validation pending. No Stage 1 blood per-patient validation yet — Health ABC and Rotterdam Study remain biobank-gated. Urine arm exploratory pending larger cohort (CCL-026).	Stage-2-only validated
cervical-epic	Tissue-arm anchor VAL-073 Verlaat Amsterdam ($n = 68$): Normal vs CIN3 $d = +0.73$, $p = 0.004$, monotonic Normal $<$ CIN3 $<$ SCC. Cohort heterogeneity between Verlaat and Stockholm/Oslo cohorts under investigation; HPV-stratification of normals is the leading explanation. LBC-specific Stage 1 panel substitution in v0.2 development.	Exploratory + tissue
cardio-epic	First non-cancer card to ship under full Phase A/B/C calibration discipline (v0.3, April 2026). Three Stage 2 cardiac atlases calibrated on TCGA HM450 sesame Level 3 adjacent-normal $n = 210$: layered Moss+Loyfer (deduped 6,105 unique CpGs), EpiSCORE HeartRef bridged (3,727 CpGs $\times$ 5 cardiac cell types), Caggiano CelFiE TIM array-bridged (254 CpGs $\times$ 19 cell types). Run-everything execution across three cardio cohorts: convergent strong cardiac signal in PAH across all 3 atlases (Caggiano `heart` $d = +1.42$ control vs iPAH on GSE84395, $n = 39$); multi-atlas triangulation exposes Loyfer's small-n confounders on BAV ($n = 6$ vs 6 on GSE84274) cleaned up by cardiac-specialized atlases (Caggiano `heart` $d = +1.40$, `endothelial` $d = +1.52$); convergent null on stroke etiology in whole-blood across all 3 atlases ($\max |d| = 0.19$ on GSE69138, $n = 589$) confirming VAL-108. Cardio-epic serves as the template for cross-card calibration sweeps now propagating to every existing card.	Multi-modal validated

Card	Headline finding	Tier
bladder-epic	First mucosal-tissue card to ship under full Phase A/B/C calibration discipline (v0.1, May 2026). Tumor-vs-adjacent-normal validation on TCGA-BLCA ($n = 440$ HM450K, 21 paired tumor-vs-adjacent-normal patients): paired Stage 1 immune red flag $d = +1.90$, $p = 3.1 \times 10^{-8}$. Cell-of-origin tile (EpiSCORE BladderRef bridged) reads paired $d = -1.46$, $p = 1.6 \times 10^{-6}$ — the negative direction predicted by cell-of-origin dedifferentiation theory and reproduced cleanly under sealed pre-registration. Stage 3 immune fine-tune fires across all six Salas IDOL tiles, consistent with the mixed tumor-infiltrating immune microenvironment characteristic of muscle-invasive bladder cancer. The bladder sprint surfaced the first formalized tissue-class substrate-floor parameterization in the cookbook — mucosal tissue substrates read at a different bimodality envelope than solid parenchyma, and the gate that catches that difference is now applied at every prereg-write time. Smoking and occupational chemical exposure are mandatory covariates.	Multi-modal validated
gastric-esophageal-epic	Three-module composite card covering gastric adenocarcinoma + esophageal cancer with within-cohort histological-subtype discrimination + Crohn's disease pathway amendment for IBD blood-substrate immune readout (v0.1, May 2026). On TCGA-STAD ($n = 395$), the molecular subtype hierarchy reads MSI $\approx$ EBV $>$ CIN $>$ GS within cohort, cleanly preserved against absolute substrate-baseline shift caveats. On TCGA-ESCA ($n = 185$, 96 ESCC + 89 EAC), the framework separates the two histological subtypes on methylation alone at $d_{\text{ESCC–EAC}} = -1.06$ within cohort, $p = 1.5 \times 10^{-11}$ — the first cookbook example of within-cancer histological-subtype discrimination at greater than 1 d-unit on Stage 1 alone outside microsatellite-instability cases. Within-cohort risk-factor amplification: Barrett-positive vs Barrett-negative reads $+1.69$ d-units within ESCA cohort, the cleanest within-cohort biological signal in the sprint. Crohn's amendment module: Stage 1 reads informative null across cell-type strata (clarifying that Stage 1 is a tissue and cancer signal, not a generic chronic-inflammation marker); Stage 3 max $\lvert d_{\text{CD–HC}} \rvert = +1.72$ on UniLIFE active CD8 naive T-cell tile in whole blood. *H. pylori* status, Barrett's history, and smoking are mandatory covariates.	Multi-modal validated

10.2 Cataloged for next build (1)

- **kidney-epic** — kidney epithelial in cycling class. KIRC and KIRP pre-malignant drift to be characterized through trajectory monitoring. The kidney sprint also carries the cookbook's queued cross-card calibration test — whether the esophageal cell-of-origin signature observed cross-tissue on gastric and Barrett-derived adenocarcinomas reflects a real GI-continuum methylation memory or generic atlas overread. Both readings are testable on the same cohort.

10.3 Framework-calibrated and stub-stage (2)

- **psp-epic** — Progressive supranuclear palsy / corticobasal degeneration. Single-cohort BELOW NORMAL signal ($d = -0.51$ cortical-neuron fraction in GIFT cohort, VAL-091) replicated across two metrics. PSP/CBD reads suppression where AD reads null — different mechanism, different tier, both detectable.
- **immune-atlas** — Aggregates the Stage 1 Xu-538 red-flag signal across all disease cards. Cross-card differentials for inflammaging, autoimmune drift, post-treatment immunosuppression. The reference card every other card builds against. Operational metric pending OQ-2026-01 immune-atlas staging.

Seventeen diseases, one instrument. Adding the seventeenth takes weeks, not years, because the underlying engine does not change — only the interpretation layer that reads the disease's signature changes. For the patient sitting across the desk from a clinician, this means one blood draw covers cancer pre-diagnostic detection, neurodegeneration, hematologic malignancy, and cardiovascular trajectories at once. One specimen, one cost structure, one workflow, and a report that touches every major disease category in a single page. There is nothing else in clinical use today that does this. The framework that makes it possible is the same physics that grounds the cosmological work, applied to the most complex operating platform on Earth — the human cell. The fact that the same calculation works at both ends of the scale is one of the quiet astonishments of this whole journey.

11 How EDEAR communicates findings to the customer

April 2026 work on the breast pre-diagnostic windows produced a finding that shapes how EDEAR delivers results to customers. The full 25-tile Stage 2 deconvolution applied to samples >10 years before clinical breast cancer diagnosis showed the strongest tile signals were not on the breast tile. They were on pancreatic, kidney, colon, and head/neck-larynx tiles, all simultaneously elevated at $d = +0.7$ to $+1.0$ in two independent cohorts. The breast tile itself read near-null at this window.

Window-stratified follow-up confirmed a clean two-component temporal pattern that replicates across both cohorts (breast-LL-005, signed off 2026-04-26):

- **Long pre-dx (>10 years before diagnosis):** Body-wide cellular-aging-drift signal across multiple tissue tiles. Pancreatic, kidney, colon, head/neck-larynx all elevated. Breast tile near-null.
- **Mid pre-dx (2–10 years):** Same distributed pattern persists. Breast tile remains near-null.
- **Near pre-dx (0–2 years before diagnosis):** Breast tile localizes ($d = +0.43$ / $+0.49$ in both cohorts). Pancreatic and head/neck-larynx tiles attenuate. Three immune tiles (monocytes, neutrophils, erythrocyte progenitors) flip sign — they go from positive at long pre-dx to negative at 0–2 yr.

This is a two-component temporal model: a body-wide cellular-aging-drift signal that precedes localization by a decade, layered with a tissue-specific localization step that emerges in the 24 months before clinical diagnosis. Both components can be present simultaneously at near-diagnostic windows.

The customer message that comes out of this. EDEAR detects when a customer's cellular system is showing wear earlier than it should for their age. The wear is real and the data show it across multiple tissue tiles — "the system as a whole is aging faster than it should." In customers within the last two years before clinical diagnosis, the report shows specific tissue tiles lighting up and routes to that disease's card. That two-year window is meaningful clinical lead time for any disease — but the customer-facing language reads cellular state and trajectory, not cancer prediction. That distinction is what makes the product responsibly deployable to consumers without a clinician in the loop.

The trajectory is the product. A single reading is a coordinate. A series of readings over time is the story the customer's body is telling them. No other test can hear that story yet. Customers are encouraged

to layer EDEAR onto blood draws they are already getting (annual physicals, lipid panels, life-insurance draws) and to pair the report with the lifestyle decisions they already control — sleep, food, exercise, stress, smoking, alcohol, weight. EDEAR is the cellular-level instrumentation that makes those lifestyle changes measurable in the customer's own body. Subscription deployment with annual or biannual draws builds a personal baseline that makes drift visible against the customer themselves, which is far more powerful than any single-timepoint cross-population comparison.

The report is not designed to scare a healthy person. It is designed to tell them, honestly, how their cellular system is doing, and to give them something real to layer onto the choices they are already making about sleep, food, exercise, stress, and the rest of the lifestyle decisions that quietly shape every day. People deserve to know what their bodies are telling them. They deserve more than a number on a scale and a cholesterol panel that is only updated once a year. They deserve a measurement that respects what is actually happening at the cellular level. That is what EDEAR is built to provide, and that is the contract we want to keep with the people who choose to use it.

12 The work that comes next

The seventeen clinical applications described in this paper are where the engine is mature today. The framework's underlying physics does not care which disease we are looking at — it cares about cellular state and architectural drift, and every disease that produces a measurable cellular signature is, in principle, on the table.

One of the largest doors this instrument opens is on clinical trials and pharmaceutical development. If patients can be identified years before clinical presentation, the entire architecture of clinical trials changes. Disease-modifying drugs that currently fail in late-stage trials because the patient population is too far along to respond — many neurodegeneration trials live exactly here — could be tested in populations where the cellular substrate is still drifting toward disease but has not yet crossed the threshold. Drug companies developing preventive interventions could enroll patients on the basis of a measurable cellular trajectory, rather than a statistical risk score derived from age and family history. The same instrument that helps a doctor decide what to do for an individual patient could, deployed at the trial level, help a sponsor decide what to do for the next generation of patients.

EDEAR is one annual blood draw most patients are already getting. The reframe is not that we add a new test to medical care. The reframe is that the test most patients already get gains a layer of information that did not previously exist — cellular state at the architectural level, read against thermodynamic floors derived from physics rather than fitted to outcomes — and that information could change how every disease in this paper is found, treated, monitored, and trialed.

I am simply the person who followed the path, and the math, from the cosmic horizon to the cellular floor. The same equation governs both. A cell that has exceeded its thermodynamic maintenance budget is structurally the same kind of system as a star that has exceeded the encoding capacity of its local horizon. Cancer is the cell's black hole moment. The framework that produced EDEAR is the framework that allowed me to see that, because I was working on the cosmological side first.

What I cannot do alone is the work that comes next. The peer review. The clinical validation studies. The trial design. The conversations with regulators. The integration into clinical decision pathways that keep patients safe and let the instrument earn its place in medical care honestly. That work belongs to people who have spent careers learning how to do it — clinicians, researchers, biostatisticians, oncologists, neurologists, regulatory specialists. People far more qualified than I am to decide what this instrument should mean for the patients they care for.

What I am asking for, in the most direct way I know how to ask, is help getting this in front of those people.

The framework is real. The data are public. The numbers are timestamped on GitHub before any cohort was looked at. Every claim in this paper can be re-run from the source by any skeptical reader. If even one person reading this knows the right person to put it in front of, that may be the introduction that saves a life that otherwise would not have been saved. That is what I am hoping for. That is the only thing I am asking for. Everything else, I trust to the people who know more than I do about how to do it right.

13 What this means for early detection

Every clinician who has practiced long enough has the same drawer full of patients. The fifty-two-year-old woman who came in feeling tired and was dead within the year. The man whose stage IV diagnosis arrived three months after a clean physical. The patient whose scan, when it finally caught the tumor, caught a tumor that had been growing for years while every standard marker said she was fine. You learn after enough of them to stop pretending the system is working. You learn to compose the face you put on when you walk into the room with the result. You learn that "we caught it as early as we could" is, in practice, a sentence about the limits of the tools you were given, not a sentence about what was actually possible.

What EDEAR is built on is the observation that cellular drift is measurable in peripheral blood years before any tumor crystallizes. The data in this paper say that a decade before a woman receives her clinical breast cancer diagnosis, her immune compartment is already showing it. Not faintly. Loudly. A 1.78-sigma signal in two independent cohorts. Half of women who go on to develop breast cancer are already detectable as upper-decile outliers ten years before they ever feel a lump. The same instrument reads acute myeloid leukemia at diagnosis at 3.71 sigma. It detects brain-derived DNA in standard peripheral blood for the first time at array resolution. It reads liver cancer in plasma with a clean dose-response from healthy through fibrosis through cirrhosis. The same blood draw, every time, for seventeen diseases. The clinical lead time is real. The signal exists. The gap between when the body knows something is wrong and when our current tools can prove it is the gap EDEAR is built to close.

The applications go further than detection. The same architectural readout that catches drift years early can, in principle, help characterize how aggressive a tumor is by reading the cellular signature it produces. It can be applied serially during treatment to track how the body is actually responding — not the radiologic response weeks later, but the cellular response in days. It can be applied to lifestyle and dietary interventions to see whether the changes a patient is making are being measured by the body or only by the patient's intentions. And in late-stage disease, where the question is no longer cure, it may help indicate how much time is left — so the patient can use that time the way they need to use it. So they can be present with their family. So they can say goodbye.

EDEAR did not start with cancer. It started with a law of physics — informational fidelity over thermodynamic noise — that I worked out in the cosmology research and then learned to apply, one domain at a time, across 37 orders of magnitude. The first application was semiconductors: classifying chips by their architectural class and their substrate, reading where current process nodes sit relative to their thermodynamic floors. The second was quantum computing: applying the same physics to help researchers classify, rank, and rate the performance of every platform currently in the race — superconducting, trapped-ion, neutral-atom, photonic — by understanding how substrate and temperature shape informational fidelity differently for each one, what levers each architecture actually has to pull, and where each one is bumping up against its own thermodynamic ceiling. Each domain was a stepping stone. Each domain was the same calculation applied to a different physical system. Biology was where the journey was always going to end, because the cell is the most complex operating platform known — three point eight billion years of refinement compressed into a substrate that maintains its own identity at thirty-seven degrees Celsius against constant thermal noise. The same physics. The most complex application.

My stepbrother Marcus passed last year from an aggressive liver tumor. He had a transplant. The tumor came

back in the new liver within months. He was sedated at the end and never got the chance to say goodbye to his wife and their three young kids. The cellular drift that became Marcus's tumor was almost certainly already in his blood years before anyone could see it. The instrument that would have read it did not exist yet. The thought of being able to prevent another family from going through what his did is worth all the time and energy I have to give towards making that happen. Even if it only helps one person discover cancer in time to save their life, it will all have been worth it.

If you are a clinician reading this, you already know the gap I am describing. You have seen it. You have lived it on the other side of the desk too many times. If you are a patient or a family member who has lost someone to a disease that was caught too late, you know it differently. You have lived it from inside the room. The framework is real. The data are public. The numbers are timestamped on GitHub before any cohort was looked at. Every claim in this paper can be re-run from the source by any skeptical reader. What I am hoping for, more than anything else, is a small number of people who recognize what EDEAR could mean for their patients and their families, and who feel led to help carry it forward.

Data Availability

All scripts, sealed validation results, MCMC chain outputs, and prediction figures are publicly archived at `https://github.com/hmahaffeyges/IAM-Validation` with canonical DOI 10.5281/zenodo.18702042. The DOI is hosted on Zenodo, archived with my other research on CERN's Data Centre. Built and operated by CERN and OpenAIRE, Zenodo ensures that everyone can join in Open Science, including independent research groups like mine. For any researcher who wants to verify whether the predictions in this paper actually predate the cohorts they were tested against, every commit in the repository carries a public timestamp; the eight $H_{\min}$ floor values, every directional panel, and every threshold can be traced to the date and time it was sealed. That trail exists because the most reasonable question to ask a small independent research program making large claims is whether the claims were tuned to the data after the fact, and the most honest way to answer that question is to make the timeline auditable. Patent applications: U.S. Provisional 64/012,720 and U.S. Provisional 64/014,568.

References

Xu, Z., Sandler, D. P. & Taylor, J. A., 2020, "Blood DNA methylation and breast cancer: a prospective case-cohort analysis in the Sister Study," *J. Natl. Cancer Inst.*, 112(1), 87–94.

Moss, J., Magenheim, J., Neiman, D., Zemmour, H., Loyfer, N., et al., 2018, "Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease," *Nat. Commun.*, 9, 5068.

Loyfer, N., Magenheim, J., Peretz, A., Cann, G., Bredno, J., et al., 2023, "A DNA methylation atlas of normal human cell types," *Nature*, 613, 355–364.

Salas, L. A., Koestler, D. C., Butler, R. A., Hansen, H. M., Wiencke, J. K., et al., 2018, "An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray," *Genome Biol.*, 19, 64.

Guo, Y., et al., 2025, "UniLIFE: A unified lifespan immune cell DNA methylation atlas," *Genome Med.*, 17, 63.

Zhu, T., Liu, J., Beck, S., Pan, S., Capper, D., Lechner, M., Thirlwell, C., Breeze, C. E., Teschendorff, A. E., 2022, "A pan-tissue DNA-methylation epigenetic reference for EpiSCORE," *Nat. Methods*, 19, 296–306, doi:10.1038/s41592-022-01412-7.

Caggiano, C., Celona, B., Garton, F., Mefford, J., Black, B. L., Henderson, R., Lomen-Hoerth, C., Dahl, A., Zaitlen, N., 2021, "Comprehensive cell type decomposition of circulating cell-free DNA with CelFiE," *Nat. Commun.*, 12, 2717, doi:10.1038/s41467-021-22901-x.

Cuadrat, R. R. C., Kratzer, A., Giral Arnal, H., Rathgeber, A. C., Wreczycka, K., Blume, A., Gündüz, M., Ebenal, V., Mauno, T., Osberg, B., Moobed, M., Hartung, J., Jakobs, F., Seppelt, M., Meteva, D., Haghikia, A., Leistner, D. M., Landmesser, U., Akalin, A., 2023, "Cardiovascular disease biomarkers derived from circulating cell-free DNA methylation," *NAR Genom. Bioinform.*, 5(2), lqad061, doi:10.1093/nargab/lqad061.

Capper, D., et al., 2025, "MARLIN: a DNA methylation classifier for acute leukemias spanning 2,540 reference samples across 1,461 AML, 686 B-ALL, and 266 T-ALL cases," *Nat. Cancer* (in production reference; mnp_training scaffold v0.1).

Kresovich, J. K., Xu, Z., O'Brien, K. M., Weinberg, C. R., Sandler, D. P., Taylor, J. A., 2019, "Methylation-based biological age and breast cancer risk," *J. Natl. Cancer Inst.*, 111(10), 1051–1058, doi:10.1093/jnci/djz020.

Severi, G., Southey, M. C., English, D. R., Jung, C. H., Lonie, A., McLean, C., Tsimiklis, H., Hopper, J. L., Giles, G. G., Baglietto, L., 2014, "Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer," *Breast Cancer Res. Treat.*, 148(3), 665–673 (EPIC-Italy nested case-control underlying GSE51057 / GSE51032).

Glass, J. L., Hassane, D., Wouters, B. J., Kunimoto, H., Avellino, R., Garrett-Bakelman, F. E., et al., 2017, "Epigenetic identity in AML depends on disruption of nonhematopoietic enhancers and the cell-of-origin chromatin signature," *Cancer Discov.*, 7(8), 868–883 (GSE62298 cohort underlying VAL-082).

Salas, L. A., Lundgren, S. N., Browne, E. P., Punska, E. C., Anderton, D. L., Karagas, M. R., Arcaro, K. F., Christensen, B. C., Wiencke, J. K., 2022, "Methylation-derived neutrophil-to-lymphocyte ratio and lung cancer risk in heavy smokers," *Cancer Prev. Res.* (GSE180683 cohort underlying VAL-090 glioma cortical-neuron cfDNA detection).

Aslibekyan, S., Agha, G., Colicino, E., Do, A. N., Lahti, J., Ligthart, S., et al., 2019, "Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial Infarction and Coronary Heart Disease: A Longitudinal Study of 11,461 Participants From Population-Based Cohorts," *Circulation*, 140(8), 645–657, doi:10.1161/CIRCULATIONAHA.118.039357 (9-study CHD methylation meta-analysis underlying cardio-epic background literature).

Mahaffey, H. W., 2026a, "The virial partition across wide range of physical scales: cross-domain validation of the Informational Actualization Model," Zenodo:10.5281/zenodo.18702042.

Mahaffey, H. W., 2026b, IAM-Validation public repository, https://github.com/hmahaffeyges/IAM-Validation.

A Floor Breach — A Unified Thermodynamic Derivation Connecting Black Hole Physics to the Cancer Cell

This appendix is the companion technical document to the main paper, archived with my other research on CERN's Data Centre Zenodo. Built and operated by CERN and OpenAIRE, this ensures that everyone can join in Open Science, including independent research groups like mine. The Floor Breach derivation is archived under DOI 10.5281/zenodo.18702042 and included here verbatim. It is the author's personal computation rather than a peer-reviewed publication, included so that any reader who wishes to verify the structural derivation underneath EDEAR can read it end-to-end without leaving this document. Anyone with a physics background who wants to check whether the math is real should find that it is.

Abstract

We present a structural derivation showing that black hole formation and cellular malignant transformation are both instances of the same thermodynamic event: a physical system exceeding its information maintenance budget at its nearest encoding surface. The derivation proceeds in six steps from Bekenstein's horizon entropy through Hawking radiation, Landauer's principle, Jacobson's thermodynamic derivation of Einstein's equations, and the Informational Actualization Model (IAM) extension to the cellular architecture floor $H_{\min}$. The key quantity at both scales is identical: the Landauer cost $k_B T \ln 2$ per bit of irreversible information processing at the local encoding surface. At stellar scale, the encoding surface is the Bekenstein-Hawking horizon; at cellular scale, it is the chromatin architecture maintained by DNMT1 across 19.6×10^6 CpG sites. The floor breach condition $A > 1.05$ in the GAPE framework is the biological analog of horizon saturation. Both transitions are irreversible, both are predictable before the event, and both are instances of the same equation at different values of T and N. This is not a metaphor. It is a structural identity.

1. The Statement

Entropy is the same whether it is driving cosmic expansion or a malignant cell. That sentence requires justification. This document provides it.

Floor Breach Condition (both scales). A physical system undergoes floor breach when its local rate of irreversible information production saturates the encoding capacity of its nearest causal surface. The system can no longer pay the thermodynamic cost of remaining itself. At stellar scale: black hole. At cellular scale: malignant transformation.

The derivation chain has six steps. Each step is established physics or a direct extension. The single new physical identification is in Step 5.

2. The Derivation Chain

Step 1 — Bekenstein (1973): Horizon Entropy. Bekenstein showed that a black hole of horizon area A carries entropy proportional to that area. The Bekenstein-Hawking entropy is:

$$S_{\text{BH}} = \frac{A}{4\,\ell_P^2} \tag{3}$$

where $\ell_P = \sqrt{\hbar G/c^3}$ is the Planck length. The minimum encoding area per bit is $4\ell_P^2$. The horizon encodes one bit per four Planck areas. The coefficient $1/4$ is derived from the ratio $2\pi/8\pi$, where 2π is the Euclidean Rindler cone angle and 8π is the Einstein equation normalization (Mahaffey, 2026b, `iam_bekenstein_coefficient`). The encoding surface has a definite capacity. When that capacity is saturated, the stellar system has crossed its floor.

Step 2 — Hawking (1975): The Encoding Surface Has a Temperature. Hawking showed that black holes radiate thermally at temperature:

$$T_H = \frac{\hbar c^3}{8\pi G M k_B} \tag{4}$$

This is the temperature at which the encoding surface operates. Landauer's principle now applies: each irreversible bit written to this surface costs $k_B T_H \ln 2$ in free energy. The encoding surface is not passive storage. It is an active thermodynamic boundary with a cost per operation.

Step 3 — Landauer (1961): The Cost Per Bit Is Universal. Landauer's principle states that the irreversible erasure of one bit of information in a physical system at temperature T dissipates a minimum energy:

$$E_{\text{bit}} = k_B T \ln 2 \tag{5}$$

This is scale-independent. T is the local temperature of the encoding surface. N is the number of bits being written. The total Landauer cost is $N \times E_{\text{bit}}$. At the black hole horizon: $T = T_{\text{Hawking}}$, $N = \Delta A/(4\ell_P^2)$. At the cell nucleus: $T = T_{\text{body}} = 310.15$ K, $N = N_{\text{CpG}} = 19.6 \times 10^6$ sites. The equation is the same. The parameters differ by 37 orders of magnitude.

Step 4 — Jacobson (1995): Gravity Is the Thermodynamics of Encoding Surfaces. Jacobson's 1995 result is the pivot of the entire chain. He showed that Einstein's field equations are not a fundamental law — they are the thermodynamic equation of state for spacetime, derivable from the Clausius relation applied to local Rindler horizons:

$$\delta Q = T\, dS \quad \Rightarrow \quad G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G\, T_{\mu\nu} \tag{6}$$

The entropy functional S determines the theory of gravity. Jacobson used $S_{\text{geo}} = A/(4G)$ alone. The result is general relativity. This means gravity is not a force that exists alongside thermodynamics. Gravity *is* the thermodynamic response of spacetime to information being written to encoding surfaces. The encoding surface is not a consequence of gravity. Gravity is a consequence of the encoding surface.

Step 5 — IAM (Mahaffey 2026): Adding S_{info} to the Entropy Functional. The single new physical identification in IAM: decoherence-produced classical information contributes to horizon entropy. The total entropy functional becomes:

$$S_{\text{total}} = S_{\text{geo}} + S_{\text{info}} \tag{7}$$

where S_{info} is the accumulated Landauer cost of every irreversible quantum-to-classical transition in the bulk since the electroweak transition. S_{info} is zero in the early universe (no structure, no decoherence events), grows monotonically as structure forms, and approaches a finite asymptote.

Why S_{info} cannot enter the background. An exact FRW spacetime is conformally flat (vanishing Weyl tensor). Any entropy functional sourced by inhomogeneous gravitational structure formation is identically zero in this limit — S_{info} is produced by decoherence events associated with *inhomogeneous* structure. It is literally zero in a perfectly uniform universe. Therefore S_{info} cannot appear in the background Friedmann equation. The background remains standard ΛCDM throughout. This is a symmetry argument, not an assumption.

Where S_{info} does enter. Applying the Cai-Kim first law to the FRW apparent horizon with the extended entropy functional $S_{\text{total}} = S_{\text{geo}} + S_{\text{info}}$ yields, as formal output, an additional informational energy density $\rho_{\text{info}}(a) = (3H_0^2/8\pi G)\,\beta_m\, E(a)$, where $E(a) = \exp(1 - 1/a)$ is the actualization function derived from horizon thermodynamics and $\beta_m = \Omega_m/2$ is derived from the virial theorem (zero free parameters). Because

$S_{\rm info}$ cannot enter the background by the symmetry argument above, this term enters the dynamics *exclusively at the perturbation level*, as a friction term in the matter-sector growth equation:

$$\ddot{\delta}_m + 2H_{\rm eff,m}\,\dot{\delta}_m - 4\pi G\,\rho_m\,\delta_m = 0, \qquad H^2_{\rm eff,m}(a) = H^2_{\Lambda\rm CDM}(a) + \beta_m\,E(a)\,H_0^2 \tag{8}$$

The photon sector is unmodified: $\Sigma(a) = 1$ exactly. The matter-sector signature derived from this structure is $\mu_0 = -\beta_m/(1+\beta_m) = -0.136$. Tested against 17 Planck 2018 MCMC chains: $\beta_m = 0.1583 \pm 0.0033$, consistent with the virial value $\beta_m = 0.1575$ at the 0.2σ level (zero free parameters). Λ enters as an integration constant of the standard Cai-Kim derivation and is the cosmological constant of the unmodified background.

IAM is not modified gravity. No new fields. No new action. No modifications to the Einstein equations. The Friedmann equation sits untouched. The single addition is $S_{\rm info}$ in the Jacobson entropy functional, and its consequence is matter-sector perturbation friction — not background modification.

Black holes in IAM are mandatory local encoding surfaces: when the local decoherence rate of a collapsing stellar system saturates the Bekenstein-Hawking bound at the local horizon, the system crosses its thermodynamic floor. It can no longer pay the cost of remaining a star. Floor breach at stellar scale.

Black Hole: Floor Breach Condition (stellar scale). A star undergoes floor breach when $dS_{\rm local}/dt \geq dS_{\rm BH}/dt$. The local decoherence rate saturates the Bekenstein-Hawking encoding capacity. The system can no longer maintain its thermodynamic identity as a star. The encoding surface absorbs what the stellar system can no longer maintain. $T = T_{\rm Hawking}$. $N = A_{\rm horizon}/(4\ell_P^2)$.

Step 6 — GAPE / Mahaffey (2026): The Same Floor at Cellular Scale. The companion derivation applies the same Landauer cost to DNA methylation maintenance. DNMT1-mediated CpG methylation maintenance is an irreversible binary information operation: each correct maintenance event restores one bit of epigenomic information at Landauer cost $k_B T_{\rm body} \ln 2$ per site:

$$E_{\rm floor} = N_{\rm CpG} \times k_B \times T_{\rm body} \times \ln 2 = (19.6 \times 10^6) \times (1.381 \times 10^{-23}) \times 310.15 \times 0.693 \tag{9}$$

$$E_{\rm floor} = 5.82 \times 10^{-14}\,\text{J per division} \approx 10^6\,\text{ATP molecules (zero free parameters)} \tag{10}$$

The minimum Shannon entropy $H_{\rm min}(\text{class})$ is the methylation entropy expression of this floor — the minimum entropy a cell of that architectural class can sustain while remaining thermodynamically itself. Calibrated via MCMC against 37 published reference cell measurements (G-002, 5 chains, R-hat < 1.001 all parameters). The A-score (actualization score) is the ratio of observed entropy to floor:

$$A = \frac{H(\beta)}{H_{\rm min}(\text{class})} \tag{11}$$

$A = 1.00$: cell is at thermodynamic floor, operating within healthy range. $A > 1.05$: floor departure — the GAPE detection threshold. $A > 1.10$: floor breach — the cell can no longer maintain its architecture class identity.

Malignancy: Floor Breach Condition (cellular scale). A cell undergoes floor breach when $H(\beta_{\rm observed}) > H_{\rm min}(\text{class}) \times 1.05$. The observed methylation entropy exceeds the thermodynamic floor of its architecture class. The cell can no longer maintain its thermodynamic identity. Malignant behavior — invasion, uncontrolled proliferation, metabolic shift — is the consequence of this thermodynamic failure, not the failure itself. $T = T_{\rm body} = 310.15$ K. $N = 19.6 \times 10^6$ CpG sites.

3. The Structural Identity

The floor breach condition at both scales is the same equation with different parameters:

Parameter	Black Hole	Cancer Cell
Encoding surface	Bekenstein-Hawking horizon	Chromatin / DNMT1 maintenance surface
Temperature T	$T_{\text{Hawking}} = \hbar c^3/(8\pi G M k_B)$	$T_{\text{body}} = 310.15$ K
Bit count N	$A_{\text{horizon}}/(4\ell_P^2)$	$N_{\text{CpG}} = 19.6 \times 10^6$ sites
Cost per bit	$k_B T_H \ln 2$	$k_B T_{\text{body}} \ln 2 = 2.97 \times 10^{-21}$ J
Floor metric	$S_{\text{BH}} = A/(4\ell_P^2)$	$H_{\min}(\text{class})$ [G-002 MCMC posterior]
Floor breach condition	$dS_{\text{local}}/dt \geq dS_{\text{BH}}/dt$	$H(\beta)/H_{\min} > 1.05$ (A-score)
Breach event	Black hole formation	Malignant transformation
Pre-breach signal	S_{info} accumulation in IAM	Rising A-score (GAPE detection)
Scale	$\sim 10^{30}$ kg (solar mass)	$\sim 10^{-12}$ kg (cell nucleus)

The structural identity: same equation, different scales.

The Identity. Entropy is the same whether it is driving cosmic expansion or a malignant cell. The floor breach condition $F = N \times k_B T \ln 2$ appears at both ends of the observable universe. T and N differ by 37 orders of magnitude. The equation does not.

4. Empirical Confirmation

The cellular floor breach condition is not hypothetical. It has been validated against 27 of 28 TCGA cancer types (4,304 matched tumor-normal pairs) with zero free parameters and zero cancer training data. The detection threshold $A > 1.05$ was derived entirely from healthy-cell $H_{\min}$ calibration before any cancer data was examined.

Validation	Result	Parameters
G-002 MCMC $H_{\min}$ calibration	R-hat < 1.001 all 8 class posteriors	Zero free parameters
G-008 TCGA cancer validation	27/28 cancer types confirmed $A > 1.05$	Zero free parameters, zero training data
GAPE-VAL-001 field effect (live, Apr 2026)	$\Delta A = +0.09$ to $+0.22$ across 6 types, $p < 10^{-9}$	365 TCGA files, public GDC portal
GAPE-VAL-002 longitudinal (live, Apr 2026)	Per13 (leukemia, active): $A = 1.18780$, rank 1/20	Raw EPIC 850K idats, GSE130748

Empirical confirmation at the cellular scale of the structural identity.

5. What This Is and What It Is Not

This derivation establishes a structural identity — the same thermodynamic equation governs floor breach at stellar and cellular scales. It does not establish a numerical unification between the two scales. The temperatures differ by 10^{10}, the bit counts differ by 10^{27}. The equation is the same; the numbers are not.

The open problem is the quantitative bridge: showing that the specific value of $H_{\min}$ for a given architecture class can be derived from first principles through the full Landauer-Jacobson-IAM chain without reference to empirical calibration. That derivation would complete the unification. The G-002 MCMC posteriors are currently calibrated against published reference cell measurements. They are consistent with the Landauer derivation and converge with R-hat < 1.001. The full first-principles derivation of $H_{\min}(\text{class})$ from the Bekenstein-Hawking coefficient through cellular scale is the remaining open problem.

Honest Caveat. The structural derivation presented here is complete and citable. The full numerical unification — deriving $H_{\min}(\text{class})$ from first principles through the complete chain without empirical calibration — is the open problem that would make this a complete theory rather than a structural result. The structural result is what is demonstrated here.

6. Implications

If the structural identity holds — and the empirical evidence across 27 cancer types and 17 MCMC chains suggests it does — then:

Malignancy has a physics definition. A cell that has exceeded its thermodynamic maintenance budget. Not a behavioral description (invasion, proliferation). Not a genetic description (oncogene mutation). A thermodynamic state: the cell can no longer pay the cost of remaining itself.

Detection precedes the event. The A-score rises before floor breach, just as S_{info} accumulates before horizon saturation. The signal is there before the malignant phenotype appears. This is why the threshold was derived from healthy-cell physics, not cancer data.

Treatment has a physics target. Return the cell to a thermodynamic state where it can maintain its own identity. Not kill the tumor. Not correct the mutation. Restore the thermodynamic budget. The E5 intervention solver in GAPE computes this target directly.

Appendix References

[1] Bekenstein, J. D. (1973). Black holes and entropy. *Phys. Rev. D* 7, 2333–2346. doi:10.1103/PhysRevD.7.2333

[2] Hawking, S. W. (1975). Particle creation by black holes. *Commun. Math. Phys.* 43, 199–220. doi:10.1007/BF02345020

[3] Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* 5, 183–191. doi:10.1147/rd.53.0183

[4] Jacobson, T. (1995). Thermodynamics of spacetime: the Einstein equation of state. *Phys. Rev. Lett.* 75, 1260–1263. doi:10.1103/PhysRevLett.75.1260

[5] Cai, R.-G. & Kim, S. P. (2005). First law of thermodynamics and Friedmann equations of FRW universe. *JHEP* 02, 050. doi:10.1088/1126-6708/2005/02/050

[6] Mahaffey, H. W. (2026a). The Geometric Origin of the Bekenstein-Hawking Entropy Coefficient. *Zenodo.* doi:10.5281/zenodo.18702042

[7] Mahaffey, H. W. (2026b). Thermodynamic Operating Constraints of Mammalian Somatic Cell Architecture Classes. *Zenodo.* doi:10.5281/zenodo.19547624

[8] Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330. doi:10.1038/nature14248

[9] TCGA Research Network (2013–2020). Pan-Cancer Atlas. Multiple journals. See https://github.com/hmahaffeyges/IAM-Validation for full list.

[10] Planck Collaboration (2018). Planck 2018 results VI: Cosmological parameters. *A&A* 641, A6. doi:10.1051/0004-6361/201833910

Floor Breach: A Unified Thermodynamic Derivation · Mahaffey (2026) · doi:10.5281/zenodo.18702042 · IAMPerformance · heath@iamperformance.net